



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**ANÁLISIS ESTADÍSTICO DE LOS DETERMINANTES DE
DESEMPEÑO DEL PROGRAMA DE ACCESO A
SERVICIOS DIGITALES EN BIBLIOTECAS PÚBLICAS**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIA

P R E S E N T A:

PATRICIA SILVA ROMÁN



**DIRECTOR DE TESIS:
M. EN D. YVON ANGULO REYES
2015**

México, D.F.



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.

Hoja de Datos del Jurado

1. Datos del alumno

Silva

Román

Patricia

56 71 65 33

Universidad Nacional Autónoma de México

Facultad de Ciencias

Actuaría

305219396

2. Datos del Tutor

M. en D.

Yvon

Angulo

Reyes

3. Datos del Sinodal 1

Dra.

María Josefa

Santos

Corral

4. Datos del Sinodal 2

Dra.

María del Pilar

Alonso

Reyes

5. Datos del Sinodal 3

M. en C.

José Salvador

Zamora

Muñoz

6. Datos del Sinodal 4

M. en C.

Alejandro

Mina

Valdés

7. Datos del trabajo escrito

Análisis estadístico de los determinantes de desempeño del Programa de Acceso a Servicios Digitales en Bibliotecas Públicas

164 p

2015

Agradecimientos

Quiero agradecer a todos los que forman parte de mi historia y comparten conmigo la alegría de este momento.

A mis amigos, familiares y directora de tesis que siempre me brindaron apoyo y sabiduría.

En especial a mis padres y hermana que fueron mi motivación y conocen el significado que este paso tiene para mí, los amo incondicionalmente.

Índice

Introducción	5
a) Objetivos	7
i. Objetivo general.....	7
ii. Objetivos específicos.....	7
CAPÍTULO 1. El Programa de Acceso a Servicios Públicos en Bibliotecas Públicas	8
a) Tecnologías de la información y comunicación y Bibliotecas públicas	8
i. ¿Qué son las TIC?	8
ii. Proyectos internacionales de acceso a las TIC	9
b) El Programa de Acceso a Servicios Digitales en Bibliotecas Públicas.....	12
i. Antecedentes	12
ii. Importancia	14
iii. Objetivos	15
iv. ¿Quiénes participan en él?.....	15
c) Proyecto de evaluación al PASDBP (realizado por el IIS)	19
i. Breve historia del proyecto de evaluación.....	19
ii. Metodología del proyecto de evaluación	20
iii. Principales resultados del proyecto de evaluación.....	23
CAPÍTULO 2. Construcción y análisis estadístico de los factores de desempeño PASDBP	25
a) Definición del Desempeño del PASDBP y de los factores que pudieran estarlo influyendo	27
i. Definición de la variable dependiente	28
ii. Definición de las variables independientes.....	30
b) Base de datos para el análisis.....	36
i. Bibliotecas de estudio para el análisis.....	37
ii. Preguntas de los cuestionarios para el análisis.....	37
c) Construcción de la variable dependiente y de las variables independientes	38
i. Índices e Indicadores Sociales.....	38
ii. Construcción del Índice “Desempeño del PASDBP”	40
iii. Resultados del Índice de “Desempeño del PASDBP”	47
iv. Indicadores de los factores que afectan el desempeño del PASDBP	54

d) Análisis exploratorio de las variables	60
i. Análisis bivariado de la variable dependiente vs. las variables independientes	60
CAPÍTULO 3. Aspectos teóricos de la Regresión Ordinal Múltiple	66
a) Técnicas Multivariadas y Regresión Estadística	66
b) Regresión Ordinal	69
i. Modelo General	69
ii. Supuestos	74
iii. Estimación del modelo	76
iv. Bondad de ajuste del modelo	79
c) Modelo de regresión ordinal aplicado al desempeño del PASDBP	84
CAPÍTULO 4. Análisis estadístico múltiple de los determinantes del desempeño del PASDBP	86
a) Construcción y análisis de los modelos de regresión ordinal para cada uno de los grupos	86
i. Modelo con las características de localidad	87
ii. Modelo con las características del MSD	95
iii. Modelo con las características del encargado de módulo	107
iv. Modelo con la difusión	117
b) Construcción del modelo de regresión ordinal final del desempeño del PASDBP	122
Conclusiones y Recomendaciones	136
Anexo 1. Construcción de la base de datos para el análisis	141
Anexo 2. Detalles de la construcción Índice “Desempeño del PASDBP”	143
Anexo 3. Métodos de Estratificación	155
Anexo 4. Pruebas de asociación para el análisis bivariado	158
Bibliografía	162

Introducción

El análisis de programas sociales es un proceso de aprendizaje para los actores involucrados, que está encaminado a determinar de forma sistemática e imparcial la eficacia y cumplimiento de los objetivos mismos de los programas.

Un aspecto importante que se puede evaluar en los programas sociales es su desempeño, ya que éste básicamente determina el éxito o fracaso del programa estudiado. Y más allá del veredicto, la información que se puede obtener al realizar el análisis del desempeño es de gran importancia porque permite retroalimentar la toma de decisiones para introducir, modificar, o mantener medidas, etc. Acompañando la implementación y desarrollo del programa, siempre y cuando se realicen antes de que concluya el programa.

Entre los diversos programas que se han implementado en México¹ para abatir la brecha tecnológica, se encuentra el Programa de Acceso a Servicios Digitales en Bibliotecas Públicas (PASDBP), el cual tuve la oportunidad de conocer en el desarrollo de mi servicio social en el Instituto de Investigaciones Sociales, UNAM² (IIS-UNAM). El PASDBP planteó como objetivo principal contribuir a mejorar la calidad de vida de las personas, principalmente en comunidades de bajos recursos, brindándoles acceso a las Tecnologías de la Información y la Comunicación (TIC) a través de la Red Nacional de Bibliotecas Públicas (Santos, De Gortari, 2006). Para lo cual, el programa brindaría computadoras a Bibliotecas Públicas; en las que se tendría acceso no sólo a paquetería sino también a internet. En el PASDBP incidieron tres componentes importantes: equipamiento, capacitación y evaluación³.

Las evaluaciones de las que fue sujeto el programa se realizaron por un equipo de investigadores del IIS-UNAM. La primera evaluación comenzó en 2004 y concluyó en 2006, y estuvo enfocada en evaluar el proceso de equipamiento con la finalidad de generar una retroalimentación que permitiera reorientar determinados aspectos. Mientras que la segunda evaluación, realizada durante el 2011, tuvo como principal objetivo evaluar sus primeros impactos a 8 años de iniciarse el PASDBP.

¹ Algunos ejemplos: e-México, el proyecto de Plazas Comunitarias del Instituto Nacional de Educación para adultos (INEA), los Centros Comunitarios de Aprendizaje (CCA) y el proyecto México Conectado.

² El Instituto de Investigaciones Sociales (IIS) de la UNAM, estuvo a cargo de la evaluación del cumplimiento de los objetivos inicialmente planteados por el programa, del cual se desprende el libro *Acceso Tecnológico: Una reinterpretación de la biblioteca mexicana*.

³ Para llevar a cabo el componente de evaluación se realizó un levantamiento de información, en el que su estrategia general se basó en: 1) Encuestas Nacionales, 2) Entrevistas Semiestructuradas y 3) Diarios de Campo.

En este estudio se pretende aportar elementos que complementen las evaluaciones anteriormente hechas al PASDBP⁴, pero en este caso analizando las características de las bibliotecas en las que el programa tuvo un buen desempeño y en las que no. Entre las múltiples características que influyen en el desempeño del programa, las que se analizan en este trabajo son las que hacen referencia a las características de la localidad, del módulo, del encargado y de difusión.

Para evaluar el impacto de este conjunto de variables (variables independientes) en el desempeño del programa (variable dependiente) se aplicó un modelo de regresión ordinal.

El trabajo de esta tesis se desarrolló en diferentes etapas, cada una con actividades específicas. En la primera se realizó la revisión de toda la documentación del programa, así como de las evaluaciones realizadas al PASDBP, y se investigó el contexto en el cual se desarrolló el programa, lo cual se presenta en el capítulo uno. En la siguiente etapa se definió teóricamente lo que se entendió como desempeño del programa, así como al conjunto de variables independientes y se construyeron empíricamente los índices e indicadores asociados a estas definiciones (capítulo dos). En la tercera etapa, que comprende los capítulos tres y cuatro, se abordaron los aspectos teóricos de la regresión ordinal y finalmente se presentan y analizan los resultados de la aplicación de esta técnica estadística a la evaluación del desempeño del PASDBP.

El análisis estadístico de este trabajo parte de un análisis exploratorio para conocer la naturaleza de las variables, después se lleva a cabo un análisis bivariado donde se determina si existe asociación entre cada una de las variables independientes y la variable dependiente, y se concluye con un análisis más complejo al desarrollar un modelo de regresión ordinal que proporciona un mejor entendimiento del desempeño del programa con el estudio de múltiples variables de manera simultánea.

Gracias a la experiencia de realizar esta tesis, tuve la oportunidad de aplicar los conocimientos adquiridos en la carrera de Actuaría mediante un análisis multivariado con herramientas estadísticas de inferencia, lo cual me permitió tener una visión más amplia de las bondades del análisis estadístico en el ámbito social.

⁴ Un punto importante para desarrollar esta tesis es que se dispone de la información necesaria para llevar a cabo un análisis estadístico multivariado, gracias al levantamiento de las tres Encuestas Nacionales dentro del componente de evaluación. Lo que generó información del programa que permite combinar métodos cualitativos y cuantitativos.

a) Objetivos

i. Objetivo general

El objetivo general de esta tesis es identificar los factores que en mayor medida determinan el nivel de Desempeño del Programa de Acceso a Servicios Digitales en bibliotecas Públicas (PASDBP). Con tal fin se desarrolló un modelo de regresión ordinal para describir el comportamiento de la variable respuesta (Desempeño del PASDBP) explicada a través de la relación con los factores que lo determinen en mayor medida.

De esta manera, se podrá tener un mejor bosquejo de los elementos que están catalizando el desarrollo del programa y de igual forma los que están frenando su operación. Se espera que este análisis aporte información para entender la dinámica de operación del Programa de Acceso a Servicios Digitales en Bibliotecas Públicas, debido a que es un programa con alcance a nivel nacional y su desempeño se da de manera diferenciada. Igualmente se espera poder sugerir recomendaciones que conlleven al logro de los resultados esperados del PASDBP o de programas sociales similares, lo que podría ayudar directamente a las personas a cargo de la operación y diseño de los programas e indirectamente a los usuarios.

ii. Objetivos específicos

- A. Generar un índice de “Desempeño del PASDBP”, es decir, medir el desempeño del programa a través de la percepción de usuarios y encargado de módulo/ bibliotecario-encargado de módulo en cada una de las bibliotecas de estudio. El índice “Desempeño del PASDBP” será nuestra variable dependiente.
- B. Realizar un análisis bivariado de la variable dependiente “Desempeño del PASDBP” con cada una de las variables independientes, para poder establecer si existe algún tipo de asociación entre éstas.
- C. Evaluar mediante una regresión ordinal la relación causal entre el “Desempeño del PASDBP” y las variables de las dimensiones:
 - a) Características de la localidad
 - b) Características del Módulo de Servicios Digitales
 - c) Características del Encargado de Módulo
 - d) Característica de Difusión
- D. Modelar el “Desempeño del PASDBP” mediante una regresión ordinal con las características que en mayor medida influyan en el funcionamiento del MSD⁵ como las variables independientes.

⁵ Módulo de Servicios Digitales.

CAPÍTULO 1. El Programa de Acceso a Servicios Públicos en Bibliotecas Públicas

a) Tecnologías de la información y comunicación y Bibliotecas públicas

i. ¿Qué son las TIC?

Las Tecnologías de la Información y la Comunicación (TIC)⁶ son un conjunto de técnicas y elementos utilizados en el proceso, manipulación y transmisión de las informaciones. En cuanto a informaciones se refiere a informática, telecomunicaciones y principalmente a internet.

Otra de sus definiciones es: "Las tecnologías de la información y la comunicación no son ninguna panacea ni fórmula mágica, pero pueden mejorar la vida de todos los habitantes del planeta. Se disponen de herramientas para llegar a los Objetivos de Desarrollo del Milenio, de instrumentos que harán avanzar la causa de la libertad y la democracia, y de los medios necesarios para propagar los conocimientos y facilitar la comprensión mutua" Kofi Annan⁷.

Se considera a las TIC como un concepto dinámico, ya que la tecnología va innovando cada vez más rápido, lo que supone que la nueva tecnología de la comunicación e información se va transformando y encuentra su significado según la definición actual, es decir, según su marco histórico⁸. Actualmente el internet es una de las TIC's más importantes y de mayor referencia porque facilita cada vez más el acceso a la información reduciendo los límites en un entorno global y el espacio-tiempo a gran velocidad, con el propósito de facilitar el acceso y la recuperación de información que requiere la sociedad actual.

La introducción de las Tecnologías de la Información y Comunicación plantea nuevos retos para las sociedades debido a que las transformaciones tecnológicas han provocado cambios relacionados con las formas de trabajo, maneras de interacción y comunicación entre los diversos sectores sociales. Las TIC también son una puerta hacia la Sociedad de la Información⁹, que es aquella sociedad en la cual las tecnologías facilitan la creación, distribución y manipulación de la información.

⁶ O bien NTIC para nuevas tecnologías de la información y de la comunicación

⁷ Secretario general de la Organización de las Naciones Unidas, discurso inaugural de la primera fase de la WSIS, Ginebra 2003.

⁸ Un marco histórico es la época en la que se desarrolla el problema, tema investigado, las características sociales, económicas, etc..

⁹ La concepción actual de lo que se entiende por sociedad de la información es influjo de la obra del sociólogo japonés Yoneji Masuda, quién en 1981 publicó (Editorial World Future Society, Colombia), traducido al castellano en 1984 como La sociedad informatizada como sociedad post-industrial (Madrid, ed.Fundesco-Tecnos).

Entonces se puede ver que los usos de las TIC no paran de crecer y de extenderse, sobre todo en los países ricos y en desarrollo, lo que supone un riesgo social y tecnológico, ya que empieza a marcarse la llamada Brecha Digital, y sus consecuencias como la desigualdad informática y analfabetismo tecnológico se ven reflejados en la falta de oportunidades laborales y educativas principalmente. No es difícil ver que el acceso a las TIC se ha convertido en una medida para evaluar el desarrollo de un país, sería imposible negar que la Tecnología posee un papel muy importante en el desarrollo de todo tipo de ámbitos.

Asimismo la educación es un elemento fundamental en el proceso de incorporar las TIC a las actividades diarias, y también determina el avance de la Sociedad de la Información (un ejemplo de esto es el e-learning¹⁰). Esto introduce también el problema de la poca capacidad que tiene la escuela para absorber las nuevas tecnologías; y en este sentido se encuentra un nuevo concepto: Nuevas Tecnologías Aplicadas a la Educación (NTAE) que son aquellos recursos para la enseñanza como medio para el aprendizaje, como medios de comunicación y como objeto de aprendizaje y reflexión.

En México, uno de los objetivos para la igualdad de oportunidades es impulsar el desarrollo de las TIC en el ramo educativo.¹¹ Por lo que las escuelas, como otras instituciones, están replanteando sus metas pedagógicas y didácticas para reinventarse alrededor de las nuevas oportunidades abiertas por la tecnología de la información. Las redes educativas virtuales se están convirtiendo en las nuevas bases del sistema educativo que incluyen nuevos escenarios educativos, nuevos instrumentos educativos y la formación de nuevos educadores especializados en la enseñanza de un nuevo espacio social.

Algunos de los beneficios más claros que las TIC, dentro de los medios de comunicación, que aportan a la sociedad son el acceso a la cultura y la educación, donde los adelantos tecnológicos y los beneficios que conlleva la era de la comunicación lanzan un balance y unas previsiones extraordinariamente positivas. En este punto es importante señalar que debe existir una relación estrecha entre la información que se provee y la capacidad de asimilación de la misma por parte de las personas expuestas a ella; por esto, es conveniente una adecuada educación en el uso de estos poderosos medios.

ii. Proyectos internacionales de acceso a las TIC

La visión del acceso a las Tecnologías de Información y Comunicación (TIC) como herramientas de desarrollo científico, cultural, social y económico es compartida por los gobiernos alrededor del mundo. Ya que sus metas no compiten con otras metas sociales que se establecen en comunidades

¹⁰ El e-learning es el tipo de enseñanza que se caracteriza por la separación física entre el profesor y el alumno, y que utiliza internet como canal de distribución del conocimiento y como medio de comunicación. Los contenidos de e-learning están enfocados en las áreas técnicas. A través de esta nueva forma de enseñar al alumno y el docente pueden administrar su tiempo, hablamos de una educación asincrónica.

¹¹ «Fuentes electrónicas» <http://pnd.calderon.presidencia.gob.mx/igualdad-de-oportunidades/transformacion-educativa.html> (Consulta 12/04/2015).

que carecen de desarrollo no solamente tecnológico, sino también carencias de servicios prioritarios.

Según información de la Fundación Gates: “En 1997, se pusieron en marcha dos grandes proyectos: el Programa de Bibliotecas en Estados Unidos (*U.S. Library Program*) y la iniciativa Internacional de Bibliotecas (*International Library Initiatives*)¹². Ambos programas fueron concebidos con la idea de que la introducción de computadoras conectadas a internet en bibliotecas, podrían ayudar a aquellas personas que de otra manera no tendrían acceso a dicha tecnología” (Santos, de Gortari, Angulo, 2006, 22). La elección de las bibliotecas públicas como recinto físico de estos programas se debe a que las bibliotecas son espacios neutrales, seguros y centrales, sus servicios son gratuitos, son portales de información y están disponibles para todas las personas.

Como resultado del proyecto de la Fundación Gates *U.S. Library Program*, actualmente en Estados Unidos las bibliotecas cuentan con computadoras como parte de su infraestructura básica y 98% cuenta con acceso a internet; mientras que el 40% de los estadounidenses no tienen acceso a internet en casa¹³. Estas cifras revelan la importancia que ha tenido el programa en E.U.A., ya que el alcance de las computadoras e internet en las bibliotecas se encuentra en su punto más alto jamás antes visto. Sin duda, en estos tiempos económicos difíciles, las bibliotecas se han convertido en un recurso esencial para la comunidad.

“Bajo la premisa de que brindar acceso igualitario a las TIC tendría un impacto social y económico importante, la Fundación Bill y Melinda Gates ha apoyado los esfuerzos de proveer y mantener el acceso público y gratuito de computadoras y conexión a internet en diez países de todo el mundo, a través de *International Library Initiatives*” (Santos, de Gortari, Angulo, 2006).

Los países que participan actualmente en *International Library Initiatives* son México, Chile, Colombia, Botswana, Lituania, Letonia, Rumania, Ucrania, Polonia, Bulgaria, Moldova, Vietnam e Indonesia. También se han coordinado programas de menor escala en Nepal, Bhután, la India, Guatemala y Honduras.¹⁴ Al ser diferentes países tienen diferentes necesidades, capacidades y prioridades, en consecuencia la iniciativa se esfuerza en crear programas que sean sensibles a las condiciones locales y nacionales. Esto ha sido posible gracias al trabajo en equipo que realiza la Fundación Gates y los gobiernos de cada país.

Por otra parte, también gracias a la experiencia de los programas, tanto en los E.U.A. como en el extranjero, se ha aprendido que es mucho más sencillo realizar la instalación de computadoras y conexiones a internet de lo que es mantenerlo. Lo que conlleva a plantear como solución que las bibliotecas fijen planes a largo plazo de financiación para mantener al día equipos y servicios de

¹² «Fuentes electrónicas» <http://www.gatesfoundation.org/es/What-We-Do/Global-Development/Global-Libraries> (Consulta (12/04/2015)

¹³ «Fuentes electrónicas» <http://www.gatesfoundation.org/libraries/> (Consulta 12/04/2015)

¹⁴ «Fuentes electrónicas» <http://www.gatesfoundation.org/es/What-We-Do/Global-Development/Global-Libraries> (Consulta (12/04/2015)

internet. Esta característica de financiamiento, pasó a formar parte importante de las estrategias internacionales de *International Library Initiatives*.

A continuación se presenta un pequeño resumen de los primeros proyectos de *International Library Initiatives* (programas homólogos al PASDBP) y de los resultados de la experiencia en estos programas.

En octubre de 1998 la Fundación Gates dio inicio a su programa de apoyo a las bibliotecas Canadienses ubicadas en zonas de bajos recursos, mediante el proyecto *Provincewide Library Grant for Public Access to Computing and the Internet*. Para el año de 1999, Canadá se convirtió en el primer país en tener todas sus bibliotecas públicas y escuelas conectadas a internet.

En 1999, la Fundación Gates estableció un convenio de colaboración *The People's Network Project* con la Comisión de Bibliotecas e Información¹⁵ Inglesa, para dotar de computadoras con acceso a internet y capacitación del personal de estas instalaciones en las zonas más marginadas de Gran Bretaña. Sin embargo, el panorama para las bibliotecas en Gran Bretaña no parece correr con la misma suerte que Canadá. Debido a problemas presupuestales (recorte del gasto público por iniciativa del Gobierno británico) en 2011, más de 300 bibliotecas públicas fueron cerradas y posiblemente otras mil más (de las 4500 existentes) siguen en la mira de tener el mismo problema¹⁶.

En América Latina, en 2001 en Chile, el proyecto *BiblioRedes: Abre tu mundo*¹⁷, estableció el acceso gratuito a internet en 381 bibliotecas públicas a lo largo del país, fue concebido en colaboración con la Dirección de Bibliotecas, Archivos y Museos de Chile (DIBAM) dentro de uno de los ejes de su fundación "Educación e información". Este fue el primer programa en América Latina por parte de *Global Libraries* y se decidió partir del modelo empleado en Estados Unidos. Además del éxito alcanzado por este programa durante años de trabajo, se ha hecho merecedor de diversos premios nacionales e internacionales¹⁸. Actualmente el programa *BiblioRedes* sigue vigente y está presente en 412 bibliotecas públicas y 18 laboratorios regionales en todo Chile. Como parte de sus resultados, desde inicios del 2008 se implementó en 60 de las bibliotecas el servicio de WIFI para mejorar el servicio específicamente en zonas turísticas¹⁹.

En 2003 la Fundación de Bill y Melinda Gates incorporó a México en el Proyecto *Global Libraries* mediante el PASDBP, con el cual a la fecha se han equipado 1,700 bibliotecas públicas²⁰

¹⁵ Esta entidad actualmente lleva el nombre de Comisión de Bibliotecas, Museos y Archivos (MLA, por sus siglas en inglés).

¹⁶ «Fuentes electrónicas» <http://www.actualidadeditorial.com/bibliotecas-gran-bretana-cierre/> y <http://www.publico.es/espana/359781/las-bibliotecas-publicas-de-gran-bretana-se-niegan-a-desaparecer/> (Consulta 12/04/2015)

¹⁷ «Fuentes electrónicas» <http://www.biblioredes.cl/> (Consulta 12/04/2015)

¹⁸ El más importante de ellos es el *Stockholm Challenge*, más conocido como el Premio Nobel de Internet, que el año 2006 fue otorgado en la categoría cultura a *BiblioRedes* por el Instituto Real de Tecnología de Estocolmo.

¹⁹ «Fuentes electrónicas» <http://www.biblioredes.cl/acceso-internet> (Consulta 12/04/2015)

²⁰ En México las bibliotecas públicas surgieron en 1920, pero no fue sino hasta 1980 cuando se estableció el Programa Nacional de Bibliotecas Públicas que aumentaron en una cifra considerable, lo cual permitió un

con módulos de acceso gratuito a equipos de cómputo e internet. Y precisamente es el programa al que se hace referencia en esta tesis.

b) El Programa de Acceso a Servicios Digitales en Bibliotecas Públicas

i. Antecedentes

A partir del año 2000, el gobierno de México empezó a promover una serie de políticas públicas encaminadas a reducir la brecha digital²¹. Se pretendía brindar el acceso de internet de forma gratuita y de calidad a toda la población, específicamente se buscaba apoyar a las personas en situación de pobreza que difícilmente podrían acceder a la tecnología.

Como parte de estas preocupaciones, en 2001 se diseñó el Sistema Nacional e-México que surgió como una Política Pública que integraría y articularía los intereses de los distintos niveles de gobierno, de diversas entidades y dependencias públicas, de los operadores de redes de telecomunicaciones, de las cámaras y asociaciones vinculadas a las Tecnologías de la Información y las Comunicaciones (TICs), así como de diversas instituciones, a fin de ampliar la cobertura de servicios básicos como educación, salud, economía, gobierno y ciencia, tecnología e industria, así como de otros servicios a la comunidad. Su objetivo principal era generar alternativas de valor a través de un sistema tecnológico con contenido social, que ofrecían las herramientas y oportunidades que son posible alcanzar por medio de las Tecnologías de la Información y las Comunicaciones para mejorar la calidad de vida de todos los mexicanos. El sistema estaba organizado en tres grandes ejes: conectividad, contenidos y sistema²², y precisamente para lograr el objetivo del eje de conectividad que promovía el acceso a internet y la creación de diversos contenidos digitales (vía internet) se instalaron Centros Comunitarios Digitales (CCD). Los CCD son módulos que cuentan con computadoras e internet, y se ubicaron estratégicamente en las cabeceras municipales, esto con la finalidad de formar una red satelital a nivel nacional.

A principio del actual gobierno, este proyecto fue transformado a “México Conectado”²³. El cual es un proyecto del Gobierno de la República que contribuye a garantizar el derecho constitucional de acceso al servicio de Internet de banda ancha (artículo 6to constitucional). Para

verdadero despliegue a nivel nacional que se completó en el año 2000 cuando alcanzaron 6,109, de las cuales el 90% estaban ubicadas en cabeceras municipales; para fines de 2006 eran 7,210 bibliotecas. Actualmente son unas 7,700 en todo el país.

²¹ En sus inicios fue descrita como el acceso desigual a las tecnologías de la información y a las computadoras y el internet, a consecuencia de factores como los ingresos, la geografía, la edad, las diferencias étnicas, etc. Con el tiempo ha evolucionado hacia un sentido más amplio que define las disparidades en el uso de la tecnología como resultado de la falta de acceso, de habilidades o de encontrarle sentido y /o interés al uso de estas herramientas.

²² *Resumen Ejecutivo e-México, Secretaría de Comunicaciones y Transportes, México D.F. 2002.* «Fuentes electrónicas» <http://www.sct.gob.mx/informacion-general/areas-de-la-sct/coordination-de-la-sociedad-de-la-informacion-y-el-conocimiento/el-sistema-nacional-e-mexico/> (Consulta 12/04/2015)

²³ «Fuentes electrónicas» <http://mexicoconectado.gob.mx/> (Consulta 12/04/2015)

lograr dicho objetivo, México Conectado promueve el despliegue de redes de telecomunicaciones que proveen conectividad en los sitios y espacios públicos tales como escuelas, centros de salud, bibliotecas, centros comunitarios o parques, en los tres ámbitos de gobierno: federal, estatal y municipal.

México Conectado lleva a cabo distintas acciones con el objetivo de potenciar el uso y aprovechamiento que se hace de la conectividad en los más de 45 mil sitios y espacios públicos del país en los que provee acceso al servicio de Internet de banda ancha. De manera paralela, en el marco de México Conectado, se llevan a cabo algunos programas de inclusión digital, que hacen uso de las tecnologías de la información y la comunicación para atender poblaciones específicas como mujeres y jóvenes. Estos programas son: Programa de formación y certificación para promotores digitales y Mujer Migrante

Otro proyecto del gobierno que tuvo inicio en 2001, son las Plazas Comunitarias del Instituto Nacional de Educación para Adultos (INEA). Las Plazas comunitarias nacen como una fórmula de combate al rezago educativo dentro del programa de mediano plazo del Consejo Nacional de la Educación para la Vida y el Trabajo (CONEVYT). Las plazas comunitarias son espacios educativos abiertos a la comunidad; en ellas se ofrecen prioritariamente programas y servicios educativos para las personas jóvenes y adultas que no han concluido su educación básica y la certificación de sus contenidos. Dentro de los servicios que ofrecen se encuentra el acceso a las nuevas Tecnologías de Comunicación e Información para el desarrollo de la lectura, escritura, comunicación, razonamiento lógico e informática que faciliten el ingreso de las personas al empleo. Este servicio se lleva a cabo en las instalaciones de Sala de cómputo, que es un aula con computadoras en red y con acceso a través de Internet a información, materiales, cursos, ejercicios, bibliotecas digitales, instituciones educativas, programas de capacitación para el trabajo y sitios electrónicos que complementan la formación de los educandos²⁴.

También los Centros Comunitarios de Aprendizaje (CCA) por parte de la Secretaría de Desarrollo Social (SEDESOL) forman parte de los programas que contribuyen a reducir la brecha digital. Un Centro Comunitario de Aprendizaje (CCA) es un espacio destinado al desarrollo de capacidades y habilidades mediante el uso de computadoras con acceso a Internet. Está enfocado principalmente a la población que habita en zonas marginadas del territorio nacional, a fin de proporcionar cursos de educación formal y para el trabajo, que de otra forma sería más difícil su acceso. Los centros se instalan en convenio con los gobiernos estatales y municipales, así como con organismos de la sociedad civil y del sector privado²⁵.

La puesta en marcha de estos programas fue el primer paso para que México creara espacios donde las personas tuvieran oportunidad de acceder a herramientas tecnológicas, en especial a las computadoras e internet. Es en este ambiente que surge el Programa de Acceso a Servicios Digitales

²⁴ «Fuentes electrónicas»

<http://www.inea.gob.mx/index.php/proyectosbc/pestrategicosplazasbc/pestplazasdescrb.html> (Consulta 12/04/2015)

²⁵ «Fuentes electrónicas» <http://www.microrregiones.gob.mx/> (Consulta 12/04/2015)

en Bibliotecas Públicas (2003), siendo un proyecto novedoso que comparte su visión e implementación en otras partes del mundo donde tiene sus predecesores y que marca un nuevo concepto en nuestro país al ser una iniciativa con una estructura bipartita, es decir, un trabajo en conjunto del gobierno federal con la Fundación Bill y Melinda Gates.

ii. Importancia

En 2003 (previo al inicio del PASDBP), sólo el 2% de las bibliotecas públicas en México contaban con computadoras de uso público²⁶. Lo que significa que para ser un espacio comunitario que apoya la educación y la cultura, estaban rezagadas y necesitaban un cambio para adecuarse a la realidad tecnológica de la información.

Por consiguiente, el Programa de Acceso a Servicios Digitales en Bibliotecas Públicas representó una puerta hacia el progreso tecnológico en México y un puente para disminuir la brecha digital, que en la última década se ha acentuado, no solamente en el contexto de nuestro país sino en el mundo²⁷. Con este programa, la infraestructura bibliotecaria del país se fortaleció y modernizó tecnológicamente, permitiendo a todos los usuarios de diferentes edades, tener acceso a computadoras, y sobre todo a las nuevas tecnologías, de manera gratuita en 1,700 de las 7,700²⁸ bibliotecas públicas de todo el país.

Si bien para entonces existían algunas ofertas tecnológicas, como las Plaza Comunitarias del INEA, los Centros Comunitarios Digitales (CCD) o los Centros Comunitarios de Aprendizaje (CCA), que pudieran ser opciones²⁹ para compensar este rezago tecnológico; los diferentes enfoques y alcances de ellas no permitían que todas las personas que lo necesitaran tuvieran la oportunidad de acceder a computadoras e internet con lo cual se limitaba el impacto de sus beneficios a toda la población. Por ejemplo, la población objetivo de las Plazas Comunitarias son los adultos que no han concluido su educación básica. Esto significa que el uso del equipo se restringe únicamente a un servicio educativo de certificación para un sector de la población.

También pareciera que la ausencia de vinculación entre esos programas impedía que se complementaran entre sí y que, en el peor de los casos, algunas localidades coexistieran con gran variedad de ofertas tecnológicas, a veces sin saberlo, y que en otras localidades hubiera una carencia total de ellas. Lo que muestra que no basta sólo con brindar la infraestructura y poner en marcha un programa, como sucedió en los casos anteriores, sino que es necesario entender el contexto y trabajar de forma conjunta con otros programas o actores para su éxito.

²⁶ «Fuentes electrónicas» http://dgb.conaculta.gob.mx/info_detalle.php?id=33 (Consulta 12/04/2015)

²⁷ Esto se debe a la desigualdad de las riquezas de los países en el mundo y dentro de ellos, lo que genera a su vez la desigualdad para poder comprar y acceder a la tecnología.

²⁸ «Fuentes electrónicas» <http://www.eluniversal.com.mx/cultura/67193.html> (Consulta 12/04/2012)

²⁹ Nos referimos a los programas especificados en la sección anterior de antecedentes en México.

iii. Objetivos

El PASDBP surge con la idea de brindar acceso a las Tecnologías de la Información y Comunicación (TIC)³⁰ a todas las personas; en especial a las que se encuentran en situación de pobreza, y que por consecuencia no tienen posibilidad de adquirir o acceder a las TIC. Al permitir este acceso tecnológico se desea cambiar de forma significativa las oportunidades de las comunidades donde son instalados los Módulos de Servicios Digitales, mejorando su calidad de vida y permitiéndoles incorporarse a la Sociedad de la Información.

Los principales objetivos del PASDBP pueden ser resumidos de la siguiente manera³¹:

1. Disminuir la Brecha Digital; el principal objetivo del programa es brindar las posibilidades tecnológicas a aquellas comunidades que se encuentran en condiciones desfavorables y que de otra forma no podrían acceder a ellas.
2. Generar cambios de índole económico, social, cultural y tecnológico en las comunidades donde esté presente el programa; transformando de esta forma la vida de las personas y brindándoles mejores oportunidades que deriven en los cambios anteriormente mencionado.
3. Redefinir el papel de las bibliotecas públicas como espacios de acceso a la Información, mejorando con la llegada del Programa en su infraestructura y actualizándose conforme a las necesidades de la cambiante realidad tecnológica.
4. Servir de apoyo en las políticas aplicadas por el gobierno y complementar los programas anteriormente puestos en marcha con objetivos comunes; para atender de una manera integral y desde otro punto de vista, que anteriormente fue implementado en otros países, la problemática del rezago tecnológico.

iv. ¿Quiénes participan en él?

En un principio, y dentro del marco del programa “Hacia un país de lectores”³² la Dirección General de Bibliotecas (DGB) se encargó de la promoción de “El Programa de Acceso a Servicios Digitales en Biblioteca Públicas”.

En este programa participan distintas instituciones internacionales, gubernamentales y tecnológicas. Consiste, a grandes rasgos, en dotar de computadoras e internet a las bibliotecas públicas seleccionadas, y dar capacitación a los encargados de los MSD para que lleven a cabo de manera adecuada el uso e instrucción de la nueva infraestructura.

Para entender la participación de los diversos actores concernientes a la estructuración del PASDBP, se le puede dividir en las siguientes etapas de desarrollo: Financiamiento, Diseño, Implementación y Operación.

³⁰ Específicamente computadoras e internet.

³¹ “Acceso Tecnológico: Una reinterpretación de la biblioteca pública mexicana”. CONACULTA.

³² El programa “Hacia un país de lectores” formó parte del Proyecto Educativo del Estado, que el Presidente Vicente Fox Quesada dio a conocer el 28 de mayo de 2002; y actualmente ya no se encuentra vigente.

Comenzando con el financiamiento del PASDBP, se encuentra que la infraestructura fue financiada por la Fundación Bill y Melinda Gates³³, gracias a un donativo de más de 30 millones de dólares para dotar de equipo de cómputo y software a través de Microsoft México como parte del proyecto de Bibliotecas Globales (*Global Libraries*), quienes insisten que la innovación es la clave para mejorar el mundo³⁴. Y como parte del plan de sostenibilidad del programa, las autoridades municipales correrían con los gastos del internet y con el pago del salario de los encargados del módulo. Sin embargo, como afirma la Dra. Santos: “Evaluaciones realizadas a finales de 2010 por Conaculta y la UNAM encontraron que un 25% de los módulos no cuentan con internet, lo que significa que sus socios (en su mayoría presidentes municipales o jefes delegacionales, según sea el caso de los lugares donde se localizan) no cumplen con los acuerdos de financiamiento (...) Aún en esta situación, el PASDBP representa una oportunidad para que las bibliotecas públicas se actualicen a un bajo costo debido a que el pago de la conexión a internet resulta ser menor al compararse con el costo necesario para actualizar el acervo de las bibliotecas, desempeñan un nuevo rol en la instrucción de estudiantes y se convierten en lugares de acceso dentro de las comunidades ”³⁵.

Por su parte, el diseño del PASDBP fue un ejercicio en conjunto de la Fundación Gates y del Consejo Nacional para la Cultura y las Artes (CONACULTA)³⁶ a través de la Dirección General de Bibliotecas (DGB)³⁷. Combinó las bases que tenía el proyecto de Bibliotecas Globales (*Global Libraries*) a partir de un prototipo de programa puesto en marcha en otros países, pero que debía adecuarse al contexto particular de las localidades de México. Para ello el PASDBP le fue asignado a CONACULTA (porque el donativo de la Fundación Bill y Melinda Gates les fue otorgado directamente), y fue el encargado de realizar la evaluación para seleccionar las bibliotecas beneficiadas.

Para la implementación del programa, la DGB trabajó de manera individual en la capacitación de los encargados del MSD y de manera conjunta con las Autoridades Municipales o Delegacionales (según fuera el caso) al realizar acuerdos como parte de la estrategia de sostenibilidad del programa. Debido a la magnitud del proyecto, el cual originalmente tenía programado equipar

³³ Que es la fundación privada de caridad más grande del mundo, fundada por Bill Gates y Melinda Gates. Su sede está en la ciudad de Seattle, Washington y fue creada en el año de 1994 con el propósito de erradicar la pobreza mundial. «Fuentes electrónicas» <http://www.gatesfoundation.org/What-We-Do/Global-Development/Global-Libraries> (Consulta 12/04/2015)

³⁴ Carta Anual de Bill Gates«Fuentes electrónicas» <http://www.gatesfoundation.org/Who-We-Are/Resources-and-Media/Annual-Letters-List/Annual-Letter-2012> (Consulta 12/04/2015)

³⁵ «Fuentes electrónicas» <http://www.iis.unam.mx/pdfs/iismedios/julio2011/Entra%20UNAM%20en%20proyecto%20para%20digitalizar%20bibliotecas%20publicas.pdf> (Consulta 12/04/2012)

³⁶ Quien fue creado con el fin de coordinar las políticas, organismos y dependencias tanto de carácter cultural como artístico. Asimismo, tiene labores de promoción, apoyo y patrocinio de los eventos que propicien el arte y la cultura. «Fuentes electrónicas» <http://www.conaculta.gob.mx/> (Consulta 12/04/2015)

³⁷ Que surge en 1985 y que actualmente está integrada a CONACULTA. Es la dirección encargada de instrumentar el PASDBP mediante la Red Nacional de Bibliotecas conformada por 31 redes estatales y 16 redes delegacionales. «Fuentes electrónicas» <http://dgb.conaculta.gob.mx/> (Consulta 12/04/2015)

1,150³⁸ bibliotecas del total 7,700 pertenecientes a la Red Nacional de Bibliotecas, se realizó un esquema de equipamiento en diversas fases, de acuerdo con el siguiente cuadro.

FASE	FECHA	Módulos instalados
Primera Fase	Noviembre de 2002 a junio de 2004	678
Segunda Fase	Enero de 2004 a junio de 2005	474
Tercera Fase	Agosto de 2005 a octubre de 2006	1576

Fuente: "Acceso Tecnológico: Una reinterpretación de la biblioteca pública mexicana". CONACULTA.

Durante la cuarta fase de equipamiento se permitió que 3,162 recintos dispongan de recursos electrónicos de información, que equivale al 43% del total de las bibliotecas de la Red Nacional.

Obviamente cada una de estas fases tuvo internamente diferentes etapas, primordialmente se distingue:

- *Selección de Bibliotecas.* El paquete se transfirió únicamente a las bibliotecas públicas que CONACULTA seleccionó a partir de la evaluación de su infraestructura, esto con el fin de elegir a aquellas bibliotecas que tuvieran las características necesarias³⁹ (y básicas) para que funcionara el equipo de cómputo y pudiera mantenerse lo mejor posible. En algunos casos la llegada de las computadoras provocó un cambio físico, como la ampliación de la biblioteca o un cambio de domicilio a un edificio con mejores instalaciones para cumplir con las funciones tanto de la biblioteca como del MSD.
- *Acuerdos con diversas autoridades.* Los presidentes municipales de los lugares donde se localizaron las bibliotecas seleccionadas se comprometieron a seguir capacitando a los Encargados de Módulo, a prestar los servicios de mantenimiento al equipo de cómputo para que no dejaran de operar o bien reemplazarlo en caso de ser necesario, pagar el servicio de internet y sueldo de los encargados de MSD⁴⁰. En consecuencia se establecieron lazos de cooperación benéficos para la biblioteca pública con otros actores de la sociedad para promover la consolidación y permanencia tecnológica y financiera a nivel local.
- *Capacitación de los Encargados de módulo.* La capacitación está orientada en torno a dos ejes básicos. El primero se refiere a la administración del Módulo de Servicios Digitales, conocimientos generales del equipo y software y solución de problemas comunes. El segundo eje se relaciona con la transmisión de capacidades técnicas, principalmente a los usuarios y a otros actores de las bibliotecas, como los bibliotecarios.

³⁸ Es necesario indicar que al final se agregaron 88 bibliotecas y otras se sustituyeron, por lo que para la tercera encuesta se contó con 1238 bibliotecas equipadas. Por otro lado, el cambio de los equipos permitió que en la tercera fase se equiparan 1576 bibliotecas, en lugar de las 478 contempladas para esta fase. "Acceso Tecnológico: Una reinterpretación de la biblioteca pública mexicana". CONACULTA.

³⁹ En algunas bibliotecas no se contaba con instalación eléctrica por ejemplo, lo cual hacía imposible la instalación de las computadoras.

⁴⁰ CONACULTA también se encargó de la evaluación de las Autoridades Municipales durante la selección de las bibliotecas beneficiarias del programa.

Un punto importante de remarcar, es que la llegada de las computadoras implicó un cambio físico en las bibliotecas públicas, esto debido que uno de los requisitos de elegibilidad de las bibliotecas era precisamente contar con un espacio físico adecuado para garantizar el buen funcionamiento del equipo, estos espacios físicos son llamados Módulo de Servicios Digitales (MSD).

Entonces, el contexto en el cual residían las bibliotecas determinó de manera significativa la elección de las mismas para participar en el programa, aún cuando se esperaría una mayor incidencia social o aprovechamiento del MSD en aquellas localidades que carecen de los medios para satisfacer dicha condición. Ya que las bibliotecas que tenían un mayor nivel de marginación presentaban dificultades para cumplir con este requerimiento por falta de recursos económicos.

Y a pesar de las medidas que se tomaron para garantizar la correcta instalación del MSD, se encontraron diversas problemáticas a la hora de la implementación. Por ejemplo, en algunos casos, el inmueble donde estaba la biblioteca no permitía la adecuación apropiada del módulo y se tuvieron que hacer remodelaciones para ampliar el lugar o bien el cambio a otro edificio que les permitiera contar con las características deseadas del MSD.

Por último, la operación del programa se refiere a la puesta en marcha del PASDBP. Ya que se generaba un nuevo espacio tecnológico⁴¹ en comunidades donde escasamente se tenía acceso y conocimiento a esta tecnología, se necesitó la incorporación de nuevos actores para el cuidado y atención del módulo. Estos nuevos actores son los *Encargados de Módulo*, que dependiendo del tamaño del Módulo, de la biblioteca y en muchas ocasiones del presupuesto de las Autoridades Municipales o Delegacionales se decidía el número de personas encargadas y las funciones a desempeñar. Debido a esta característica se dividió a los encargados de módulo en dos tipos:

- *Encargado de Módulo*. Se refiere a la persona a cargo del módulo que únicamente atienden el Módulo de Servicios Digitales.
- *Encargado de Módulo-Bibliotecario*. Se refiere a la persona a cargo del módulo y de la biblioteca simultáneamente.

Los encargados de módulo (ya sea sólo del módulo o del módulo y biblioteca) son aquellas personas cuyas funciones principales son: ser los responsables del funcionamiento y mantenimiento del equipo del Módulo de Servicios Digitales tanto de las computadoras asignadas por el programa como de otros equipos con que se haya complementado, brindar asesoría a los usuarios para que puedan hacer uso de las herramientas tecnológicas, lo cual implica asistir a los usuarios desde las funciones más básicas como encender una computadora hasta el uso de programas más especializados⁴².

Respecto a los encargados de módulo la problemática se relacionaba principalmente con su capacitación, la cual estuvo a cargo de la Dirección General de Bibliotecas (DGB). Ya que a pesar de las facilidades y apoyos económicos que se les ofrecían para que concluyeran su capacitación, en

⁴¹ Nos referimos al Módulo de Servicios Digitales.

⁴² Esto obviamente dependerá de los conocimientos que tenga el encargado de módulo.

muchos casos se presentó el ausentismo. Por otro lado el nivel de rotación, derivado de cambios en la estructura de los gobiernos municipales, o bien la inestabilidad del personal, obstaculizó una capacitación completa.

c) Proyecto de evaluación al PASDBP (realizado por el IIS)

i. Breve historia del proyecto de evaluación

Como se mencionó anteriormente, en el PASDBP incidieron tres componentes con una visión y concepción común de la tecnología: el equipamiento, la capacitación y la evaluación. Este último componente de evaluación es el que dio lugar a la concepción del Proyecto de evaluación.

El proyecto de evaluación arrancó a fines de 2004, a petición de la Fundación Bill y Melinda Gates y CONACULTA; y estuvo a cargo de investigadores⁴³ del Instituto de Investigaciones Sociales (IIS) de la UNAM, que trabajan en líneas de investigación sobre tecnología y cultura, sociología de la ciencia, innovación y desarrollo tecnológico; y contó con la participación de investigadores especialistas en áreas como el análisis demográfico, la brecha digital y el acceso y uso de las nuevas tecnologías de comunicación e información. Se tuvo asesoría técnica y académica de la Unidad de Estudios sobre la Opinión (UDES) adscrita a la IIS-UNAM. Su objetivo principal era cubrir las necesidades de información de tres actores distintos: la Fundación Bill y Melinda Gates, CONACULTA y la de los investigadores participantes en la evaluación. Las necesidades de información resultaron variadas, ya que la Fundación Gates principalmente requería datos cuantitativos que permitieran la comparación de resultados con otros países donde se había instrumentado el proyecto; por su parte CONACULTA necesitaba información que evaluara el proceso de equipamiento con el fin de corregir (en caso de ser necesario) y afinar la operación del programa y por último los investigadores requerían datos que se relacionaran con sus líneas de trabajo.

La evaluación del PASDBP tendió a reflejar la complejidad del entorno social del Programa de Acceso a Servicios Digitales en Bibliotecas Públicas más que a calificarlo, debido a esto el proyecto de evaluación se construyó desde una perspectiva que implicaba la necesidad de considerar tanto los elementos técnicos con los sociales, como su relación entre ellos.

⁴³ María Josefa Santos Corral, Investigadora Titular "C", T. C., Definitiva con Doctorado en Antropología, sus líneas de investigación son: Tecnología y cultura; innovación tecnológica; redes y antropología industrial. Rebeca de Gortari Rabiela, Investigadora Titular "C", T.C. Definitiva con Doctorado de Tercer Ciclo en Historia; sus líneas de investigación son: Sociología de la ciencia y la innovación; innovación y desarrollo tecnológico; redes y regiones.

Yvon Angulo Reyes, Investigadora Asociada "C", T. C., Interina Maestra en Demografía con líneas de investigación: Marginalidad; Redes sociales y políticas de bienestar; Diseño muestral; Análisis estadístico

ii. Metodología del proyecto de evaluación

La metodología base propuesta por la Fundación Gates, con la cual se inició el proyecto, estaba sustentada en la que se diseñó para comparar los programas norteamericano y chileno⁴⁴, orientada a la construcción de indicadores cuantitativos que mostraran impactos a mediano y largo plazo.

Sin embargo, para evaluar el PASDBP en México el equipo de investigadores del IIS-UNAM propuso uno diferente, se replanteó la metodología en la que se incluyó trabajo cualitativo, para adecuarse a las necesidades de información que implicaba el nuevo objetivo de la evaluación que era satisfacer las necesidades de información de los tres actores (Fundación Gates, Conaculta e investigadores del IIS-UNAM) y los retos que surgieran en su ejecución.

Así que finalmente se optó por una metodología que combinó ambas perspectivas, por lo que se aplicaron métodos cualitativos y cuantitativos en su desarrollo. La estrategia propuesta para la evaluación fue un análisis longitudinal para evaluar cada etapa de equipamiento, lo que generó información en diversos momentos en el tiempo. A su vez, cada etapa de la evaluación representó una evaluación parcial y en su conjunto conformaron la evaluación global.

Los elementos que formaron cada una de estas etapas son los siguientes:

- **Encuestas.** Su objetivo era generar información que permitiera analizar de forma cuantitativa los posibles efectos del PASDBP.
- **Diarios de campo**⁴⁵. Permitirían conocer y entender el entorno de la comunidad y la dinámica que rodea la biblioteca y el MSD.
- **Entrevistas a profundidad**⁴⁶. Su objetivo era conocer la percepción específica y detallada de algunos actores en torno a los servicios digitales y su rol en el funcionamiento del mismo.

La parte cualitativa se desarrolló principalmente con la información obtenida en Entrevistas a profundidad y diarios de campo, mientras que la parte cuantitativa se apoyó de la información obtenida en las encuestas, y a partir de éstas se hicieron tablas de frecuencias y cruces.

Características de las Encuestas Nacionales

Para comenzar la descripción de las características de las Encuestas Nacionales se presenta un cuadro con las fechas específicas del levantamiento de la información.

⁴⁴ El programa americano es U.S. Library Program y el programa chileno es Biblioredes: Abre tu mundo; ambos apoyados por la Fundación Gates.

⁴⁵ En total se realizaron 35 etnografías. En el primer año se seleccionaron 20 comunidades y 15 comunidades en el segundo año, cuyas bibliotecas se encuentran en diferentes fases de equipamiento del programa.

⁴⁶ En total se aplicaron 350 entrevistas a profundidad a usuarios, no usuarios, bibliotecarios, encargados de módulo, autoridades municipales y encargados de la instrumentación del proyecto.

Encuesta	Fecha
Primera Encuesta	Febrero de 2005
Segunda Encuesta	Septiembre de 2005
Tercera Encuesta	Abril – Mayo de 2006

Fuente: "Acceso Tecnológico: Una reinterpretación de la biblioteca pública mexicana". CONACULTA.

A continuación se describen los objetivos de cada encuesta en las diferentes etapas de evaluación.

- *Primera Encuesta:* Generación de información para evaluar la primera etapa de equipamiento, información para comparar las dinámicas de bibliotecas equipadas y no equipadas, analizar usuarios y no usuarios de los MSD, generar información de las bibliotecas no equipadas.
- *Segunda Encuesta:* Generar información para evaluar la segunda etapa de equipamiento, proporcionar información para analizar el proceso de antes y después del equipamiento⁴⁷, generar información para el análisis de bibliotecas equipadas⁴⁸ que fueron encuestadas en la primera encuesta.
- *Tercera Encuesta:* Generar información para evaluar la tercera etapa de equipamiento; para el análisis de seguimiento de bibliotecas equipadas en la primera fase e incluidas en la primera y en la segunda encuesta⁴⁹; y proporcionar información para el análisis de seguimiento de las bibliotecas equipadas en la segunda fase e incluidas en la segunda encuesta.

Por último, para resumir la información del diseño de las muestras se presenta el siguiente cuadro:

⁴⁷ Para ellos se consideraron las bibliotecas no equipadas de la primera encuesta y que serían encuestadas en la segunda encuesta. Con esta información se puede evaluar posibles cambios en el funcionamiento de las bibliotecas, que pudieran ser atribuibles al equipamiento de las mismas.

⁴⁸ De esta forma, se incluyen en esta segunda muestra bibliotecas equipadas incluidas en la primera encuesta.

⁴⁹ Con la finalidad de identificar efectos a mediano y largo plazo.

Encuesta	Esquema de Muestreo	No.de bibliotecas	Estratos	Dominios	No. de Bibliotecas Encuestadas
Primera	Estratificado ⁵⁰ , Polietápico ⁵¹ y con selección aleatoria	1150	Según condición de equipamiento y número de computadoras asignadas	bibliotecarios, encargados de módulo, usuarios y no usuarios de las bibliotecas, y autoridades municipales	185
Segunda		1150	Grado de Marginación	bibliotecarios, encargados de módulo, usuarios y no usuarios de las bibliotecas, autoridades municipales, maestros y líderes comunitarios	185
Tercera		1238			304

Fuente: "Acceso Tecnológico: Una reinterpretación de la biblioteca pública mexicana". CONACULTA.

Como se puede observar en el cuadro anterior, para cada una de las encuestas se tomó una muestra probabilística con características similares debido a que tuvieron el mismo esquema de muestreo. Sin embargo, debido al aprendizaje obtenido en la experiencia de cada una de las encuestas se realizaron cambios en algunos aspectos de su diseño. Por ejemplo: agregar nuevos actores en el dominio (maestros y líderes comunitarios) al reconocer su importancia en la dinámica del programa.

⁵⁰ El muestreo estratificado consiste en la división previa de la población de estudio en grupos o clases que se suponen homogéneos respecto a la característica a estudiar.

⁵¹ Lo polietápico se refiere a las diferentes etapas de muestreo que se siguen para configurar la muestra.

iii. Principales resultados del proyecto de evaluación

De acuerdo con la primera evaluación⁵² al PASDBP, se encontraron los siguientes resultados (Santos, de Gortari, Angulo, 2006):

- La llegada de computadoras a las bibliotecas Públicas no representó, en la mayoría de los casos, el primer contacto con las TIC para los usuarios. Esto quizá, fue contrario a lo que se esperaba ya que las condiciones eran de alta marginación en la mayoría de los casos.
- A pesar de la existencia de la brecha digital en México, en la mayor parte de las comunidades se encontró oferta de lugares de acceso a las computadoras, la mayoría de ellas privadas en la que los cibercafés juegan un papel determinante.
- Que el PASDBP se distingue de otros proyectos gubernamentales⁵³ y de la sociedad civil⁵⁴ por dos aspectos fundamentales: la gratuidad en el uso de las computadoras y los esfuerzos realizados para formar capacidades tecnológicas en sus usuarios⁵⁵.
- Aunque se hicieron esfuerzos para que todos los bibliotecarios completaran la capacitación del Programa, su formación es un proceso que no fue concluido en muchos casos.
- Un grave problema fue la falta de conectividad, que se encontró afectaba a poco menos de la mitad de las bibliotecas equipadas. La falta de internet deja fuera una gran cantidad de recursos informativos, sociocomunicativos y recreativos; que contribuirían a transformar el papel de las bibliotecas con la llegada del MSD.
- Uno de los resultados más importantes que se obtuvieron en este proyecto es encontrar que el PASDBP representa la única posibilidad de acceso gratuito a las TIC para muchos habitantes de la República Mexicana. Esto da cuenta de la importancia del programa, que llega a comunidades rurales y marginadas, donde en ocasiones tampoco se cuenta con ninguna otra oferta de acceso a Internet⁵⁶.
- El perfil de usuarios y de las bibliotecas ha variado muy poco desde la instalación de los MSD; y el uso que se les da a las tecnologías sigue siendo primordialmente escolar.
- La instalación de los MSD permite que las bibliotecas sean más visibles en los lugares donde están localizadas, esto se refleja con el aumento en el flujo de usuarios y en el tiempo de uso de las instalaciones.

En conclusión, el primer capítulo ha aportado los antecedentes para apuntar y comprender los aspectos que parecen ser los de mayor importancia en el desempeño del Programa de Acceso

⁵² La primera evaluación, realizada por el IIS-UNAM, desde su definición se contemplaron solamente dos de las tres fases de equipamiento, debido a que se obedeció a los plazos que eran metodológicamente abordables, en función de los tiempos definidos para dicha evaluación.

⁵³ Dentro de estos proyectos destacan las Plazas Comunitarias del Instituto Nacional de Educación para Adultos (INEA) y los Centros Comunitarios Digitales instalados directamente por la Secretaría de Comunicaciones y Transportes, la cual es la entidad gubernamental responsable de e-México.

⁵⁴ Estos proyectos se refieren primordialmente a los Cibercafés.

⁵⁵ Dentro de los usuarios encontramos a los bibliotecarios y encargados de módulo, que no solamente son los asesores, sino que son los primeros en utilizar el equipo de cómputo.

⁵⁶ Uno de los factores que han convertido esas comunidades rurales, en lugares sin otra oferta tecnológica es su baja densidad de población.

a Servicios Digitales en Bibliotecas Públicas. Por una parte, se ha descrito la concepción del programa desde su diseño en la preocupación del combate a la creciente brecha digital, como parte de un proyecto a escala mundial (*Global Libraries*), hasta los problemas que se han presentado en la operación del programa debido a la complejidad del entorno social en que es implantado.

Ahora que se tiene noción de cómo fue planteado el programa y los resultados que se han obtenido a partir de la primera evaluación del IIS-UNAM, se puede hacer un bosquejo con cuatro elementos que parecen representarlo: el módulo como recinto físico del programa, los actores a cargo de su operación, el contexto en que es instalado el MSD y, por último, la visibilidad que tiene en la localidad.

CAPÍTULO 2. Construcción y análisis estadístico de los factores de desempeño PASDBP

Los programas sociales son la traducción operacional de las políticas sociales. Dicha operación tiene diversas etapas anteriores a ella, en las que se plantean estrategias que conlleven al logro de sus metas sociales. Y precisamente la operación de estos programas tiene influjo decisivo en su impacto, debido a que a menudo se producen situaciones que generan efectos negativos o inesperados, incluso en aquellos programas bien concebidos. Por ejemplo, un criterio de selección de las bibliotecas para el equipamiento por parte del PASDBP era adecuar un espacio físico especial para la llegada de las computadoras, con la finalidad de que se garantizara el buen funcionamiento del equipo; esto dio como resultado que muchas poblaciones marginadas no fueran seleccionadas para el equipamiento debido a que no se satisfacía este requisito por falta de recursos económicos de la localidad.

De lo anterior se puede advertir cómo obstáculos inesperados afectan el cumplimiento de los objetivos de los programas. Por otra parte, frecuentemente se desconoce el desempeño de los programas sociales, es decir, no se sabe si sirven o no, mucho menos está cuantificado en qué medida. Para conocer el desempeño del programa, primero se tiene que puntualizar qué es lo que se debe medir y cómo es que se hará.

Con frecuencia se evalúa mediante la construcción de una serie de índices e indicadores donde se resumen los aspectos fundamentales para su entendimiento como concepto en términos cuantitativos. Un ejemplo de estos indicadores sería el tiempo de experiencia de un programa, que puede ser medido en días, meses o años. O bien, la medida de satisfacción de los usuarios con respecto a los servicios ofrecidos, mediante valores en una escala definida a partir de conceptos teóricos.

En el presente capítulo y en el siguiente se expone el análisis estadístico de los factores que se considera determinan el desempeño del PASDBP. En específico, este capítulo marca el inicio de la metodología aplicada a la construcción de índices e indicadores y su posterior análisis univariado y bivariado, que aborda las variables desde una perspectiva más simple antes del modelo de regresión ordinal multivariado.

A partir de lo introducido en el primer capítulo se puede entender el contexto en el cual se desarrolla el programa y los elementos que lo caracterizaron, así como los principales resultados obtenidos de los ejercicios de evaluación anteriores. Por lo tanto se tiene la información necesaria para abordar la problemática de medir el desempeño del PASDBP, así como el conjunto de factores que pudieran estar influyéndolo. De acuerdo con la selección de los factores que se mencionaron en el primer capítulo, las variables que se consideran para realizar el análisis estadístico del desempeño del programa se agrupan en las siguientes dimensiones: las características de localidad, características del MSD, características del encargado de módulo y la difusión del programa.

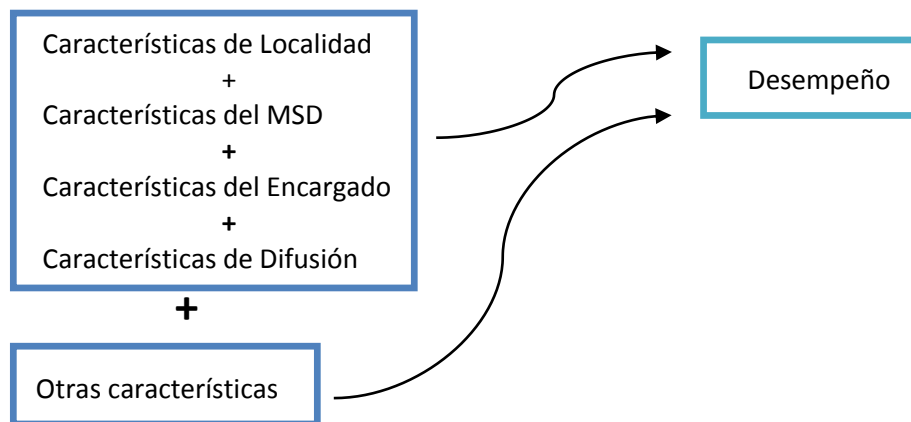
Por lo tanto, el modelo general que se plantea para el análisis estadístico se aborda desde una perspectiva causa-efecto, es decir, se entiende el desempeño del PASDBP como el resultado de múltiples factores.

Un modelo causal⁵⁷ puede ser definido como la expresión matemática que explica la causa del comportamiento de la variable dependiente relacionándola con un grupo de variables independientes. Es un modelo donde se puede estimar el resultado de nuestra variable dependiente, también llamada variable de respuesta, a partir de un conjunto de las variables independientes, llamadas también variables explicativas o predictivas.

En otras palabras, un suceso o fenómeno puede ser explicado como consecuencia de diversos factores. En donde algunos factores son más influyentes que otros en el resultado, es decir, unos tienen un efecto más fuerte que otros. A pesar de esto, en conjunto todas las variables lo modelan. Por otra parte, es importante precisar que al realizarse la modelación de un fenómeno, es imposible trabajar con todos los factores que intervienen en él, debido a la complejidad que significa la obtención de esta información. Sin embargo, que no sea posible incluir la totalidad de los factores como variables en un análisis, no implica que no existan o que no influyan en el fenómeno que se intenta modelar.

Por lo tanto, para el caso particular que ocupa esta tesis la variable dependiente a modelar es el Desempeño del PASDBP y las variables independientes son aquellas que conforman las Características de la Localidad, las Características del MSD, las Características del Encargado del Módulo y la Difusión por parte del Encargado.

MODELO CAUSAL



⁵⁷ Históricamente, los modelos causales tienen su origen en los trabajos de Wright (1934). Esta metodología utilizada por primera vez en biología fue importada a las CC. Sociales y Humanas por Simom (1957) y Blalock (1961, 1964, 1969) y popularizada por Ducan (1966) y Jöreskog y Sörbom (1978). Su aplicación permite a los investigadores no sólo validar un modelo causal basándose en los datos de correlación, sino también evaluar si los conceptos hipotetizados en el modelo son congruentes con las variables medidas en el modelo. Para una información más específica sobre el tema, consultar las obras de Asher (1989), Bisquerra (1989), Fuentes (1986), Heise (1975).

Como se visualiza en el diagrama anterior, no es posible incorporar al modelo todas las variables que influyen en el desempeño del PASDBP. Y a su vez las variables se pueden clasificar en diferentes tipos: por su grado de abstracción, naturaleza, etc.

Según su grado de abstracción se puede concebir la existencia de variables teóricas y empíricas. Aquellas variables teóricas no pueden ser medidas con precisión, es decir, sólo se puede tener una aproximación, o no son observables como por ejemplo: percepción del módulo por parte la población de la localidad donde es instalado el MSD. Mientras que las variables empíricas son aquellas variables que sí pueden ser medidas, por ejemplo: tiempo transcurrido desde la instalación del MSD.

Por otra parte, la elección de las variables para ser incluidas en un modelo depende de la utilidad y disponibilidad que se tenga de ellas. Respecto a la disponibilidad de información, se pueden clasificar como: aquellas variables de las que se cuenta con la información y de las que no se ha realizado su medición o no se tiene la información de su medición. Ejemplos de la segunda serían: número de ofertas tecnológicas en la localidad, cercanía con la cabecera municipal, entre otras.

A continuación se define conceptual y empíricamente la manera en que serán consideradas la variable dependiente y el conjunto de variables independientes de cada dimensión.

a) Definición del Desempeño del PASDBP y de los factores que pudieran estarlo influyendo

Después de comprender la complejidad de la realidad que existe en la relación de la variable dependiente con todas las variables que lo influyen. A continuación se definen las variables que utilizaremos en el análisis estadístico de los factores determinantes del desempeño del PASDBP. Esto es decisivo, ya que al tener una visión más específica de lo que se desea obtener y de lo que se necesita para realizar el análisis estadístico, se podrá pasar de los conceptos concebidos teóricamente a los conceptos empíricos.

A continuación se definirá lo que se entiende como la variable de respuesta “Desempeño del PASDBP” de acuerdo con un análisis teórico y apoyándose en resultados de evaluaciones anteriores al PASDBP previamente expuestos, se realizará una propuesta de medición con los datos disponibles. Posteriormente, y siguiendo la misma metodología, se determina cuáles son las variables predictivas que se designaron en el apartado anterior, con su respectiva propuesta de medición.

i. Definición de la variable dependiente

Para interpretar correctamente la variable dependiente o de respuesta, lo primero que se debe hacer, es puntualizar a qué se desea referir con “desempeño”. A continuación se presentan algunas de sus definiciones:

- La realización de las funciones propias de un cargo o trabajo⁵⁸.
- Habilidad en la realización de una actividad⁵⁹.
- Es la manera como alguien o algo trabaja, juzgado por su efectividad⁶⁰.

El concepto de desempeño ha sido tomado del inglés *performance* o de *perform*. Aunque admite también la traducción como rendimiento, será importante conocer que su alcance original tiene que ver directamente con el logro de objetivos (o tareas asignadas).

Ahora bien, al referirse al desempeño de un programa se encuentra que esencialmente es hacer una medición del cumplimiento de los objetivos de acuerdo a su estrategia, recursos y alcances.

Para utilidad de esta tesis se define de la siguiente forma “Desempeño del Programa de Acceso a Servicios Digitales en Bibliotecas Públicas” como: El cumplimiento de las funciones propias del Módulo de Servicios Digitales, es decir, prestar los servicios necesarios para acercar las TIC a las comunidades; contar con un personal adecuadamente capacitado para el aprovechamiento de la oferta tecnológica y lograr mayor asistencia por parte de los habitantes de la comunidad para aprovechar los dos anteriores.

Desglosando los elementos clave de la definición anterior correspondiente a la variable respuesta se puede tener una interpretación más detallada del “Desempeño del PASDBP”. Que al mismo tiempo delimita los indicadores por los que está compuesto y que más adelante servirán como estructura en el ejercicio de medición de la variable dependiente. En este sentido, se distinguen tres dimensiones del “Desempeño del PASDBP”:

- 1) Servicios
- 2) Habilidades tecnológicas de los encargados, y
- 3) Asistencia

La dimensión de servicios que ofrecen las bibliotecas equipadas, presenta el panorama de las nuevas posibilidades que han abierto estos Módulos de Servicios Digitales (MSD) a las comunidades donde fueron instalados, de la misma forma que complementan los que anteriormente ofrecía la Biblioteca. Al tener una mayor cantidad de servicios disponibles, se acondiciona una biblioteca equipada con mejores herramientas para poder acceder a nuevos usuarios, mejorar las habilidades tecnológicas de los usuarios e incluso del personal.

⁵⁸ Definición del Diccionario de la Lengua Española.

⁵⁹ «Fuentes electrónicas» <http://es.thefreedictionary.com/desempe%C3%B1o> (Consulta 20/04/2015)

⁶⁰ «Fuentes electrónicas» http://www.mp.peru-v.com/desempeno/ed_principal.htm (Consulta 20/04/2015)

Pese a que el PASDBP surge con el objetivo de proporcionar el servicio gratuito de acceso a computadoras e internet, las evaluaciones anteriores al PASDBP (por parte del IIS-UNAM) muestran que existe un grave problema en la conectividad. Por lo que parece conveniente medir dentro del indicador de servicios la conectividad y algunos otros elementos del equipo de cómputo como paquetería, ya que influyen de manera directa en las actividades del MSD. Y justamente el módulo como espacio físico, es considerado un servicio más que brinda la biblioteca. Si las instalaciones se encuentran en buen estado, otros servicios tecnológicos además de las computadoras e internet pueden contar con espacio adecuado que permita un buen funcionamiento de los equipos y al mismo tiempo genera interés en la comunidad para aprovechar todos los recursos que se ofrecen.

También existe otro tipo de servicios que se brindan en las bibliotecas equipadas y que dependen de actores más que de herramientas tecnológicas. Con ello se refiere a la asesoría que proporcionan los encargados de módulo y a las clases que se imparten para capacitar o actualizar a los usuarios y/o personal de la biblioteca.

En conclusión los servicios que consideraremos se refieren a: las herramientas tecnológicas, la asesoría y/o capacitación y a la infraestructura.

Las habilidades tecnológicas de los encargados es otra dimensión importante para medir el desempeño del PASDBP, ya que son los encargados de módulo el medio responsable de fomentar su uso adecuado y apropiación de las TIC del Módulo a través de sus habilidades tecnológicas, que puede tener como consecuencia una mejor atención debido a una mejor capacitación. Esta última puede ser resultado de una buena preparación académica, de aprendizaje en cursos por parte de la DGB o particulares, de la experiencia en el mismo MSD, etc. pero que, independiente a su origen, es significativa porque posibilita al encargado a ofrecer un mejor servicio y por lo tanto tener mejores resultados en el programa.

Como se evidenció en los resultados de las evaluaciones anteriores, la llegada de las computadoras no representó, en la mayoría de los casos, el primer acercamiento de los usuarios a las TIC debido a que existía una considerable oferta de lugares de acceso público (cibercafés). Esto nos indica dos cosas, la primera es que en el MSD, a diferencia de otros lugares de acceso (cibercafés), hay un actor cuya función específica es asesorar a los usuarios, y la segunda que se puede suponer que muchos de los usuarios que ocupan las computadoras del módulo se encuentran familiarizados con el equipo, al menos para sus funciones más básicas.

Entonces se espera que el requerimiento de la asesoría de los encargados de módulo por parte de los usuarios exija un buen nivel de las habilidades tecnológicas de los encargados para un mejor desempeño. Sin ellas, el aprovechamiento de las herramientas tecnológicas corre el riesgo de quedar estancado, sin importar el interés por parte de los usuarios y la buena actitud de los encargados. Es decir, a pesar de que haya buenas intenciones por parte de los encargados de módulo, no se puede prescindir de conocimiento sobre programas que ellos deben de tener para cumplir con los requerimientos en las metas del programa, que finalmente se reflejan en un buen desempeño.

Otra dimensión es la asistencia de usuarios, que frecuentemente se asocia a la visibilidad del programa en la localidad donde es instalado, y que se considera un indicador esencial en la medición del desempeño porque involucra directamente al actor usuario.

Para poder realizar la medición y delimitar la interpretación del indicador de asistencia, es conveniente recordar que los objetivos propuestos en el programa hacen referencia a mejoras económicas y sociales, por lo que si se están utilizando sólo por usuarios tradicionales (escolares) difícilmente podrán ser alcanzados estos objetivos, pues se está supliendo únicamente una carencia escolar.

Por lo que conocer qué tipo de usuarios asisten al módulo para aprovechar los recursos que éste provee, exhibe si el tipo de usuarios al que fue dirigido el programa es al que está llegando. Por ejemplo, en la evaluación del IIS-UNAM se encontró que el perfil de usuarios de las bibliotecas ha variado muy poco desde la instalación de los MSD, el cual es mayoritariamente de personas en edad escolar.

De igual manera, conocer si el nivel de asistencia que se tiene es el adecuado o no para los recursos económicos, tecnológicos y humanos disponibles en el módulo, está relacionado con los objetivos del PASDBP, ya que la asistencia puede indicar, además de eficiencia, el interés por parte de la comunidad al programa (e indirectamente utilidad⁶¹ en la comunidad).

Entonces, se debe ser cuidadoso con la perspectiva que se desee medir de asistencia porque de ello dependerá la construcción de su indicador y su interpretación como medida en el desempeño del PASDBP. Por ejemplo, si hay aumento en la asistencia al módulo implicaría una mayor visibilidad del programa en la comunidad; pero si al mismo tiempo la demanda de usuarios al módulo supera la capacidad que tiene el MSD, esto derivaría en una reducción del tiempo del que se puede disponer de las computadoras y de atención por parte del encargado lo que conlleva a un bajo desempeño.

ii. Definición de las variables independientes

A continuación se definirán las variables independientes que se considera pueden influir en los niveles de desempeño del programa. Se tomará como base la información generada en las evaluaciones realizadas por el IIS-UNAM, del propio diseño del PASDBP y del contexto en el que se desarrolla el programa, a partir de lo cual podemos identificar un conjunto de características en cuatro diferentes grupos que pudieran estar influyendo el desempeño del programa. Estos cuatro grupos o ámbitos son:

a) Características de la Localidad

⁶¹ Es decir, el interés de una comunidad hacia un programa está relacionado con la utilidad que pueda ofrecerle, pero no es determinante. Por ejemplo, si se implementa un programa social de apoyo a la agricultura en una zona donde su principal actividad sea la pesca, tal vez no haya tanto interés de la población debido a que les sería más útil un programa con apoyos dirigidos a la pesca. Pero no es determinante, porque puede ser que no exista una gran demanda a este programa social debido a otros factores como poca difusión del programa.

- b) Características del MSD
- c) Características del Encargado de Módulo
- d) Difusión

Enseguida se detalla cada uno de estos ámbitos.

El primer grupo que se abordará es el de las Características de la Localidad. Se define como un conjunto de factores que describen el entorno donde fue implementado el PASDBP. Las características de la localidad dan un panorama de las condiciones de las comunidades donde se encuentra ubicado el MSD. Por ejemplo, el acceso que se tiene a la tecnología puede manifestar el valor de los servicios ofrecidos por el programa, es decir, si es poca o nula la oferta de lugares de acceso a computadoras e internet en la localidad, el MSD cobra mayor importancia. Otros ejemplos de factores que pueden influir en el desempeño y forman parte de este grupo son: nivel de marginación, cercanía de la localidad a la cabecera municipal, densidad demográfica⁶², tipo de localidad, etc.

Las características de las localidades que se consideran en este análisis son:

Tipo de Localidad⁶³. Se supone como un factor que influye significativamente en el desempeño, ya que por lo general en localidades rurales se encuentra una menor oferta de acceso a lugares públicos de acceso a las TIC, y en consecuencia el MSD puede adquirir un valor más alto dentro de la comunidad en comparación con las localidades urbanas, donde se tiene una mayor oferta e incluso se tiene mayor posibilidad de acceder desde las casas a las TIC. Por otra parte, también se espera que los recursos económicos sean menores en las localidades rurales y en consecuencia el mantenimiento del módulo, la variedad de servicios y el pago a los encargados de módulo se vea en riesgo. Por lo anterior se considera que el programa tiene diferente desempeño en las localidades rurales comparado con las localidades urbanas.

Nivel de marginación de la localidad⁶⁴. Este factor delimita en mayor o menor grado, las posibilidades de acceso a la tecnología y las oportunidades de educación, cultura, entretenimiento,

⁶² La densidad demográfica es la medida expresada por la relación entre la población y el área ocupada, generalmente establecida en habitantes por kilómetro cuadrado. «Fuentes electrónicas» <http://cuentame.inegi.org.mx/poblacion/densidad.aspx?tema=P> (Consulta 20/04/2015) INEGI

⁶³ De acuerdo con el INEGI, una población se considera Rural si tiene menos de 2500 habitantes, mientras que la Urbana es aquella donde viven 2500 personas o más. «Fuentes electrónicas» http://cuentame.inegi.org.mx/poblacion/rur_urb.aspx?tema=P (Consulta 20/04/2015)

⁶⁴ El índice de marginación es una medida de déficit y de intensidad de las privaciones y carencias de la población en dimensiones relativas a las necesidades básicas establecidas como derechos constitucionales. El índice de marginación es el resultado de una estimación por componentes principales de cuatro dimensiones y nueve indicadores: educación (analfabetismo y población sin primaria completa); viviendas (ocupantes en viviendas sin agua entubada, sin drenaje ni servicio sanitario, con piso de tierra, sin energía eléctrica y hacinamiento); ingresos (población ocupada que gana hasta dos salarios mínimos); y distribución de la población (población en localidades con menos de 5 mil habitantes). «Fuentes electrónicas» http://www.conapo.gob.mx/es/CONAPO/Indices_de_Marginacion_Publicaciones (Consulta 20/04/2015)

comunicación, etc. de los habitantes de la comunidad. Si la localidad presenta un grado de marginación alto significa que dicha localidad posee características referentes a su ingreso, educación, vivienda y distribución poblacional que lo limitan y rezagan del resto de las localidades con menor grado de marginación. Se observa que estas características, con las que se realiza la clasificación del grado de marginación, se relacionan con el desarrollo económico y social primordialmente.

Por consiguiente la necesidad de tecnología y la asociación social de ésta, podría verse reflejada en el aprovechamiento de la nueva oferta tecnológica que representa el MSD, por ser un espacio de acceso gratuito a computadoras e internet. Sin embargo, también se podría relacionar con un alto nivel de marginación a un bajo presupuesto o, en el peor de los casos, al incumplimiento de los acuerdos establecidos entre el programa y las autoridades municipales para el mantenimiento del MSD y pago de otros servicios como internet y sueldo del encargado de módulo, vitales para un buen funcionamiento del módulo que derivaría en un buen desempeño.

El siguiente grupo se refiere a las Características del MSD, el cual se define como un conjunto de factores que expone las condiciones en que trabaja el módulo y elementos que lo conforman. Algunas de las características del módulo referentes a las condiciones son: mantenimiento del equipo de cómputo y de las instalaciones, características de las herramientas tecnológicas, etc. Otros factores son el tiempo de la biblioteca en el programa, la antigüedad de la biblioteca, financiamiento por parte de autoridades municipales y otras instituciones, entre otras.

Si bien el programa estableció requisitos como la adecuación de un lugar físico especial para la instalación de módulo y acuerdos con autoridades municipales de dar el apoyo necesario para mantener y continuar el programa, esta situación no se cumplió en muchos casos. Se encontró en los resultados de las evaluaciones realizadas por el IIS-UNAM, muchas bibliotecas en las que no se ha dado mantenimiento a la infraestructura o que no cuentan con las condiciones indispensables para su funcionamiento. Incluso en algunas bibliotecas las computadoras están fuera de servicio o no se albergan en el módulo.

Las características de los módulos que se consideran en el análisis son:

Experiencia/Tiempo en el programa⁶⁵. Un factor que se puede considerar fundamental en el desempeño de cualquier programa social es la experiencia del programa, refiriéndonos a la experiencia como un término que mide el tiempo de estar en el programa. En el caso del PASDBP se espera un mejor desempeño transcurrido algún lapso de tiempo debido a que pasando un tiempo, el módulo ya sea conocido por la comunidad, los encargados conozcan mejor el programa y los usuarios aprovechen mejor las herramientas tecnológicas y servicios que se ofrecen en el

Según CONAPO, hay 5 niveles de marginación: Muy Alto, Alto, Medio, Bajo y Muy Bajo; sin embargo para este trabajo se considera la misma síntesis utilizada la base de datos de las encuestas donde únicamente se clasifican en 3 niveles: Alto, Medio y Bajo.

⁶⁵ Esta medida se refiere al tiempo transcurrido desde la creación del módulo hasta la realización de la encuesta.

MSD. Por lo tanto, se supone que conforme pase el tiempo haya mayor optimización en servicios, técnicas de apoyo, difusión, interés por parte de la comunidad, etc. que den como resultado un mejor desempeño del programa.

Antigüedad de la biblioteca⁶⁶. Se piensa que el tiempo que lleve funcionando la biblioteca puede modificar el nivel de desempeño del programa al aprovechar a los usuarios asiduos a la biblioteca como primeros usuarios, pero también como promotores del módulo. Es decir, si una biblioteca presenta una mayor antigüedad pudiera tener un mayor arraigo y visibilidad por parte de los habitantes de la comunidad, en consecuencia podría atraer un mayor número de beneficiarios e influir en el desempeño del programa.

Tamaño de la biblioteca: Debido a que en su mayoría los módulos fueron instalados dentro de la biblioteca, el tamaño de éstas limita el espacio que se le da al módulo, al número máximo de computadoras que se le pueden asignar y el número de usuarios que podría atender. Si el tamaño del MSD es adecuado respecto al tamaño de la biblioteca, se puede tener un buen desempeño del programa ya que se podría atender a más usuarios de manera adecuada. Por otra parte, el tamaño de la biblioteca puede estar relacionado con la visibilidad del programa en la comunidad y causar efecto en el indicador de asistencia del desempeño.

Número de computadoras asignadas. Éste parece ser un factor que determina de manera importante el desempeño del PASDBP, al delimitar si son suficientes o no las computadoras con respecto a la demanda por parte de la localidad del servicios gratuito de computadoras e internet. Podría ser que se asignen correctamente el número de computadoras respecto al espacio físico que ocupa el módulo dentro de la biblioteca, pero al mismo tiempo sean insuficientes para todas las personas que desean aprovechar los servicios que se ofrecen.

De manera que si la necesidad de una comunidad sobrepasa la capacidad de la que puede encargarse el módulo para prestar todo tipo de servicios (herramientas tecnológicas, atención de los encargados, etc.), entonces tendrá un bajo desempeño.

Sostenibilidad: Este factor pudiera influir de manera importante el desempeño del PASDBP porque indica si existe un panorama futuro para continuar la operación del programa o si aún no se concibe, lo cual reduce las probabilidades de mantener funcionando el MSD y de seguir brindando a la comunidad el servicio. Por otro lado, al tener un plan para continuar el programa, se tiene conciencia de las acciones que se deben realizar o las medidas que se deben tener para continuarlo y a la vez se obtenga un mejor desempeño.

Financiamiento. Un factor que puede influir en el desempeño del programa es el tipo de financiamiento con el que cuenta el módulo por parte de las autoridades de la localidad (municipales o delegacionales, según sea el caso) y la certeza del mismo. Se espera afecte principalmente el

⁶⁶ Esta medida se refiere al tiempo transcurrido desde el año de Inauguración de la biblioteca a la fecha de la realización de la encuesta, según encuestas nacionales.

indicador de servicios del desempeño del PASDBP porque de él depende el mantenimiento y pago de salarios de los encargados de módulo.

En cuanto a las Características del Encargado del Módulo, se define como un conjunto de variables que describen algunos rasgos particulares del responsable del módulo. El cual se encarga del apoyo, colaboración y dirección del módulo para que los recursos tecnológicos que otorga el programa sean aprovechados y asimilados en la mejor forma posible por los usuarios.

Es decir, sería deseable que el encargado del módulo tuviera conocimiento tanto del manejo básico de las computadoras como de algunos paquetes de software específicos. La información básica que nos proporciona este acercamiento es la edad del encargado, en cierta medida su condición socioeconómica, pero sobre todo la escolaridad, ya que se espera que tenga mayor conocimiento de estas herramientas; lo cual podría potencializar el uso de los servicios del módulo.

Conocer las características de la persona que labora en el módulo puede proveer información de aspectos demográficos para identificar las características de un personal adecuado para las actividades del MSD. De igual manera, la experiencia y capacitación que los encargados van adquiriendo en la operación diaria del MSD, puede influir de manera importante en el funcionamiento del módulo. Ejemplo de ellos son el tiempo de trabajar en la biblioteca y/o en el módulo, asistencia a los cursos de capacitación que brindó la DGB, rotación del personal en la biblioteca y MSD, entre otros.

Las características de los encargados de módulo que se consideran en el análisis son:

Tipo de encargado del módulo. Un aspecto que podría influir de manera importante en el funcionamiento del módulo, es si el responsable divide su tiempo entre la biblioteca y el MSD o si se dedica exclusivamente a este último.

Es decir, si se cuenta con una persona especializada en las actividades correspondientes al módulo se puede prestar un mejor servicio, pues dispondría de mayor tiempo para la atención de los usuarios, para la revisión y mantenimiento del equipo de cómputo e inclusive para mejorar las habilidades tecnológicas de los usuarios. Si por el contrario, el MSD cuenta con un encargado-bibliotecario, este personaje tendría que realizar un mayor número de actividades y disminuiría la atención a los asuntos propios del módulo. Consecuentemente lo anterior se reflejaría en un menor desempeño, a que si tuviera un encargado exclusivo del MSD.

Tiempo de trabajar (encargado del módulo). Este factor de tiempo de trabajar en el MSD/Biblioteca puede conllevar una optimización de las habilidades tecnológicas y mayor eficacia para manejar el módulo. De igual manera, también se puede propiciar mayor acercamiento con los usuarios, y se genere un ambiente de confianza donde se unen esfuerzos de parte de la comunidad y del personal para mantener funcionando el módulo y, en el mejor de los casos, dotarlo de una mejor infraestructura o recursos.

Escolaridad del encargado del módulo: Se espera que si el encargado de módulo tiene un alto grado de escolaridad esté mejor capacitado para desarrollarse en el trabajo del MSD, ya que contará con ventajas sobre las personas que no estuvieran apropiadamente preparadas para brindar una mejor atención a los usuarios. Por ejemplo, la asesoría en paquetería, la actualización y el mantenimiento del equipo de cómputo requieren de conocimientos técnicos en computación. Sin embargo, es importante precisar que igualmente la preparación académica debe estar relacionada con las TIC, computación, informática u otras relacionadas.

Capacitación recibida (DGB): Otro factor que se cree contribuye en el desempeño del programa es la capacitación que recibe el encargado de módulo por la DGB, como un complemento a las habilidades que se supone posee debido a su formación académica.

Se esperaría que la asistencia a los cursos de la DGB, los cuales fueron diseñados especialmente para el PASDBP, proporcione al encargado del módulo aditamentos para lograr un mejor cumplimiento de sus labores dentro del MSD, y por consiguiente afecte de forma positiva el desempeño del programa.

Finalmente, el último grupo hace referencia a la Difusión, que se define como el conjunto de factores que caracterizan la promoción del módulo en la localidad. Si bien las bibliotecas son espacios públicos que promueven la cultura, su interacción en la comunidad rara vez es distinta al cumplimiento de tareas de estudiantes de educación básica por lo que difícilmente la visita a la biblioteca representa una oportunidad de reconocimiento del MSD por parte de personas en edad no escolar.

Uno de los principios del PASDBP establece que el uso de las computadoras está abierto a todo tipo de público y para todo tipo de uso (educación, capacitación, comunicación, entretenimiento, etc.). Sin embargo, en la evaluación del IIS-UNAM se encontró que el uso del MSD sigue siendo primordialmente escolar, es decir, el mismo tipo de usuarios de la biblioteca se comparte con el del módulo. Esto puede deberse al desconocimiento de la comunidad de la nueva oferta tecnológica, así como de sus funciones. Lo cual pareciera que con una adecuada promoción genere el interés de nuevos usuarios del MSD.

La característica de difusión que se considera en el análisis son:

Difusión del módulo por parte del encargado: Se supone como un factor que influye significativamente en el desempeño del programa porque por medio de la continua promoción del módulo se puede acceder tanto a los usuarios tradicionales⁶⁷ de la biblioteca como a nuevos, dando a conocer la oferta tecnológica a la comunidad. Con ello se refiere tanto al acceso gratuito de computadoras con internet como a cualquier actividad que se realice dentro de las instalaciones del módulo, ya sea que utilicen el internet o únicamente paquetería del equipo

⁶⁷ Se refiere a las personas en edad escolar, principalmente jóvenes y niños.

Parece lógico que sea el actor a cargo del módulo el que promoció el MSD por ser él quien conoce mejor los servicios que ofrece y quien adecua las actividades que se realizan en él, según las características y demanda de los usuarios. Se entiende que la difusión es importante ya que no basta ofrecer una variedad de servicios dentro del MSD si no son conocidos y aprovechados por la localidad. Por consiguiente, este factor influye directamente en el indicador de asistencia que refleja sus efectos en el desempeño.

Después de definir cada uno de los factores que utilizaremos en el análisis estadístico, se puede iniciar el análisis empírico donde se procederá a la operacionalización de cada uno de los conceptos teóricos. La información que se utilizará será generada por las tres encuestas realizadas por el IIS-UNAM, y se establecerá como propuesta de medición la construcción de indicadores para cada uno de los factores, mediante la selección de variables (preguntas) en los cuestionarios.

b) Base de datos para el análisis

Gracias a la información levantada en las tres encuestas nacionales, como parte de los ejercicios de evaluación realizados por el equipo del Instituto de Investigaciones Sociales (UNAM), se tiene acceso a la información pertinente para realizar este análisis que tiene por objetivo determinar los factores influyentes en el desempeño del PASDBP.

A partir de la información de las tres encuestas se conformó una sola base de datos con información de todas las bibliotecas equipadas, que hubieran sido incluidas al menos en una de las muestras de las encuestas.

Para estructurar la base de datos se tomaron en cuenta los siguientes aspectos:

1. Las encuestas se desarrollaron en tres momentos diferentes.
2. Se aplicaron los mismos cuestionarios. Por lo tanto, para todos los módulos en alguna de las encuestas se tiene la misma información.
3. Dada la metodología de evaluación, hay módulos que fueron encuestados en más de una ocasión.
4. En la primera etapa de la evaluación del IIS-UNAM se consideraron bibliotecas equipadas y no equipadas.

Por lo tanto, se construyó una base de datos que tuviera toda la información acerca del PASDBP y de los factores propuestos como influyentes del desempeño. Es importante resaltar que esta nueva base contiene información de los MSD a partir de la visión del encargado del módulo y de los usuarios, es decir, únicamente se utilizaron los cuestionarios de cada encuesta correspondientes a estos actores.

A continuación se presenta la metodología de la construcción de la base de datos para el análisis estadístico.

i. Bibliotecas de estudio para el análisis

Las bibliotecas que se consideran en este análisis son las bibliotecas equipadas que hayan formado parte de alguna de las muestras de la encuestas de la evaluación realizada por el IIS-UNAM. Para la formación de la base de datos primeramente se consideran todas aquellas bibliotecas equipadas encuestadas en alguna de las etapas de las evaluaciones del IIS-UNAM. Dado el diseño de evaluación y la conformación de las bases de datos, algunas bibliotecas fueron encuestadas en más de una ocasión⁶⁸. En total, en las tres encuestas se visitaron 472 bibliotecas equipadas, de las cuales 86 fueron encuestadas dos veces y 9 hasta en tres ocasiones.

Tipo de Biblioteca	Número de Bibliotecas
Participaron en una encuesta	377
Participaron en dos encuestas	86
Participaron en tres encuestas	9
Total	472

De las bibliotecas que se encuestaron una sola vez no hubo problema, porque se tiene la información de la única encuesta y esa es la que se utiliza. El problema surge para las bibliotecas encuestadas en más de una ocasión (95 bibliotecas), ya que se tenía que decidir cuál información sería más adecuado considerar: si la de la primera, segunda o incluso hasta la tercera encuesta. En otras palabras, se tuvo que determinar cuál de las encuestas en las que participó se debería tomar.

Debido a que se tienen 377 con una sola encuesta, se decide que la mejor opción sería elegir la información de la primera vez que todas fueron encuestadas, ya que además de poder unificar todas las bibliotecas en un grupo: “Encuestadas por primera vez”, se pueden controlar sesgos que pudieran ser atribuibles al conocimiento del cuestionario.

Por lo tanto, la información del conjunto de bibliotecas consideradas para el análisis se generó con información de la primera aplicación del cuestionario en los módulos de las 472 bibliotecas de las tres encuestas, lo que implicó un trabajo de unir la información de sus bases de datos.

ii. Preguntas de los cuestionarios para el análisis

A pesar de que los cuestionarios de las tres encuestas son comparables en grandes temas, existen pequeñas diferencias en cuanto al orden de las preguntas, opciones de respuestas y en algunas

⁶⁸ Como parte de la estrategia de evaluación se propuso la generación de información de seguimiento de algunas bibliotecas, con ellas se construyeron las series de datos de seguimiento para la primera y segunda encuesta. Por su parte la tercera encuesta seleccionó de manera aleatoria 300 bibliotecas (24.2% del total), de las cuáles se encontraron bibliotecas anteriormente encuestadas por lo que se continuo con la estrategia de seguimiento.

incluso en redacción. Esta situación dificultó que las bases de datos con la información de las encuestas fueran unidas de manera directa, y por lo tanto, se tuvo que realizar un análisis de cada uno de los cuestionarios utilizados en las encuestas y estructurar las bases para que pudieran ser unidas.

De manera general el procedimiento que se siguió para hacer la base de datos contempló los siguientes pasos.

1. Revisión de cada uno de los cuestionarios para identificar diferencias en preguntas u orden. Se identificaron las preguntas con seguimiento, es decir, se consideraron aquéllas que estuvieran presentes en todas las evaluaciones (las tres encuestas realizadas por el IIS-UNAM).
2. Recodificación de variables para que fueran comparables⁶⁹. Si las preguntas seleccionadas en el primer paso presentaban disparidad en redacción u opciones de respuesta, se unificaban.
3. Generar la estructura igual de las bases de datos, es decir, generar una base de datos con las variables que se utilizaron en cada una de las encuestas.
4. Unión de las bases. Finalmente se unieron las tres bases de datos.

Como resultado, se obtiene una base de datos específicamente creada para desarrollar el análisis estadístico de los determinantes del desempeño del PASDBP.

c) Construcción de la variable dependiente y de las variables independientes

Debido a que el desempeño del PASDBP es un concepto complejo que no es posible medir de manera directa mediante una única variable, se consideró necesario construir un índice para su adecuada medición y utilizar indicadores para las variables dependientes.

A continuación se describe un procedimiento por medio del cual se puede construir un índice social.

i. Índices e Indicadores Sociales

La definición de los indicadores depende de los aspectos metodológicos y teóricos que se planteen de éstos, y obedece, más exactamente, al objetivo mismo del indicador. Se puede identificar que los indicadores sociales tienen la finalidad operativa de brindar información concreta sobre aspectos determinantes de los procesos sociales, para planificar políticas públicas en torno a una problemática específica. (Foessa, 1967, citado por Rodríguez, 2000)

⁶⁹ A pesar de que los cuestionarios se diseñaron para poder realizar un análisis longitudinal, se realizaron algunos cambios para ajustarla a lo que parecía mejor de acuerdo con la experiencia de cada encuesta.

Una de las definiciones más utilizadas por diferentes organismos y autores es la que Bauer dio en 1966: “Los indicadores sociales (...) son estadísticas, serie estadística o cualquier forma de indicación que nos facilita estudiar dónde estamos y hacia dónde nos dirigimos con respecto a determinados objetivos y metas, así como evaluar programas específicos y determinar su impacto”⁷⁰

En general los indicadores sociales son instrumentos contruidos a partir de un conjunto de valores numéricos o de categorías ordinales o nominales que sintetizan aspectos importante de un fenómeno con propósitos analíticos (Cecchini, 2005). Y dentro de un enfoque técnico descriptivo se entiende por indicador social una medida estadística directa y válida que permite conocer los cambios que ha tenido una problemática social a través del tiempo (Rodriguez, 2000)

Por su parte un índice puede definirse como: “(...) un número estadístico que intenta resumir la información proporcionada por uno o más indicadores de un concepto”⁷¹.

A primera vista pareciera que índice e indicador se refieren al mismo concepto pero no deben confundirse porque existen diferencias importantes entre ellos. La principal es que los indicadores forman parte de la estructura de un índice. Es decir, un índice está conformado por una serie de indicadores.

Se puede observar que en la gran mayoría de investigaciones, que de una u otra forma proponen la medición de fenómenos sociales, comparten el objetivo de llegar a definir un índice que resuma el tema sometido a examen. La construcción de estos índices sociales es elemental para evaluar, dar seguimiento y predecir tendencias de la situación de un país, un estado o una región en lo referente a su economía, sociedad, desarrollo humano, etc., así como para valorar el desempeño de alguna institución o programa encaminado a lograr las metas y objetivos.

Entonces, para poder abordar un procedimiento para la construcción de un índice, es necesario llevar a cabo el proceso de operacionalización, por medio del cual un hecho social deja de ser una mera observación y pasa a ser objeto de investigación en el que una idea general se transcribe a lenguaje científico. Este proceso consta de 4 fases (Lazarsfeld, P., 1985:36-40):

- Primera Fase: Representación literaria del concepto, en la primera fase se construye una imagen simple de lo que se quiere analizar. Una vez que el concepto ha sido seleccionado, el siguiente paso es exponer una definición teórica del mismo, es decir, explicar en los términos más simples los posibles significados del concepto.
- Segunda Fase: Especificación del concepto, aquí se identifican los “componentes”, “aspectos” o “dimensiones” de los aspectos sociales. Las dimensiones de un concepto son definidas como: “(...) los distintos aspectos en que puede ser considerado un concepto, representado así los componentes del concepto”⁷² No existen reglas teóricas

⁷⁰ Horn, Robert V. Statistical indicators for the economic and social sciences. Cambridge, University Press, Hong Kong, 1993, p. 147.

⁷¹ Díez j. 1967 Op. cit., pp.195.

⁷² Gónzalez Blasco, P., 1994 :282

para delimitar las dimensiones del concepto; porque éstas se establecen sometiendo al mismo a un análisis teórico, o bien empíricamente aplicando resultados de estudios previos⁷³.

- Tercera Fase: Elaboración de indicadores, en esta tercera fase se cuantifican las dimensiones conceptuales anteriormente descritas. Al tener que elegir un número de indicadores perdemos matices del concepto pero ganamos la posibilidad de medición. Sin embargo, la selección de los indicadores depende primordialmente de los propósitos y necesidades de la investigación o análisis; sin pasar por alto las posibilidades reales con las que contamos.
- Cuarta Fase: Formación de índices, con lo que se sistematizan los datos obtenidos en las etapas anteriores.

En el siguiente apartado se aplican las cuatro fases para la construcción de un índice de “Desempeño del PASDBP” que permita obtener la variable dependiente.

ii. Construcción del Índice “Desempeño del PASDBP”

Como se ha planteado anteriormente, lo que se desea explicar es el desempeño del programa mediante el análisis de aquellos factores que lo determinen en mayor medida.

Entonces, como parte de la metodología que se ha establecido en este capítulo, corresponde establecer una propuesta de medición que se adecue y aproxime a la definición de “Desempeño del PASDBP” precisada en el apartado de la definición de la variable dependiente, siguiendo las cuatro fases para la construcción de un índice..

La primera fase que se refiere a la “Representación literaria del concepto”, ya fue aplicada en el inciso “Definición de la variable dependiente” como: “ El desempeño del Programa de Acceso a Servicios Digitales en Bibliotecas Públicas es el cumplimiento de las funciones propias del Módulo de Servicios Digitales, es decir, prestar los servicios necesarios para acercar las TIC a las comunidades; contar con un personal adecuadamente capacitado para el aprovechamiento de la oferta tecnológica y lograr una asistencia por parte de los habitantes de la comunidad para aprovechar los dos anteriores”.

Continuando con la segunda fase de “Especificación del concepto”, se definieron las dimensiones del “Desempeño del PASDBP” como: servicios, habilidades tecnológicas de los encargados y asistencia. Cada una de estas dimensiones a su vez abarca diversos aspectos, que en algunos casos constituyen subdimensiones. Es decir, dependiendo de la necesidad de fragmentar el concepto de la dimensión en más subtemas para su mejor entendimiento y utilidad dentro del índice se pueden obtener varios niveles dentro de una misma dimensión.

A continuación se presentan los aspectos que conforman cada una de las dimensiones del desempeño del PASDBP

⁷³ Nosotros tomaremos como referencia los resultados obtenidos en la Evaluación realizada por el IIS-UNAM.

DIMENSIÓN I. SERVICIOS

Dentro de la dimensión de servicios podemos encontrar las siguientes subdimensiones.

Dimensión	Subdimensión	Indicadores
Servicios	Herramientas Tecnológicas	Internet
		Paquetería
	Asesoría y Capacitación	Asesoría
		Cursos
	Infraestructura	Otros servicios
	Condiciones Físicas del MSD	

En seguida se especifica cada uno de los aspectos que corresponden a la dimensión de servicios.

Herramientas Tecnológicas.

- **Internet.** Éste fue un aspecto básico para lograr las metas del programa, ya que significa una puerta a diversas posibilidades de educación, comunicación, cultura, capacitación, etc. El fundamento que plantea el Programa de Acceso a Servicios Digitales en Bibliotecas Públicas es proveer de internet y computadoras a las bibliotecas públicas, pero hay diversas complicaciones para la instalación y sustentabilidad del internet, por lo que se quiere evaluar algunos aspectos de él:
 - Proporción de computadoras conectadas a internet respecto a las asignadas por el programa al MSD.⁷⁴
 - Velocidad/Calidad de internet
- **Paquetería.** Se desea averiguar si los equipos de cómputo cuentan o no con paquetería. En los casos en los que no se cuenta con acceso a internet, la paquetería se vuelve el principal soporte para tratar de satisfacer las necesidades de los usuarios.

Asesoría y Capacitación.

- **Asesoría.** El hecho de que los MSD fueran acondicionados con la infraestructura, no conlleva su aprovechamiento. Por lo tanto, contar con la asesoría adecuada es un componente que diferencia al PASDBP de otros programas similares que brindan acceso a las TIC, ya que también considera la asimilación tecnológica. Por lo tanto, lo que se quiere medir es la calidad de la asesoría que reciben los usuarios por parte del Encargado de Módulo.
- **Cursos.** Con el fin de complementar la asesoría a los usuarios, se propuso impartir cursos para mejorar y/o crear las habilidades tecnológicas necesarias para que los usuarios se

⁷⁴ Es importante mencionar que se refiere a la proporción de computadoras conectadas a internet, respecto de las computadoras asignadas al MSD por parte del programa. Sin embargo, debe aclararse que en los cuestionarios no se encontró ninguna pregunta referente a cuantas computadoras otorgadas por el programa estaban dentro del MSD y conectadas a internet. Es decir, es decir que sin hacer diferencia del origen de las computadoras nos importa la conectividad que se tiene en proporción a las asignadas.

adueñaran de las TIC (es importante mencionar que estos cursos no son impartidos necesariamente por los encargados de módulo, ya que existe la posibilidad de contratación de alguna persona ajena al MSD para su colaboración en esta actividad).

Infraestructura

- **Otros Servicios.** Gracias a la adaptación de un espacio físico adecuado para la instalación de computadoras e internet dentro de la biblioteca, se tiene la oportunidad de complementar el MSD y mejorar sus resultados con la inclusión de otros servicios, como son: impresión de texto, acceso a programas en CD, quemado de discos, entre otros.
- **Condiciones físicas del MSD.** Por medio de una calificación que le asignen los usuarios en cada uno de los aspectos: mobiliario, espacio físico, iluminación, ruido y ventilación, etc. se podrá tener una aproximación a las condiciones físicas en que se encuentra el módulo. Esta perspectiva refleja el punto de vista del actor usuario que puede especificar si es adecuado o no, para las necesidades no sólo de equipo sino como espacio de trabajo, estudio, comunicación, capacitación, etc.

DIMENSIÓN II. HABILIDADES TECNOLÓGICAS DEL ENCARGADO

Dentro de la dimensión de habilidades tecnológicas del encargado se encuentra un solo indicador

Dimensión	Indicador
Habilidades Tecnológicas del Encargado	Conocimiento/experiencia en el uso de los programas

- **Conocimiento/ experiencia en el uso de los programas.** En este indicador se desea conocer qué programas saben utilizar los encargados de módulo. Al tener mejores conocimientos de los programas, se tiene un mejor intermediario de la tecnología y los usuarios.

DIMENSIÓN III. ASISTENCIA

Dentro de la dimensión de asistencia podemos encontrar los siguientes indicadores:

Dimensión	Indicador
Asistencia	Tipo de Usuarios
	Frecuencia

A continuación se especifican los dos indicadores que conforman la dimensión de asistencia.

- **Tipo de Usuarios.** Para determinar el tipo de usuarios se tomará como referencia la edad. Para poder simplificar la categorización se dividió en 3 grupos de edad. De esta forma se

sabe si la asistencia se da por parte de los usuarios habituales⁷⁵ o si hay diversidad en los usuarios del módulo. Lo que daría razón de un mayor alcance y acercamiento del programa a la comunidad donde se instala el MSD.

- **Frecuencia.** Se desea conocer el nivel de asistencia a la biblioteca y al MSD para observar los cambios que hubo desde la instalación del MSD, tanto en la biblioteca como en el módulo.⁷⁶

Ya que se han definido las dimensiones del PASDBP, se pasa a la metodología de cuantificación en la tercera fase de “Elaboración de indicadores”, donde se debe proponer la forma en que se deben cuantificar las dimensiones conceptuales anteriormente declaradas, es decir, determinar un procedimiento adecuado para realizar la medición de las dimensiones.

Uno de los problemas para la construcción de índices con diversas variables (indicadores) es qué peso darle a cada una en el índice. Una posibilidad es sumar directamente, lo cual supone que todas las variables tienen la misma importancia. Sin embargo, esto no sería correcto porque generalmente no es así, entonces un procedimiento común es la asignación de ponderaciones diferenciadas, es decir, asignar pesos dependiendo de la importancia que tenga cada una de las variables en el índice, y estos pesos se determinan principalmente con base en la teoría.

Procedimiento para el cálculo de ponderaciones

En este proceso de asignación de pesos el principal problema que se presenta en la cuantificación es cómo elegir o qué valor otorgar a las ponderaciones de las dimensiones o variables que participan en el índice a calcular. Su importancia radica en que de ello depende que el acercamiento que se persigue a través de la medición del índice se aproxime lo más posible a la realidad que se intenta explicar.

Por ello, el objetivo de los coeficientes de ponderación es dar diferentes pesos a reacciones, actitudes, acciones o situaciones, que de hecho, pueden diferenciarse de algún grado o forma.

En investigaciones de algún fenómeno o proceso social, la elección de los pesos de las ponderaciones suele ser difícil debido a la compleja realidad social que se intenta medir. No obstante, los coeficientes de ponderación deben cumplir los siguientes requisitos con el fin de facilitar e incrementar la exactitud de los índices elaborados:

⁷⁵ Se entiende por usuarios habituales a los “estudiantes”, ya que son estos los que concurren regularmente a la biblioteca para realizar trabajos y tareas escolares.

⁷⁶ Está información no se obtiene al comparar en dos momentos diferentes la asistencia de la biblioteca, sino que se tomará el dato del testimonio de los encargados del módulo con respecto al número de computadoras del cuestionario, obteniendo un promedio diario de usuarios por computadora para el A4 y para el cuestionario A2 se tomará la opinión con respecto a la frecuencia de su asistencia a partir de la creación del MSD.

1. Los coeficientes de ponderación deben representar lo más fidedignamente posible a la variable que pondera y, a las diferencias exactas de sus indicadores⁷⁷.
2. El coeficiente ha de ser lo más sencillo posible⁷⁸.
3. El coeficiente puede ser positivo y/o negativo.
4. Las frecuencias o atributos iguales han de ponderarse igual⁷⁹.

Una vez que se determinan los coeficientes de ponderación, cada coeficiente o peso se multiplica por la proporción obtenida de esa variable o dimensión.

Aunque para la construcción del Índice de desempeño del PASDBP existe sencillez en el cálculo matemático⁸⁰, la elección de los pesos de las variables y dimensiones representó una tarea complicada, debido a que la asignación de un peso a un concepto teórico, siempre deja dudas sobre si la ponderación está siendo afectada por un criterio personal o si existe abuso en cuestiones de índole teórica o matemática. De cometer este tipo de errores, podría comprometerse los resultados del análisis estadístico.

Después de realizar varios ejercicios con el fin de retroalimentar la toma de decisiones, se logra coordinar la sencillez de lo numérico y lo teórico. Y finalmente para la cuantificación de dimensiones del “Desempeño del Programa de Acceso a Servicios Digitales en Bibliotecas Públicas” se decidió otorgar los siguientes coeficientes de ponderación.

⁷⁷ En algunos casos es relativamente sencillo establecer las diferencias entre categorías, porque se pueden representar numéricamente. En otros casos, la relación se ve dificultada por el nivel de abstracción que contiene en sus categorías. Por lo que trabajar con conceptos sociales de mayor grado de dificultad o abstracción, el investigador debe recurrir a sus conocimientos teóricos de la variable estudiada, recurrir a su intuición o a su experiencia para tomar esta decisión sobre qué y cómo ponderar.

⁷⁸ Preferiblemente que sea un número entero y lo más pequeño posible.

⁷⁹ De este modo, las comparaciones entre los índices es mayor.

⁸⁰ Debido a que se propone un índice sumado ponderado.

Dimensiones		Subdimensiones e indicadores	
Servicios 50%	Herramientas Tecnológicas 50%	Internet 75%	Acceso 75%
			Velocidad 25%
		Paquetería 25%	
	Asesoría y Capacitación 25%	Asesoría 50%	
		Cursos 50%	
	Infraestructura 25%	Condiciones Físicas 50%	
Otros Servicios 50%			
Habilidades Tecnológicas de los Encargados 25%	Uso/Experiencia en programas 100%		
Asistencia 25%	Frecuencia 75%		
	Tipo de Usuarios 25%		

Los niveles se asignaron dependiendo el grado de importancia que tienen las variables o dimensiones por sobre otra(s) ⁸¹. A modo de ejemplo se explicará la elaboración de la subdimensión “Herramientas Tecnológicas”⁸².

De acuerdo con el cuadro anterior, donde se exhiben las ponderaciones del índice, la dimensión de Servicios (sombreada en gris) se conforma con las siguientes subdimensiones: “Herramientas tecnológicas”, “Asesoría y capacitación” e “Infraestructura”. A su vez la subdimensión de “Herramientas tecnológicas” se disgrega en los siguientes indicadores: “Internet” y “Paquetería”. La asignación de sus pesos utilizando los cuatro niveles se justifica de la siguiente forma.

El peso más importante cae sobre Internet (75%) y se le deja la proporción restante (25%) a paquetería, esto se fundamenta en que la base de este paquete tecnológico son las computadoras e internet. El internet es la herramienta tecnológica clave, sin ella, el aprovechamiento de las computadoras y su alcance como TIC estaría incompleto.

⁸¹ Consideramos únicamente cuatro valores para ponderar:

- a) 25% para ponderar como poco importante,
- b) 50% para ponderar como medianamente importante,
- c) 75% para ponderar como muy importante y
- d) 100% cuando sólo se tiene una sola variable.

⁸² La construcción de los otros indicadores puede checharse en el anexo “Construcción del índice de Desempeño del PASDBP”.

Por otra parte, la paquetería⁸³ es un complemento tecnológico en la computadora con el cual se pueden desarrollar las funciones más básicas y usadas de la computadora, por lo indispensables. Cobra más importancia cuando hay ausencia del acceso a internet o su proporción de acceso con respecto al total de computadoras en el MSD, es bajo.

Debido a su importancia, se utiliza nuevamente el criterio de los cuatro niveles para dividir en dos variables a “Internet”: Acceso a internet y velocidad de la conexión a internet.

Donde tener acceso⁸⁴ recibió el mayor coeficiente de ponderación (75%) dentro de Internet. Esto se decide al tomar en cuenta el contexto de las localidades donde fueron ubicados los módulos, se entiende que al ser un servicio gratuito, lo que se prefiere es que cuenten con el servicio y pasa a segundo término la calidad, porque la ausencia de servicio implicaría que tendrían que acudir a otro lugar y pagar por él. En consecuencia la calidad o velocidad de la conexión recibió una ponderación menor (25%).

En algunas ocasiones para un indicador se contó con información desde el punto de vista del encargado del módulo y del usuario. Como por ejemplo la frecuencia de asistencia. En estos casos, se decidió asignarle mayor peso a la información del usuario (75%) que a la del encargado del módulo (25%). El punto fundamental de este criterio, es considerar que los usuarios son los beneficiarios del programa y su opinión debería ser crítica. Y del lado de los encargados de módulo se determinó que puede verse afectada al sentir que la calificación que él asigna, implícitamente califica su desempeño laboral en el módulo.

Para asignar los coeficientes de ponderación de los actores en caso de participar ambos en un indicador, nuevamente se utiliza la estrategia de los cuatro niveles para ponderar el peso que se le asignaría a cada uno.

Información proveniente de:	Coeficiente de Ponderación
Usuarios	75%
Encargado de Módulo	25%

Ejemplo de la asignación de pesos en el indicador “Velocidad de internet”.

Indicador	Coeficiente de Ponderación del Indicador	Actor	Coeficiente de Ponderación del actor
Velocidad de Internet	25%	Usuario	75%
		Encargado de Módulo	25%

⁸³ Se refiere a una lista de programas y librerías de software.

⁸⁴ Con acceso se refiere a la proporción de computadoras conectadas a internet, con respecto del total de computadoras presentes en el MSD.

Como último paso para la construcción del índice, resta aplicar la cuarta fase “Formación de índices”, donde el objetivo es sistematizar los datos obtenidos en las fases anteriores. Entonces se conforma el índice de “Desempeño del PASDBP” mediante los indicadores construidos en la fase tres.

Para integrar los indicadores de las tres dimensiones, nuevamente se utilizan los cuatro niveles para asignarle a cada uno de ellos su ponderación al interior del índice. Se define que la mayor parte del índice se construye de la dimensión Servicios, otorgándole un 50%. En un segundo lugar de importancia se encuentran las Habilidades Tecnológicas de los Encargados del Módulo y Asistencia con un 25% cada una.

iii. Resultados del Índice de “Desempeño del PASDBP”

Finalmente el Índice de “Desempeño del PASDBP” queda expresado con las dimensiones de la siguiente manera:

$$\text{Índice Desempeño} = 0.50\text{Servicios} + 0.25\text{Habilidades Tecnológicas del Encargado} + 0.25\text{Asistencia}$$

En donde cada una de las dimensiones se expresa como:

$$\text{Servicios} = 0.50\text{Herramientas Tecnológicas} + 0.25\text{Asesoría y Capacitación} + 0.25\text{Infraestructura}$$

$$\text{Habilidades Tecnológicas del Encargado} = \text{Uso y Experiencia con los Programas}$$

$$\text{Asistencia} = 0.75\text{Frecuencia} + 0.25\text{Tipo de Usuarios}$$

Dentro de la dimensión Servicios, cada una de las subdimensiones se expresa de la siguiente forma:

$$\text{Herramientas Tecnológicas} = 0.75(0.75\text{Acceso a Internet} + 0.25\text{Velocidad de Internet}) + 0.25\text{Paquetería}$$

$$\text{Asesoría y Capacitación} = 0.50\text{Asesoría} + 0.50\text{Cursos}$$

$$\text{Infraestructura} = 0.50\text{Condiciones Físicas del MSD} + 0.50\text{Otros Servicios}$$

Por último presentamos la expresión del Índice de “Desempeño del PASDBP” utilizando el total de sus indicadores y queda como:

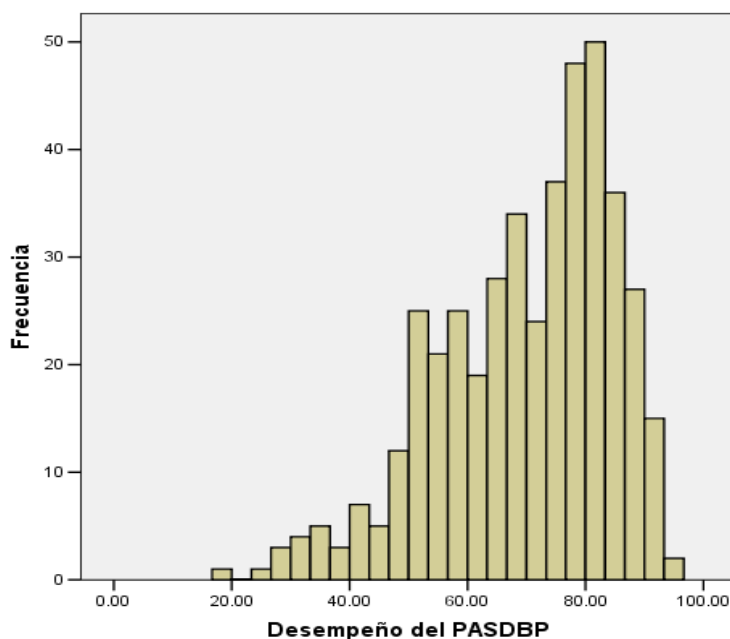
$$\text{Índice Desempeño del PASDBP} = 0.50 (0.50 (0.75 (0.75\text{Acceso a Internet} + 0.25\text{Velocidad de Internet}) + 0.25\text{Paquetería}) + 0.25 (0.50\text{Asesoría} + 0.50\text{Cursos}) + 0.25 (0.50\text{Condiciones$$

Físicas del MSD+0.50Otros Servicios))+0.25(Usos y Experiencia con los Programas)+0.25(0.75Frecuencia+0.25Tipo de Usuarios)

A continuación se presenta la información⁸⁵ obtenida del índice “Desempeño del PASDBP”, por medio de algunas medidas de tendencia central y dispersión para describir los resultados obtenidos al calcular el índice.

De las 432 bibliotecas para las que se realizó el cálculo del índice para medir su desempeño, se encuentra que la calificación más baja obtenida fue de 18.6 y la más alta obtenida fue de 95.1 dentro del intervalo [0, 100]. Es decir, la calificación mínima que podía ser obtenida en el índice era cero y la máxima 100 por lo tanto no se alcanza ni el mínimo ni el máximo, pero se acerca más al máximo. Por otra parte se encontró que el promedio de desempeño obtenido fue de 69.9 y su desviación estándar fue de 14.8.

Para obtener una representación gráfica de la distribución del índice, a continuación se presenta el histograma⁸⁶ correspondiente al índice “Desempeño del PASDBP”, con el cual se continuará la descripción del índice final obtenido.



⁸⁵ Las frecuencias de la variable “Desempeño” se pueden consultar en el Anexo “Construcción del Índice de Desempeño del PASDBP”.

⁸⁶ El histograma es una representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados. En el eje vertical se representan las frecuencias, y en el eje horizontal los valores de las variables, normalmente señalando las marcas de clase, es decir, la mitad del intervalo en el que están agrupados los datos.

Lo primero que se nota en el histograma es que el índice presenta una distribución asimétrica, unimodal y sesgada a la izquierda. Obsérvese que la moda (refiriéndonos a datos agrupados) del índice del “Desempeño del PASDBP” se encuentra en el intervalo modal [80, 83.333].

Otra estadística descriptiva que se puede utilizar para describir el índice final, son los cuantiles⁸⁷, para el índice de “Desempeño del PASDBP” se obtiene específicamente los percentiles 0.25, 0.50 y 0.75 también llamados cuartiles. La siguiente tabla muestra los valores obtenidos.

Cuartiles	
0.25	59.61
0.50 ⁸⁸	72.96
0.75	81.88

A partir de estos datos, se puede confirmar que es baja la incidencia de un bajo desempeño, ya que menos de una cuarta parte se encuentra por debajo de una calificación de 59; a diferencia de que poco más de un 25% tiene una calificación mayor a 80. La información de los cuartiles sirve para complementar la información sobre el tipo de distribución que se observa en el histograma, como un índice que presenta un sesgo negativo (sesgada a la izquierda). Por lo que concentra un mayor número de bibliotecas en niveles de desempeño alto.

Análisis exploratorio de la variable dependiente

Ahora que se ha descrito la variable dependiente mediante algunas medidas de tendencia central y dispersión, se procede a realizar un análisis del comportamiento de la variable “Desempeño del PASDBP”.

Este es un análisis que a pesar de ser simple, es recomendable realizar antes de comenzar cualquier tipo de estudio bivariado o multivariado para evitar posteriores fallas o complicaciones en los procedimientos.

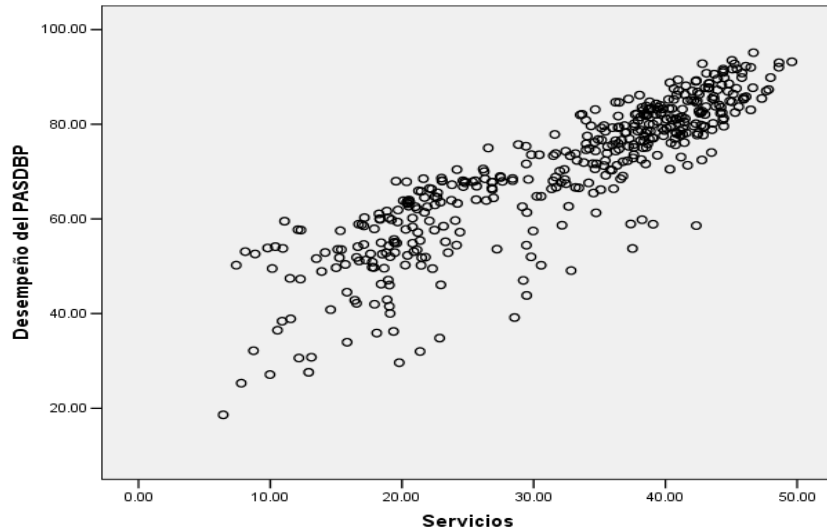
Para el análisis exploratorio de la variable dependiente “Desempeño” se realizará una validación interna. La validación interna de un índice se realiza con el análisis de cruces:

- Con variables que lo conformaron. En este caso utilizaremos las dimensiones: Servicios, Herramientas Tecnológicas de los Encargados de Módulo y Asistencia.

Para comenzar la validación interna se analiza el cruce con la dimensión Servicios a través de un diagrama de dispersión para observar la relación que existe.

⁸⁷ Son medidas de posición no central que permiten conocer puntos característicos de la distribución que no son valores centrales.

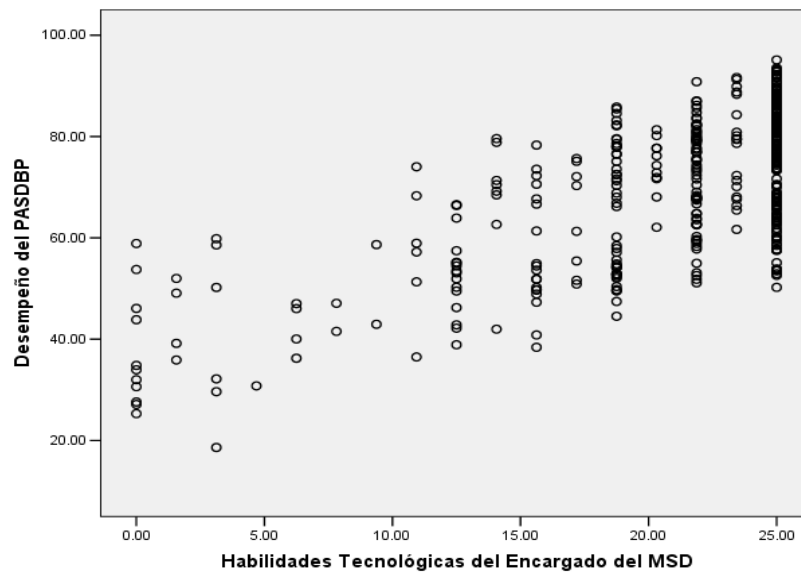
⁸⁸ El percentil 0.50, es también llamado Mediana.



Como se puede observar en el diagrama de dispersión, a mayor calificación en la dimensión Servicios, se espera un mejor nivel de desempeño.

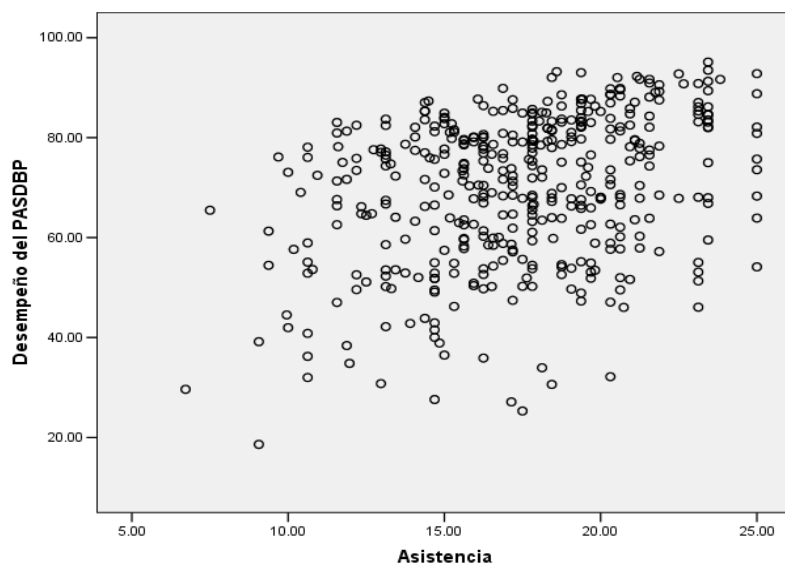
Se puede observar cómo la mayor concentración de puntos en el diagrama se encuentra en las calificaciones de Servicios y Desempeño más altas. También existen algunos casos que sugieren que la calificación de los Servicios es baja pero la del Desempeño es más alta, lo que sugiere que se compensa con las otras dimensiones. Y otros pocos, por el contrario, sugieren que se apoyan en la Dimensión de Servicios para compensar bajas calificaciones en las otras dos dimensiones.

Con respecto a la dimensión de Herramientas Tecnológicas de los Encargados se tiene el siguiente diagrama de dispersión.



Nuevamente se puede observar la relación que existe entre el desempeño del PASDBP y las Habilidades Tecnológicas de los encargados parecida a la anteriormente analizada de los Servicios y Desempeño del programa, sin embargo, se puede advertir cómo los puntos del diagrama se encuentran más dispersos pero siguen presentando la misma tendencia. Además se puede notar la concentración de casos en niveles altos de habilidades tecnológicas de los encargados.

Pasando a la dimensión de Asistencia se tiene el siguiente diagrama de dispersión.



Se observa que, a diferencia de los dos diagramas de dispersión anteriores, éste presenta casos más dispersos y que la mayor concentración de casos se encuentra en los valores 15-20 de Asistencia y niveles altos de Desempeño. Esto indica que la mayoría de las bibliotecas con mejores niveles de desempeño no alcanzan un nivel de asistencia al MSD óptimo, y por consiguiente se apoyan en Servicios y Habilidades Tecnológicas de los Encargados para un mejor funcionamiento del equipo.

Gracias a este análisis exploratorio de la variable respuesta se puede tener una panorámica del “Desempeño del PASDBP” donde se revelan sus fortalezas y señalan sus debilidades al interior del concepto, lo que nos ayuda a comprender mejor cómo se estructura del desempeño PASDBP.

Variable ordinal “Desempeño del PASDBP”

Después de estudiar los resultados obtenidos para cada biblioteca del índice del “Desempeño del PASDBP”, es importante comenzar a considerar el tratamiento más adecuado para la variable dependiente obtenida de acuerdo al modelo que se aplicará para el análisis de los determinantes del desempeño del programa.

Para elegir el modelo se considerarán los siguientes elementos:

- 1) Se tiene una variable dependiente numérica de razón o proporción, y

- 2) La mayoría de las variables independientes son categóricas, ya sea nominales u ordinales.

Por otra parte, asignar la calificación del desempeño de una biblioteca con un valor numérico (en este caso dentro de un intervalo [0,100]) no puede proveer directamente de un criterio que permita comprender qué tan bueno o malo es su desempeño, para resumir de manera simple el resultado de la variable dependiente. Este criterio es importante por tratarse de una variable que busca cuantificar un fenómeno social.

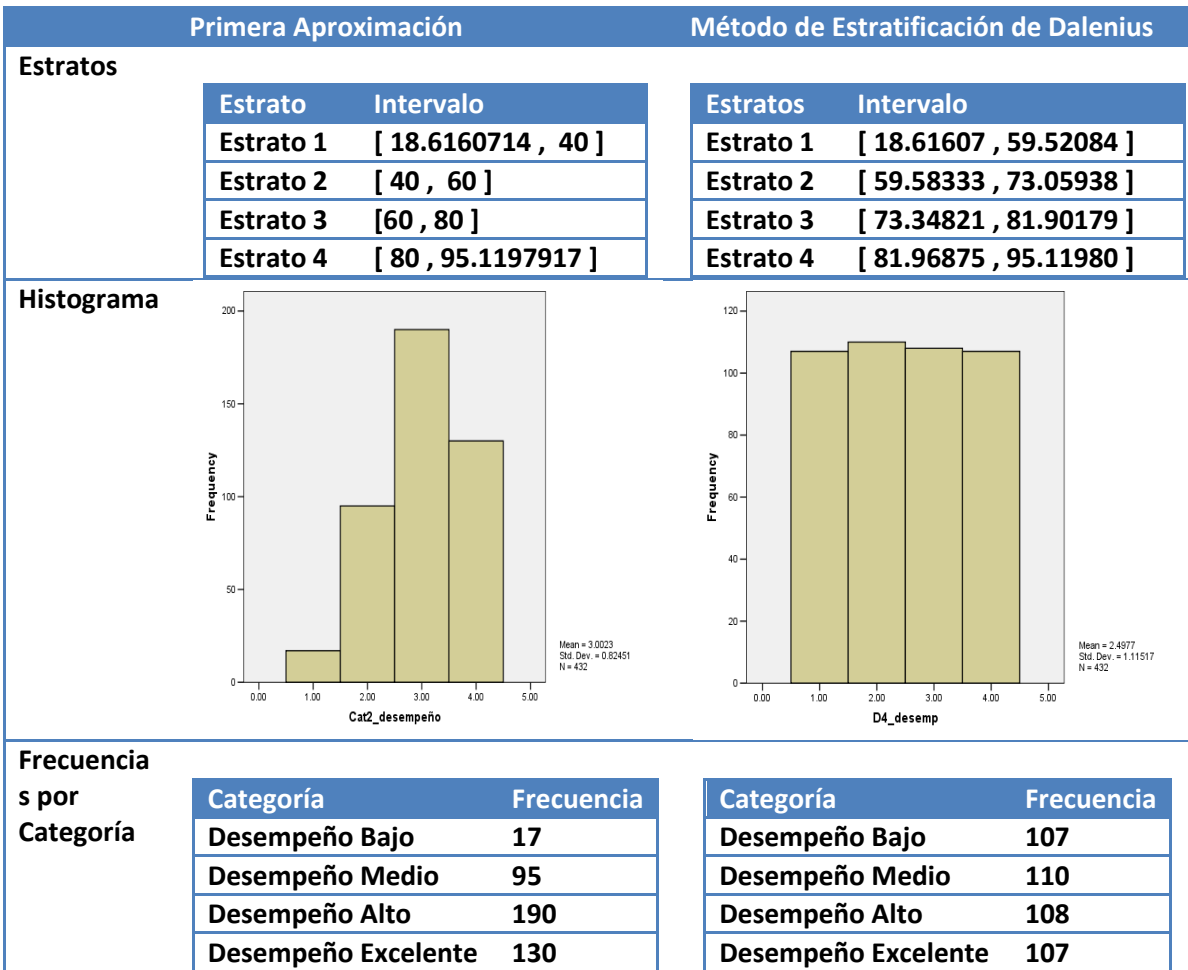
Si bien la variable numérica “Desempeño del PASDBP” presenta ventajas en el análisis estadístico, la naturaleza categórica de las variables independientes pudiera no ser adecuada en diversos métodos estadísticos que utilicen una variable dependiente numérica. Lo que lleva al cuestionamiento de si es la mejor opción conservar la variable dependiente como numérica o si bien, debe considerarse la transformación de la variable “Desempeño del PASDBP” a una de tipo categórica.

Para decidir qué es más conveniente, se analizaron los métodos estadísticos que podrían utilizarse para el análisis que tiene por objetivo esta tesis. Al final se decide que la regresión ordinal, que es un método que utiliza una variable categórica ordinal y variables independientes que pueden ser numéricas o categóricas, es la mejor opción.

Entonces, para poder llevar a cabo la regresión ordinal se debe categorizar la variable dependiente “Desempeño del PASDBP” para convertirla en una variable ordinal. A pesar de que se pierde información en el procedimiento, no se pierden todas las bondades que tiene la variable numérica ya que se conserva el orden. Por lo que el siguiente paso es decidir cómo categorizarla.

Para ello, se propone comparar dos métodos: uno basado en una categorización empírica “Primera aproximación” en la que se realizan cuatro cortes iguales en función del valor máximo y mínimo alcanzado, y otro basado en la teoría de estratificación “Método de Estratificación de Dalenius”⁸⁹. En la siguiente tabla se presenta información obtenida a partir de su aplicación.

⁸⁹ Para mayor información sobre este método se puede consultar el anexo 3 sobre los métodos de estratificación.



Debido a que en la primera estratificación con el método de “Primera Aproximación” no se considera la distribución de los casos, sino únicamente el rango de los valores, se obtienen como resultado estratos muy diferentes en cuanto al número de casos. En especial, la primera categoría presenta una frecuencia considerablemente baja en comparación con las demás.

Mientras que con el Método de Estratificación de Dalenius se puede observar cómo la variable recodificada bajo este método presenta una distribución más uniforme. Esto se debe a que la “Primera Aproximación” es un método meramente empírico que define las categorías a partir de un criterio de longitud del intervalo.

Tomando en cuenta que el método de regresión ordinal utiliza tablas de contingencia para manejar la información, es importante considerar la frecuencia de las categorías de la variable respuesta con el fin de prever el comportamiento que podrían mostrar en tablas de contingencia⁹⁰.

⁹⁰ La tabla de contingencia es una tabla de doble entrada donde se representa simultáneamente dos variables en una misma población. Si las variables son X y Y , y el tamaño de la muestra n . Las clases de X se escriben como c_1, c_2, \dots, c_i y la de Y como d_1, d_2, \dots, d_j . Entonces se tiene una tabla de $i \times j$. Cada fila y columna corresponden a una submuestra particular.

Entonces si la variable “Desempeño” presenta categorías con frecuencia baja, es posible que se obtengan casillas vacías⁹¹. Por lo tanto parece más conveniente tener variable dependiente que contenga una frecuencia similar en todas las categorías.

Finalmente, se considera que para este caso el Método de Estratificación de Dalenius es el mejor método de estratificación y se toma la variable ordinal, que se obtiene al aplicarlo, como la variable dependiente para el análisis de los factores determinantes del desempeño del PASDBP.

iv. Indicadores de los factores que afectan el desempeño del PASDBP

A continuación se presenta un análisis descriptivo de las variables independientes consideradas para el análisis estadístico. En algunos casos la información se pudo obtener de manera directa y en otros se construyeron índices.

Si bien, es permitido dentro de un modelo de regresión ordinal tener variables independientes de diferente naturaleza, se tiene la problemática de generar tablas de contingencia extensas cuando se trabaja con variables numéricas o con variables categóricas con un número grande de categorías.

Otro inconveniente relacionado a las tablas de contingencia múltiples extensas es generar un número considerable de celdas vacías que provoca problemas en la aplicación de algunas pruebas estadística, por lo que se debe tener cuidado al trabajar con un gran número de variables independientes y la naturaleza de éstas.

Entonces, se considera conveniente transformar las variables numéricas en variables categóricas de escala ordinal. Con ello, además de reducir las casillas vacías, también se resume mejor la información que presenta la variable y por consiguiente se podrá entenderla más fácilmente.

También se decide recategorizar aquellas variables con un excesivo número de categorías o aquéllas en las que pueda presentarse con mayor claridad la información, es decir, la interpretación de la variable se vuelve más sencilla si se reduce el número de categorías. Pero es importante remarcar que sólo debe realizarse en aquellas variables donde la unión de categorías, sinteticé de forma adecuada la información siguiendo criterios teóricos que ayuden a representar de forma lógica la información expuesta en el indicador. Para la recodificación o recategorización se utilizaron los siguientes procedimientos: “Método de Estratificación de Dalenius”, “Criterio Teórico” y “Criterio Empírico⁹²” según fuese el caso.

⁹¹ Las casillas, también llamadas celdas, vacías son un problema a la hora de realizar algunas pruebas estadísticas. Por ejemplo, en la Prueba Ji-Cuadrada es necesario que todas las celdas cuenten con al menos 5 casos para poder realizarla correctamente.

⁹² Para este método la recategorización se realiza con criterios en base a la observación, experiencia y sentido de las respuestas.

En los siguientes cuadros se presentan las variables independientes por los grupos anteriormente definidos⁹³. El primer grupo que se presenta será el de las Características de la Localidad, el cual tiene sólo dos variables independientes.

a. Características de la Localidad.

Variables Independientes	n	Distribución porcentual
Tipo de Localidad		
• Urbano	328	75.9%
• Rural	104	24.1%
Total	432	100%
Nivel de Marginación		
• Bajo	268	62.0%
• Medio	91	21.1%
• Alto	73	16.9%
Total	432	100%

De acuerdo al tipo de localidad, el 75.9% de las bibliotecas está en localidades urbanas y el resto (24.1%) en localidades rurales. En cuanto al nivel de marginación, el 62% de las localidades en donde se han equipado las bibliotecas, tienen un nivel de marginación bajo; el 21.1% medio y el 16.9% alto.

El siguiente es el grupo de características del Módulo de Servicios digitales, conformado por seis variables independientes.

b. Características del MSD

Variables Independientes	n	Distribución porcentual
Experiencia/ Tiempo en el PASDBP⁹⁴		
• Muy poca experiencia	88	22.2%
• Poca experiencia	124	31.3%
• Mediana experiencia	100	25.3%
• Mucha experiencia	84	21.2%
Total	396	100%
Antigüedad de la Biblioteca⁹⁵		

⁹³ Los grupos fueron definidos en el ii. Definición de las variables independientes del a) Definición del desempeño del PASDBP y de los factores que pudieran estarlo influyendo, del presente capítulo.

⁹⁴ Originalmente la variable “Experiencia/Tiempo en el PASDBP” era numérica, y mediante “Método de Estratificación de Dalenius” se realizó la categorización presentada en la tabla.

⁹⁵ Originalmente la variable “Antigüedad de la Biblioteca” era numérica, y mediante “Método de Estratificación de Dalenius” se realizó la categorización presentada en la tabla.

• Hasta 14 años	64	18.4%
• Más de 14 y hasta 20 años	96	27.5%
• Más de 20 y hasta 23 años	122	35.1%
• Más de 23 años	66	19.0%
Total	348	100%
Tamaño de la Biblioteca⁹⁶		
• Adecuado para el MSD	289	66.9%
• Inadecuado para el MSD	143	33.1%
Total	432	100%
Número de Computadoras Asignadas⁹⁷		
• 2 o 3 computadoras	158	36.6%
• 4 – 6 computadoras	221	51.2%
• 9 – 13 computadoras	53	12.3%
Total	432	100%
Sostenibilidad⁹⁸		
• Cuenta con plan para continuar el MSD	171	39.6%
• No cuenta con plan para continuar el MSD	261	60.4%
Total	432	100%
Financiamiento⁹⁹		
• Sí tienen financiamiento por parte de las autoridades municipales	322	74.5%
• No tienen financiamiento por parte de las autoridades municipales	110	25.5%
Total	432	100%

⁹⁶ Originalmente la variable “Tamaño de la biblioteca” era ordinal de tres categorías. En una de sus categorías definida como “Mas o menos adecuado para el MSD” presenta una baja frecuencia, por lo que se decidió realizar una recategorización con la finalidad de generar una variable más uniforme y más simple. Mediante un “Criterio Empírico” se realizó la recategorización que redujo la variable a dos categorías.

⁹⁷ Originalmente la variable “Número de computadoras asignadas” era numérica, y mediante el “Criterio Empírico” que tomó en cuenta la clasificación otorgada por el PASDBP de: Atención Especial, Cobertura y Soporte se realizó la categorización.

⁹⁸ Originalmente la variable “Sostenibilidad” era de tipo ordinal, y mediante un “Criterio Empírico” se decidió juntar las categorías de “Cuenta más o menos con un plan para continuar el programa” y “No sabe”, al presentar 7 y 45 casos respectivamente, con la categoría de “No cuenta con un plan para continuar el programa” se realizó la categorización.

⁹⁹ Originalmente la variable “Financiamiento” era ordinal y por medio de un “Criterio Empírico” que decidió juntar las categorías de “En parte” y “No”, al presentar 68 y 45 se realizó la categorización presentada en la tabla

De acuerdo a la experiencia o tiempo que llevan las bibliotecas en el programa del PASDBP, se encuentra que un 22.2% tiene muy poca experiencia, el 31.3% tiene poca experiencia, el 25.3% tiene mediana experiencia y el 21.2% cuenta con mucha experiencia en el programa.

Por otra parte, un 18.4% de las bibliotecas 14 años de antigüedad o menos , un 27.5% tiene más de 14 y hasta 20 años de antigüedad, el 35.1% de más 20 y hasta 23 años de antigüedad y por último sólo un 19% tiene más de 23 años de antigüedad.

En cuanto al tamaño de la biblioteca, sólo el 33.1% considera que el espacio asignado para el módulo es inadecuado para la biblioteca y que el restante 66.9% sí lo considera adecuado.

Otro factor de las características del MSD es el número de computadoras que les fueron asignadas por parte del PASDBP. Los diferentes tipos de asignación del programa son los siguiente: 3 computadoras a las bibliotecas clasificadas como “Atención Especial¹⁰⁰”, 6 computadoras a las bibliotecas catalogadas como de “Cobertura¹⁰¹” y 12 a las bibliotecas clasificadas como “Soporte¹⁰²”. Por lo tanto, si existen valores diferentes a 3, 6 y 12 pudiesen ser resultado de que la respuesta de los encargados de módulo pudiese reflejar un desconocimiento o confusión de contabilizar las computadoras que al momento tiene el MSD, ya sean más o menos de las asignadas inicialmente¹⁰³. Entonces se tiene que el 36.6% de las bibliotecas se les asignó 2 o 3 computadoras, al 51.2% se le asignaron de 4-6 computadoras y al 12.3% se les asignó de 9-13 computadoras.

En cuanto a la sostenibilidad, se encuentra que el 60.4% de las bibliotecas no cuentan con un plan para continuar el programa, mientras que el restante 39.6% sí cuenta con el plan. Es decir, más

¹⁰⁰ Las Bibliotecas de tipo “Atención Especial” tienen las siguientes características:

- La superficie de la Biblioteca es de 40-60 metros cuadrados.
- Capacidad de atención simultanea de hasta 20 usuarios.
- Cuentan con 1-2 encargados de módulo.
- El espacio dedicado al Módulo de Servicios Digitales es de 10-15 metros cuadrados.

¹⁰¹ Las Bibliotecas de tipo “Cobertura” tienen las siguientes características:

- La superficie de la Biblioteca es de 60-200 metros cuadrados.
- Capacidad de atención simultanea de hasta 60 usuarios.
- Cuentan con 2-7 encargados de módulo.
- El espacio dedicado al Módulo de Servicios Digitales es de 15-25 metros cuadrados.

¹⁰² Las Bibliotecas de tipo “Soporte” tienen las siguientes características:

- La superficie de la Biblioteca es de 250-1000 metros cuadrados.
- Capacidad de atención simultanea de hasta 200 usuarios.
- Cuentan con 10-20 encargados de módulo.
- El espacio dedicado al Módulo de Servicios Digitales es de 25-40 metros cuadrados.

¹⁰³ Es decir, podría ser que las computadoras asignadas a la biblioteca, hayan sido trasladadas para otro uso fuera del módulo, estén descompuestas o en reparación y no estén físicamente en el módulo al momento de la encuesta, que las computadoras presentes en el módulo sean las otorgadas por el programa y también estén presentes las que hayan sido otorgadas por el gobierno o cualquier otro patrocinador antes o después del programa. Y también existe la posibilidad de que pasen todas las anteriores situaciones al mismo tiempo o algunas de ellas.

de la mitad de las bibliotecas no tiene todavía un plan a largo plazo para continuar el funcionamiento del MSD.

Y por último, la mayoría de las bibliotecas (74.5%) tiene financiamiento por parte de las autoridades municipales o delegacionales (según sea el caso de la localidad) y sólo un 25.5% no cuenta con financiamiento por parte de las autoridades.

En el siguiente grupo de las Características del Encargado del Módulo, se tiene cuatro variables independientes.

c. Características del Encargado de Módulo

Variables Independientes	n	Distribución porcentual
Tipo de Encargado		
• EM	227	52.5%
• EM-B	205	47.5%
Total	432	100%
Tiempo de trabajar en el MSD¹⁰⁴		
• Menos de un año	114	29.4%
• De 1 año y menor a 2.5 años	123	31.7%
• De 2.5 y menor a 9 años	79	20.4%
• De 9 años o más	72	0.185%
Total	388	100%
Escolaridad del Encargado¹⁰⁵		
• Educación Básica	59	13.8%
• Educación Media Superior	244	57.3%
• Educación Superior	123	28.9%
Total	426	100%
Capacitación recibida por la DGB¹⁰⁶		
• 0 -1 curso	101	23.4%
• 2 – 4 cursos	187	43.3%
• 5 -7 cursos	144	33.3%
Total	432	100%

¹⁰⁴ Originalmente la variable “Tiempo de trabajar del encargado” era numérica, y mediante “Método de Estratificación de Dalenius” se realizó la categorización presentada en la tabla.

¹⁰⁵ Originalmente la variable “Escolaridad del Encargado” era ordinal, y mediante un “Criterio Teórico” se realizó la categorización presentada en la tabla.

¹⁰⁶ Originalmente la variable “Capacitación recibida por la DGB” era de tipo numérica, y por medio de un “Criterio Empírico” se realizó la categorización presentada en la tabla.

De acuerdo al tipo de encargado del módulo, el 52.5% de las bibliotecas cuentan con una persona encargado exclusivamente de las actividades del MSD y el restante 47.5% de las bibliotecas tiene que compartir el encargado del módulo con la biblioteca.

En cuanto al tiempo de trabajar del encargado, se tiene que un 29.4% de las bibliotecas cuentan con un encargado que tiene menos de un año trabajando, un 31.7% cuentan con un encargado que ha trabajado de 1 año y menor a 2.5 años, 20.4% con uno que ha trabajado de 2.5 a menos de 9 años y sólo 18.5% de las bibliotecas cuentan con encargados que tenga 9 años o más trabajando.

Por otra parte, con respecto a la escolaridad del encargado el 13.8% de las bibliotecas tiene a cargo a una persona con escolaridad de Educación Básica (esta categoría incluye: primaria incompleta, primaria completa, secundaria incompleta y secundaria completa), en el 57.3% de las bibliotecas el encargado tiene nivel de escolaridad de Educación Media Superior (esta categoría incluye: carrera técnica, carrera comercial, preparatoria incompleta y preparatoria completa) y en el restante 28.9% el encargado es una persona con nivel de Educación Superior (esta categoría incluye: licenciatura incompleta, licenciatura completa y posgrado).

Y finalmente se llega al grupo de Difusión que únicamente contiene un variable referente a la difusión por parte del encargado del MSD.

d. Difusión

Variables Independientes	n	Distribución porcentual
Difusión por parte del Encargado¹⁰⁷		
• Sí	366	84.7%
• No	66	15.3%
Total	432	100%

De acuerdo con la variable “Difusión por parte del Encargado” en el 84.7% de las bibliotecas la persona a cargo del módulo realiza difusión de las actividades y servicios que presta el MSD, mientras que el restante 15.3% no tiene difusión por parte del actor Encargado del MSD.

En resumen, a continuación se presenta un cuadro con información de todas las variables independientes que se utilizarán en el análisis estadístico multivariado. Gracias al proceso de categorizar o recategorizar (según sea el caso de la variable), se cuenta con variables independientes más apropiadas al método de regresión ordinal que se empleará para modelar el “Desempeño del PASDBP” a través de los factores propuestos.

¹⁰⁷ Originalmente la variable “Difusión por parte del encargado” era ordinal con tres categorías, y mediante un “Criterio Empírico” se realizó la categorización presentada en la tabla al juntar las categorías “En parte” y “No”.

Grupo de Variables Independientes	Variable Independiente	Tipo de Variable	Número de Categorías
Características de la Localidad	Tipo de Localidad	Nominal	2
	Nivel de Marginación	Ordinal	3
Características del Módulo de Servicios Digitales	Experiencia/Tiempo en el PASDBP	Ordinal	4
	Antigüedad de la Biblioteca	Ordinal	4
	Tamaño de la Biblioteca	Ordinal	2
	Número de computadoras asignadas	Ordinal	3
	Sostenibilidad	Ordinal	2
	Financiamiento	Ordinal	2
Características del Encargado de Módulo	Tipo de Encargado de Módulo	Nominal	2
	Tiempo de trabajar del Encargado de Módulo	Ordinal	4
	Escolaridad del Encargado de Módulo	Ordinal	3
	Capacitación recibida por la DGB	Ordinal	3
Difusión	Difusión por parte del Encargado de Módulo	Ordinal	2

d) Análisis exploratorio de las variables

En este apartado se quiere realizar una primera aproximación a la relación entre la variable dependiente “Desempeño del PASDBP” y cada una de las variables explicativas.

De modo que sea posible afirmar que las variables independientes que son propuestas para ser incluidas en el modelo de regresión ordinal, como factores influyentes en el desempeño del PASDBP, muestran relación con la variable dependiente.

i. Análisis bivariado de la variable dependiente vs. las variables independientes

En este apartado se analiza si la variable dependiente “Desempeño del PASDBP” está asociada con cada una de las variables independientes.

El análisis bivariado que se efectúa es únicamente de carácter exploratorio. Porque sólo se quiere conocer si existe algún grado de asociación entre la variable dependiente y cada una de las independientes.

Debido a que se tienen dos tipos de variables independientes: Nominal y Ordinal. Se harán dos tipos de pruebas de asociación para poder explotar de manera óptima los tipos de variable. Ya

que la variable dependiente con la que se trabajará es ordinal, las pruebas de asociación que se realizarán son *ordinal vs nominal* y *ordinal vs ordinal*.

Por lo tanto, las pruebas de asociación¹⁰⁸ adecuadas para realizar el análisis bivariado son: la Prueba de χ^2 (Ji – Cuadrada)¹⁰⁹ en caso de presentar una variable independiente nominal y la Prueba de Gamma (de Goodman y Kruskall) si se tiene una variable independiente ordinal.

Y se establece una $\alpha = 0.05$ para realizar el contraste de las siguientes hipótesis:

$$H_0: \text{Independencia de } X_i \text{ e } Y^{110} \quad \text{vs} \quad H_a: \text{Asociación de } X_i \text{ e } Y$$

Donde Y se refiere a la variable “Desempeño del PASDBP” y X_i se refiere a la variable independiente i con $i = 1, 2, \dots, 13$ ¹¹¹

Entonces si se obtiene un $p - \text{value} < \alpha = 0.05$ se rechaza la hipótesis nula y se acepta que existe asociación entre el “Desempeño del PASDBP” y la variable independiente X_i .

Con el fin de comparar la intensidad de la asociación de cada una de las variables independientes con el desempeño del programa (si es que existe asociación), se realizará además el cálculo del *Coefficiente de contingencia*, el cual nos dará un valor C entre $0 \leq C < 1$, para obtener así una medida de la fuerza de asociación.

A continuación se muestra el resumen de los valores obtenidos en las pruebas de asociación correspondientes a nuestra variable dependiente con cada una de las variables independientes de las características de la localidad, características del MSD, características del encargado de módulo y difusión.

Para comenzar presentamos el siguiente cuadro con los resultados de las pruebas aplicadas al grupo de características de la localidad.

¹⁰⁸ En caso de requerir información detallada sobre ambas pruebas de asociación : Ji-Cuadra y Gamma (de Goodman y Kruskall) revisar el Anexo: “Pruebas de Asociación para el análisis bivariado”.

¹⁰⁹ La prueba de asociación de la Ji-Cuadrada se puede realizar con variables ordinal vs ordinal, sin embargo se obtendría un resultado con menor eficacia al desestimar la cualidad de orden que posee.

¹¹⁰ Otra forma de decir que existe independencia entre X e Y , es que “ X no está asociada con Y ” o bien que “no hay relación entre las variable X e Y ”.

¹¹¹ Las trece variables independientes pueden ser consultadas en el último cuadro del apartado anterior.

Resultados de las pruebas de asociación entre el desempeño del PASDBP y las variables de características de la localidad

Grupo	Variable	Tipo de Variable	Ji-Cuadrada / Gamma	Grados de Libertad	P-value	Coefficiente de Contingencia
Características de la Localidad	Tipo de Localidad	Nominal	20.937	3	0.000	0.215
	Nivel de Marginación	Ordinal	-0.143	-	0.030	0.147

○ **Características de la localidad.**

Para este grupo las hipótesis a probar son:

H_0 : No existe asociación entre el desempeño del PASDBP y X_i

vs

H_a : Existe asociación entre el desempeño del PASDBP y X_i

Con $i = 1, 2$ donde:

X_1 = Tipo de Localidad

X_2 = Nivel de Marginación

Como se puede observar en el cuadro anterior, en ambos casos se rechaza la hipótesis nula y se acepta que existe asociación entre el desempeño del programa y X_i , con $i = 1, 2$.

Sin embargo, la fuerza de asociación no fue la misma. El *Coefficiente de Contingencia* para la variable Tipo de Localidad resultó ser mayor que para el Nivel de Marginación, lo que sugiere que la relación es más fuerte para la variable Tipo de Localidad.

A continuación, se presentan los resultados de las pruebas de asociación entre las variables que componen el grupo de características del módulo de servicios digitales y el desempeño del PASDBP.

Resultados de las pruebas de asociación entre el desempeño del PASDBP y las variables de características del módulo de servicios digitales

Grupo	Variable	Tipo de Variable	Gamma	P-value	Coefficiente de Contingencia
Características del Módulo de	Experiencia/Tiempo en el PASDBP	Ordinal	0.180	0.001	0.236

Servicios Digitales	Antigüedad de la Biblioteca	Ordinal	0.121	0.044	0.184
	Tamaño de la Biblioteca	Ordinal	-0.064	0.401	0.043
	Número de Computadoras Asignadas	Ordinal	0.327	0.000	0.290
	Sostenibilidad	Ordinal	-0.192	0.007	0.158
	Financiamiento	Ordinal	-0.189	0.024	0.139

○ **Características del módulo de servicios digitales.**

Para este grupo las hipótesis a probar son:

$$H_0: \text{No existe asociación entre el desempeño del PASDBP y } X_i$$

vs

$$H_a: \text{Existe asociación entre el desempeño del PASDBP y } X_i$$

Con $i = 1, 2, \dots, 6$ donde:

X_1 = Experiencia / Tiempo en el PASDBP

X_2 = Antigüedad de la biblioteca

X_3 = Tamaño de la biblioteca

X_4 = Número de computadoras asignadas

X_5 = Sostenibilidad

X_6 = Financiamiento

Como se puede observar en el cuadro anterior, en cinco casos se rechaza la hipótesis nula y se acepta asociación con el Desempeño del programa. Las variables con las que se encontró asociación con el desempeño del PASDBP son las siguientes:

- “Experiencia /Tiempo en el Programa” y “Número de Computadoras Asignadas” además obtuvieron los *Coficiente de Contingencia* más altos, lo que significa que son las dos variables independientes del grupo que presentan una asociación más fuerte con la variable dependiente.
- “Antigüedad de la Biblioteca”, “Sostenibilidad”, y “Financiamiento”.

Debido a que la variable “Tamaño de la Biblioteca” no rechaza la hipótesis nula de independencia con el “Desempeño del PASDBP” será tratada con mucho cuidado en el capítulo siguiente.

En conclusión, la mayoría de las variables pertenecientes a este grupo presentaron algún grado de asociación.

Resultados de las pruebas de asociación entre el desempeño del PASDBP y las variables de características del encargado del MSD

Grupo	Variable	Tipo de Variable	Ji-Cuadrada / Gamma	Grados de Libertad	P-value	Coefficiente de Contingencia
Características del Encargado del Módulo	Tipo de Encargado	Nominal	35.838	3	0.000	0.277
	Tiempo de trabajar del Encargado de Módulo	Ordinal	-0.083	-	0.146	0.245
	Escolaridad del Encargado de Módulo	Ordinal	0.309	-	0.000	0.241
	Capacitación recibida por la DGB	Ordinal	0.249	-	0.000	0.226

○ **Características del encargado de módulo.**

Para este grupo las hipótesis a probar son:

$$H_0: \text{No existe asociación entre el desempeño del PASDBP y } X_i$$

vs

$$H_a: \text{Existe asociación entre el desempeño del PASDBP y } X_i$$

Con $i = 1, \dots, 4$ donde:

X_1 = Tipo de Encargado

X_2 = Tiempo de trabajar en el Módulo

X_3 = Escolaridad del Encargado de Módulo

X_4 = Capacitación recibida por la DGB

Como se puede ver en cuadro anterior, en tres casos se rechazan la hipótesis nula y se aceptan asociación con el Desempeño del programa. Las variables con las que se encontró asociación con el desempeño del PASDBP, y su fuerza de asociación, son las siguientes:

- Con mayor fuerza de asociación se tiene a la variable: “Tipo de Encargado”, seguido de “Escolaridad del Encargado de Módulo” y por último, la variable con el valor más bajo obtenido en el Coeficiente de Contingencia, “Capacitación recibida por la DGB”.

Sin embargo, cabe resaltar que no existe una diferencia considerable entre las fuerzas de asociación de las variables mencionadas.

Por otro lado, no resulta haber asociación entre la variable “Tiempo de trabajar en el Módulo” y el “Desempeño del PASDBP” por lo que será tratada con mucho cuidado en el capítulo siguiente.

Por último presentamos el cuadro con la información de la prueba aplicada al grupo de Difusión.

Resultados de las pruebas de asociación entre el desempeño del PASDBP y la difusión

Grupo	Variable	Tipo de Variable	Gamma	p-value	Coefficiente de Contingencia
Difusión	Difusión	Ordinal	-0.284	0.004	0.138

○ **Difusión.**

Para este grupo las hipótesis a probar son:

$$H_0: \text{No existe asociación entre el desempeño del PASDBP y } X_1$$

vs

$$H_a: \text{Existe asociación entre el desempeño del PASDBP y } X_1$$

Donde $X_1 =$ Difusión

Como se observa en el cuadro anterior, al aplicar la *Prueba de Gamma*, de “Sostenibilidad”, rechaza la hipótesis nula y se supone que existe una asociación entre la “Difusión” y el “Desempeño del PASBP”.

Gracias a este análisis bivariado de la variable dependiente “Desempeño del PASDBP” con cada una de las variables independientes se han identificado algunas asociaciones (once de las trece variables independientes presentaron asociación con la variable dependiente), que sirven como una primera aproximación a la relación que tienen los factores propuestos como determinantes del desempeño con éste.

Se encontró que seis variables tiene una buena asociación con el desempeño del PASDBP, y en especial cuatro de ellas: “Número de Computadoras”, “Tipo de Encargado”, “Escolaridad del Encargado” y “Experiencia/Tiempo en el PASDBP” presentan la asociación más fuerte con la variable dependiente. Por lo tanto, se espera sean variables estadísticamente significativas para el desarrollo de los modelos multivariados.

Sin embargo, los resultados que arrojaron las pruebas de asociación (bivariadas) no definen cuáles son las variables independientes que formarán parte de los modelos de regresión ordinal. Es decir, la influencia que tiene una variable independiente en el desempeño del programa puede cambiar al formar parte de un conjunto de variables que influyen simultáneamente, por lo que el efecto de ésta hacia el mismo desempeño puede mantenerse, modificarse o incluso no ser significativo.

CAPÍTULO 3. Aspectos teóricos de la Regresión Ordinal Múltiple

En este capítulo se exponen los conceptos teóricos necesarios para entender cómo es que se construye un modelo de regresión ordinal, sus fundamentos y su cálculo. Tener clara la teoría conlleva a una adecuada interpretación del modelo, y por lo tanto del tema de estudio.

La regresión ordinal es un método relativamente nuevo¹¹² y con un enfoque muy útil en investigaciones económicas y sociales, principalmente por el tipo de datos que modela.

A continuación se presentan una pequeña introducción a las técnicas multivariadas, para entender en qué tipo de métodos se encuentra clasificada la regresión ordinal y qué la diferencia de otras técnicas estadísticas. Posteriormente se describe detalladamente la teoría concerniente a la regresión ordinal múltiple.

a) Técnicas Multivariadas y Regresión Estadística

Cuando se realiza el estudio de un fenómeno mediante métodos estadísticos para recoger, organizar y analizar sus datos, así como para sacar conclusiones válidas y tomar decisiones razonables a partir de ello, se lleva a cabo un análisis estadístico¹¹³.

El análisis multivariado es una rama de la estadística cuya finalidad es analizar simultáneamente medidas múltiples (características o variables) de cada objeto o individuo sometido a investigación. Esto mediante diferentes métodos o técnicas estadísticas elegidas de acuerdo a la información de la que se dispone y los fines que se persiguen. Es decir, la naturaleza y condiciones de los datos, de las variables y del modelo que se desea obtener delimitan qué tipo de método se podrá utilizar.

Entonces los métodos estadísticos multivariados son herramientas estadísticas cuyo propósito principal es brindar un mejor entendimiento del fenómeno de estudio. Debido que al contemplar un mayor número de variables y estudiarlas simultáneamente se puede obtener un modelo con mayor significancia.

Podemos clasificar las técnicas multivariadas en dos grandes grupos (Hair, Anderson, Tatha, 1998):

¹¹² El diseño de la regresión ordinal se basa en la metodología de McCullagh (1980 y 1998).

¹¹³ Podemos distinguir tres tipos de análisis estadístico según el número de variables (características o aspectos) que se analicen conjuntamente: univariado (una sola variable), bivariado (dos variables) y multivariado (más de tres variables). Y es precisamente este último el que desarrollaremos en este capítulo para precisar cuáles son los factores que determinan el desempeño del Programa de Acceso a Servicios Digitales en Bibliotecas Públicas.

- 1) **Métodos de Dependencia.** Un análisis de dependencia se define como aquél en que una o un conjunto de variables se identifican como variables *dependientes*, las cuales son explicadas a través de otro conjunto de variables llamadas *independientes*. Algunos ejemplos de este tipo de métodos son los de Regresión. A su vez, este grupo puede dividirse en dos tipos debido a las siguientes características:
 - a) **Número de variables dependientes.** Se pueden clasificar los métodos de dependencia en aquéllos que únicamente quieren explicar una variable dependiente con múltiples independientes, y en aquellos métodos que cuentan con varias variables dependientes. Lo que puede generar varias variables independientes en una relación única o múltiples relaciones de variables dependientes e independientes.
 - b) **Tipos de escalas de medida utilizadas en las variables tanto dependiente(s) como independientes.** Dependiendo del tipo de escala de medida de las variables, se podrá optar por un método más adecuado para el tratamiento de ese tipo de datos. Es decir, que dependiendo si son variables métricas o no métricas (nominales u ordinales) se podrán aplicar métodos estadísticos que aprovechen de forma óptima la naturaleza de las variables de estudio.
- 2) **Métodos de Interdependencia.** Éste es un análisis en que ninguna variable o grupo de variables es definido como dependientes e independientes. El procedimiento de un análisis de interdependencia implica el análisis de todas las variables del conjunto simultáneamente. Algunos ejemplos de este tipo de técnicas de interdependencia son: Análisis Factorial¹¹⁴ y Análisis de Cluster¹¹⁵. Este tipo de métodos multivariados se puede dividir dependiendo:
 - a) **A quien esté dirigido el análisis de la estructura de las relaciones.** Esto puede ser estudiar la estructura de la relaciones entre variables, casos/encuestados u objetos.

Para fines de esta tesis son de interés los métodos de Dependencia, ya que debe recordarse que a lo largo de todo el capítulo dos se construyeron las trece variables explicativas y una variable dependiente para realizar el análisis estadístico de los factores determinantes en el desempeño del PASDBP. Dentro de este tipo de técnicas multivariadas, la más versátil y ampliamente usada es el análisis de Regresión¹¹⁶.

¹¹⁴ El análisis factorial es un nombre genérico que se le da a una clase de métodos estadísticos multivariantes cuyo propósito principal es definir la estructura subyacente en una matriz de datos. Aborda el problema de cómo analizar la estructura de las interrelaciones (correlaciones) entre un gran número de variables con la definición de dimensiones subyacentes comunes, conocidas como factores.

¹¹⁵ El análisis de Cluster es la denominación de un grupo de técnicas multivariantes cuyo principal propósito es agrupar objetos basándose en las características que poseen. Este análisis clasifica objetos de tal forma que cada objeto es muy parecido a los que hay en el conglomerado con respecto a algún criterio de selección predeterminado.

¹¹⁶ El término regresión fue introducido por Francis Galton en su libro *Natural inheritance* (1889) y fue confirmada por su amigo Karl Pearson. Su trabajo se centró en la descripción de los rasgos físicos de los descendientes (variable A) a partir de los de sus padres (variable B). Estudiando la altura de padres e hijos a partir de más de mil registros de grupos familiares, se llegó a la conclusión de que los padres muy altos tenían

La regresión se utiliza para predecir una medida basándose en el conocimiento de otra(s). Se conoce como análisis de regresión múltiple al método estadístico que permite establecer una relación matemática entre un conjunto de variables X_1, X_2, \dots, X_n (variables independientes, covariables, factores) y una variable dependiente Y. La finalidad del análisis de regresión múltiple es usar varias variables independientes, cuyos valores son conocidos para predecir la única "variable criterio"¹¹⁷ seleccionada por el investigador. Se utiliza fundamentalmente en estudios en los que no se pueden controlar por diseño los valores de las variables independientes (porque ya están dados). Los objetivos de un modelo de regresión puede ser dos:

- **Predicción:** se desea obtener una ecuación que permita "predecir" el valor de Y (variable dependiente o endógena) una vez conocidos los valores de X_1, X_2, \dots, X_n (variables exógenas o independientes), es decir, su finalidad es poder estimar el valor de Y; por lo que se conocen como modelos predictivos.
- **Explicativo:** se quiere cuantificar la relación entre X_1, X_2, \dots, X_n y la variable Y (variable endógena, explicativa o dependiente) con el fin de conocer o explicar mejor los mecanismos de esa relación. Se trata de modelos explicativos, muy utilizados cuando se busca encontrar qué variables afectan a los valores de otra variable.

Por lo tanto el objetivo de esta tesis se inscribe dentro de los objetivos explicativos, por tener como principal objetivo analizar el desempeño del PASDBP para identificar los factores que determinan su funcionamiento.

Dentro de los métodos de regresión existen diferentes variantes donde se trabaja con distintos tipos de variables. Así por ejemplo están la regresión lineal, la regresión logística, la regresión ordinal, entre otros.

En particular el método de regresión ordinal empleado en este trabajo, es adecuado cuando la variable dependiente es una variable no métrica¹¹⁸ y las variables independientes pueden ser métricas o no métricas. La variable dependiente debe cumplir que además de ser una variable no métrica, tenga la particularidad de que exista orden entre sus categorías.

Podemos ver el Modelo de Regresión Ordinal como la siguiente relación:

$$\begin{array}{ccc} Y & \leftarrow & X_1, X_2, X_3, \dots, X_{n-1}, X_n \\ \text{(No Métrica)} & & \text{(Métricas/ No Métricas)} \end{array}$$

una tendencia a tener hijos que heredaban parte de esta altura, pero que revelaban también una tendencia a regresar a la media. Galton generalizó esta tendencia bajo la "ley de la regresión universal": «Cada peculiaridad en un hombre es compartida por sus descendientes, pero en media, en un grado menor.»

¹¹⁷ La variable dependiente también puede llamarse variable criterio.

¹¹⁸ La variable dependiente no métrica debe ser codificada numéricamente.

Es decir, la Regresión Ordinal le da importancia a la existencia de orden entre las categorías de la variable dependiente. En el caso de la variable dependiente “Desempeño del PASDBP” se cumple el requerimiento de existir orden en sus cuatro categorías.

Considerando las características de las técnicas estadísticas de los diferentes tipos de regresión, así como el tipo de variable dependiente y de las variables independientes, se decidió aplicar el modelo de Regresión Ordinal para el análisis estadístico de los factores que determinan el desempeño del Programa de Acceso a Servicios Digitales en Bibliotecas Públicas porque se ajusta mejor a los datos y al objetivo que se plantea.

b) Regresión Ordinal¹¹⁹

La regresión ordinal permite dar forma a la dependencia de una respuesta ordinal politémica¹²⁰ sobre un conjunto de predictores, que pueden ser factores o covariables. El diseño de la regresión ordinal se basa en la metodología de McCullagh (1980, 1998).¹²¹

i. Modelo General

Para comenzar la descripción del planteamiento del modelo general, se le debe asignar un orden a las categorías de la variable dependiente ordinal.

Considere el ejemplo de una variable dependiente que califica el aprovechamiento de un servicio con cinco categorías: muy malo, malo, regular, bueno y muy bueno. De modo que el orden de estas variables es el siguiente:

$$\text{muy malo} < \text{malo} < \text{regular} < \text{bueno} < \text{muy bueno}$$

En donde, se le asigna como $Y_1 = \text{muy malo}$, $Y_2 = \text{malo}$, $Y_3 = \text{regular}$, $Y_4 = \text{bueno}$, $Y_5 = \text{muy bueno}$.

¹¹⁹ Ordinal Regression Chapter 4, [en línea] USA Consultado el 12 de abril de 2015 en http://www.norusis.com/pdf/ASPC_v13.pdf

Jose Barrera (2009, 20 de mayo). Introducción a la Regresión Ordinal, [en línea] Consultado el 30 de abril de 2012 en http://www.mat.uab.cat/~jbarrera/Teaching_files/Regresion_Ordinal.pdf IBM. Página oficial [en línea] Consultado el 12 de abril de 2015 en <http://www.ibm.com/us/en/> SPSS Inc. (2003). SPSS Advanced Models 12.0 [en línea] Chicago, USA Consultado el 12 de abril de 2015 en <http://www.docs.is.ed.ac.uk/skills/documents/3639/SPSSAdvancedModels12.0.pdf>

¹²⁰ Una variable cualitativa ordinal politémica es una variable que puede tomar tres o más valores posibles, los cuales pueden ser ordenados siguiendo un criterio establecido por una escala ordinal, la cual se caracteriza porque no es preciso que el intervalo entre mediciones consecutivas sea uniforme.

¹²¹ Para su cálculo en el programa SPSS la sintaxis se hace referencia al procedimiento como PLUM, que es una extensión del Modelo Lineal General para datos categóricos ordinales.

Como se observa en el ejemplo anterior la categoría 1 se asigna al valor más bajo o negativo y 5 a la categoría más alta o positiva. Generalizando esto se obtiene que la variable dependiente Y con n categorías, donde $n \geq 3$, obtiene el siguiente orden en sus categorías (orden ascendente):

$$Y_1 < Y_2 < Y_3 < \dots < Y_{n-1} < Y_n$$

También se define el vector X como el conjunto de m variables independientes:

$$X = (X_1, X_2, X_3, \dots, X_{m-1}, X_m)$$

Las variables independientes pueden ser variables factores o covariables. Las primeras deben ser variables categóricas (no métricas), y las covariables están definidas como variables numéricas (métricas).

Ahora bien, para comenzar el planteamiento del modelo general se debe definir el evento de interés. Para la Regresión Ordinal este evento es la observación de un resultado en particular o de aquéllos de menor orden, o visto de otra forma, se obtiene la categoría j o las categorías menores a ella. Entonces se tienen las siguientes probabilidades:

$$\theta_1 = P(Y \leq Y_1 | X) = P(Y = Y_1 | X)^{122}$$

$$\theta_2 = P(Y \leq Y_2 | X)$$

$$\theta_3 = P(Y \leq Y_3 | X)$$

:

$$\theta_n = P(Y \leq Y_n | X) = 1^{123}$$

En general,

$$\theta_j = P(Y \leq Y_j | X)$$

Con $j = 1, 2, 3, \dots, n$

A esta probabilidad se le conoce como probabilidad acumulada debido a que se obtiene la probabilidad de que se dé Y_j más las probabilidades de que ocurra Y_1, Y_2, \dots, Y_{j-1} .

Si se desea obtener la probabilidad absoluta, es decir, la probabilidad de que únicamente ocurra un evento en particular, se deben realizar los siguientes cálculos:

$$\tau_1 = P(Y = Y_1 | X)$$

$$\tau_2 = P(Y = Y_2 | X) = P(Y \leq Y_2 | X) - P(Y \leq Y_1 | X)$$

¹²² Esta igualdad se debe a que Y_1 siempre es la primera categoría, es decir, la categoría de orden menor.

¹²³ La probabilidad de $\theta_n = 1$, esto ocurre naturalmente porque es equivalente a que ocurra cualquier resultado de la variable dependiente.

$$\tau_3 = P(Y = Y_3 | X) = P(Y \leq Y_3 | X) - P(Y \leq Y_2 | X)$$

:

$$\tau_n = P(Y = Y_n | X) = P(Y \leq Y_n | X) - P(Y \leq Y_{n-1} | X)$$

En general,

$$\tau_j = P(Y = Y_j | X) = P(Y \leq Y_j | X) - P(Y \leq Y_{j-1} | X)$$

Con $j = 2, 3, \dots, n$

En realidad se desea obtener, para cada una de las categorías de la variable dependiente (menos la más alta) Y_1, Y_2, \dots, Y_{n-1} , un modelo que estime la probabilidad de ocurrencia de dicha categoría dado un vector $X = (X_1, X_2, X_3, \dots, X_{m-1}, X_m)$, mediante una ecuación lineal que pondere en qué medida influye cada variable independiente y un umbral referente a la categoría.

Entonces se tiene que el predictor lineal que se busca es:

$$f(\theta_j(X)) = f(P(Y \leq y_i | X)) = \alpha_j + \beta'X$$

Con $j = 2, 3, \dots, n - 1$. Donde α_j ¹²⁴ y $\beta = (\beta_1, \beta_2, \dots, \beta_m)'$ son los parámetros a estimar.

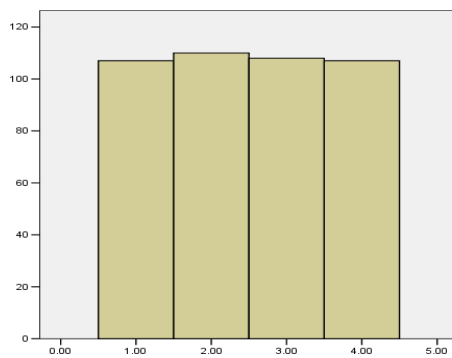
Como se puede notar, aún falta conocer la función con la que, a partir de las probabilidades anteriormente definidas (probabilidades acumuladas), se obtienen los parámetros β s y α s para la construcción del modelo de regresión ordinal.

Esta función f se conoce como función liga, y es una transformación de las probabilidades acumuladas que permiten la estimación del modelo. También se puede definir como una función que describe la relación funcional entre la función de las variables explicativas que se usa como predictor y el valor esperado de la distribución de probabilidad de la variable de respuesta. Las funciones ligas que pueden ser utilizadas para obtener el modelo son las siguientes:

¹²⁴ Se refiere a los umbrales de localización.

Función	Fórmula	Aplicación típica
Logit ¹²⁵	$Log\left(\frac{\xi}{1-\xi}\right)$	Cuando las categorías se encuentran distribuidas de forma uniforme
Log-log	$Log(-Log(1-\xi))$	Donde las categorías más altas son más probables
Log-log negativo	$-Log(-Log(\xi))$	Donde las categorías más bajas son más probables
Probit	$\phi^{-1}(\xi)$	La variable latente ¹²⁶ sigue una distribución normal

Con el análisis exploratorio que previamente se realizó a la variable dependiente, se elige la función liga que mejor se adecue a la distribución. Por lo tanto se trabajó con la función liga **logit** debido a que se encontró que la respuesta está uniformemente representada. Enseguida se muestran el histograma de la variable respuesta y la tabla con sus frecuencias correspondientes, para corroborar la elección de la función liga:



En el histograma no se observan picos, por el contrario, dada la forma en que se categorizó el índice todas las categorías tienen un número de observaciones similar, es decir, se asemeja a una distribución donde todas las categorías comparten la misma probabilidad de ocurrencia.

Categoría	Frecuencia
Desempeño Bajo	107
Desempeño Medio	110
Desempeño Alto	108
Desempeño Excelente	107

¹²⁵ La función liga *logit* generalmente es asociada al método de Regresión Logística, sin embargo, su uso se extiende más allá de este método estadístico.

¹²⁶ En estadística las variables latentes son aquellas que no son directamente observables, pero que pueden ser inferidas (a través de un modelo matemático) de otras variables que sí son observables.

Como se observa en la tabla, las frecuencias correspondientes a cada una de las categorías en la variable “Desempeño del PASDBP” son muy parecidas, al presentar como máximo 3 observaciones de diferencia entre categorías (diferencia mínima, tomando en cuenta que el total es de 432 observaciones y solamente hay 4 categorías). Por lo tanto se concluye que la distribución de la variable de respuesta se asemeja a la distribución uniforme.

Entonces para la función *Logit*, donde $\xi = P(Y \leq Y_j | X)$, se obtiene que:

$$f(\theta_j(X)) = f(P(Y \leq y_i | X)) = \text{Log} \left[\frac{P(Y \leq Y_j | X)}{1 - P(Y \leq Y_j | X)} \right] = \alpha_j + \beta'X$$

A partir de esta relación, se realizan los siguientes cálculos para obtener los parámetros α_j y $\beta = (\beta_1, \beta_2, \dots, \beta_m)'$:

$$\left[\frac{P(Y \leq Y_j | X)}{1 - P(Y \leq Y_j | X)} \right] = e^{\alpha_j + \beta'X} \rightarrow P(Y \leq Y_j | X) = e^{\alpha_j + \beta'X} [1 - P(Y \leq Y_j | X)]$$

$$\rightarrow P(Y \leq Y_j | X) = e^{\alpha_j + \beta'X} - e^{\alpha_j + \beta'X} [P(Y \leq Y_j | X)]$$

$$\rightarrow P(Y \leq Y_j | X) + e^{\alpha_j + \beta'X} [P(Y \leq Y_j | X)] = e^{\alpha_j + \beta'X}$$

$$\rightarrow P(Y \leq Y_j | X) [1 + e^{\alpha_j + \beta'X}] = e^{\alpha_j + \beta'X} \rightarrow P(Y \leq Y_j | X) = \frac{e^{\alpha_j + \beta'X}}{1 + e^{\alpha_j + \beta'X}}$$

$$\rightarrow P(Y \leq Y_j | X) = \frac{1}{\frac{1 + e^{\alpha_j + \beta'X}}{e^{\alpha_j + \beta'X}}} \rightarrow P(Y \leq Y_j | X) = \frac{1}{e^{-(\alpha_j + \beta'X)} + 1}$$

Finalmente se obtiene un modelo de regresión ordinal que proporciona las probabilidades para cada una de las categorías de la variable dependiente (nótese que se manejan probabilidades acumuladas y probabilidades absolutas):

$$P(Y \leq Y_j | X) = \left[\frac{1}{1 + e^{-(\alpha_j + \beta'X)}} \right]$$

$$P(Y = Y_j | X) = \left[\frac{1}{1 + e^{-(\alpha_j + \beta'X)}} \right] - \left[\frac{1}{1 + e^{-(\alpha_{j-1} + \beta'X)}} \right]$$

Un detalle importante del modelo de regresión ordinal, es la inclusión de diferentes parámetros α_j para $j = 2, 3, \dots, n - 1$. La no inclusión de diferentes umbrales de α_j implicaría que:

$$P(Y \leq Y_j | X) = P(Y \leq Y_k | X), \quad \text{con } j \neq k$$

Y por lo tanto:

$$P(Y = Y_j | X) = \left[\frac{1}{1 + e^{-(\alpha_j + \beta'X)}} \right] - \left[\frac{1}{1 + e^{-(\alpha_{j-1} + \beta'X)}} \right] = 0$$

Debido a que los $\beta = (\beta_1, \beta_2, \dots, \beta_m)'$ se definen como el mismo conjunto en cada una de las categorías de la variable dependiente.

ii. Supuestos

Un paso previo a la estimación de los parámetros es revisar el cumplimiento de los supuestos del modelo, ya que de ello depende que tenga sentido y validez los resultados de un método estadístico.

Para poder realizar el método estadístico de regresión ordinal se requiere satisfacer supuestos en las variables y su distribución, que a continuación se especifican:

- Sólo se debe de tener una variable de respuesta Y (variable dependiente), la cual debe estar definida como una variable con escala de medida ordinal $Y_1 < Y_2 < Y_3 < \dots < Y_{n-1} < Y_n$ donde Y_j es la j -ésima categoría de Y con $j = 2, 3, \dots, n$.
- Para cada patrón distinto de valores en las variables independientes $X_1, X_2, X_3, \dots, X_{m-1}, X_m$, se supone que las respuestas son variables multinomiales independientes.
- Se supone una condición sobre los datos a modelar: odds o momios¹²⁷ proporcionales o líneas paralelas. Donde se asume que la relación que existe entre las variables independientes y las categorías es la misma para cada una de las categorías de la variable dependiente. Esto se prueba con la *Prueba de líneas paralelas*¹²⁸, la cual es una prueba correspondiente a la hipótesis de que los parámetros de localización son equivalentes en todos los niveles de la variable dependiente.

Se considera la posibilidad de un modelo alternativo para el modelo acumulativo, especificado al permitir diferentes parámetros de regresión

$\beta^j = (\beta_1^j, \beta_2^j, \dots, \beta_m^j)'$ con $j = 2, 3, \dots, n$ en las primeras $n - 1$ categorías de respuesta:

$$f(\theta_j(X)) = f(P(Y \leq Y_j | X)) = \alpha_j + \beta^j X$$

Es decir, que cada categoría de la variable dependiente tendrá valores diferentes en β' s.

$$f(\theta_1(X)) = f(P(Y \leq Y_1 | X)) = \alpha_1 + \beta^1 X$$

$$f(\theta_2(X)) = f(P(Y \leq Y_2 | X)) = \alpha_2 + \beta^2 X$$

:

¹²⁷ En estadística la odds ratio (término en inglés de traducción discutida; se ha traducido como disparidad, razón de posibilidades, razón de oportunidades, razón de momios, razón de odds) es el cociente de dos razones: el numerador es la razón del riesgo de que un evento suceda y el riesgo de que no suceda bajo ciertas condiciones y el denominador es la razón del riesgo de que dicho evento suceda y el riesgo de que no suceda bajo las condiciones complementarias. Por lo que es una medida de tamaño de efecto.

¹²⁸ «Fuentes electrónicas»

http://publib.boulder.ibm.com/infocenter/spsstat/v20r0m0/index.jsp?topic=%2Fcom.ibm.spss.statistics.help%2Falg_csordinal_testparallel.htm (Consulta 30/04/2015)

$$f(\theta_{n-1}(X)) = f(P(Y \leq Y_{n-1} | X)) = \alpha_{n-1} + \beta^{n-1}'X$$

El modelo alternativo contiene los parámetros con valores de α_j y parámetros de regresión. El modelo acumulativo es una restricción del modelo alternativo, basado en el supuesto de líneas paralelas, que corresponde a decir que:

$$\begin{aligned} f(P(Y \leq Y_1 | X)) &= \alpha_1 + \beta'X \\ f(P(Y \leq Y_2 | X)) &= \alpha_2 + \beta'X \\ &\vdots \\ f(P(Y \leq Y_{n-1} | X)) &= \alpha_{n-1} + \beta'X \end{aligned}$$

Lo que nos lleva a que las hipótesis de la Prueba de Líneas Paralelas son:

$$H_0: \beta^1 = \beta^2 = \dots = \beta^{n-1} \text{ vs } H_a: \beta^i \neq \beta^j \text{ para algún } i \neq j$$

Se lleva a cabo esta prueba mediante la estimación de los parámetros del modelo alternativo y la aplicación de una Prueba tipo Wald para $L\beta_a = 0$ con la matriz L de contraste dada por:

$$L = \begin{bmatrix} L^1 \\ \vdots \\ L^m \end{bmatrix} \text{ donde cada } L^t, \text{ con } t = 1, 2, \dots, n \text{ es una matriz de } (n-2)(m+1)(n-1) \text{ que}$$

contiene contrastes pairwise¹²⁹ para el parámetro t entre el primero y el resto de las ecuaciones de regresión para las repuestas correspondientes.

Entonces, si se rechaza la hipótesis nula se viola el supuesto de momios proporcionales y se concluye que el modelo de regresión ordinal no es adecuado. De lo contrario, se considera el cumplimiento del supuesto de Líneas Paralelas y se asume que la relación que existe entre las variables explicativas y las categorías de la variable de respuesta es la misma en cada una de ellas.

Como se puede percibir, el método de Regresión Ordinal es una técnica estadística flexible ya que requiere el cumplimiento de pocos supuestos¹³⁰. Pero no sólo el número de supuestos lo convierte en un método más adaptable, sino que moderada la rigidez de ellos y permite que sea una técnica apropiada para aplicarse a problemas de diversa naturaleza (social, económica, médica, etc.).

Si bien, se ha remarcado las ventajas referentes a los supuestos de la Regresión Ordinal, también es importante reconocer cuáles son las limitaciones que presenta el modelo, para tomar

¹²⁹ Pairwise es definido en estadística como un tratamiento a pares.

¹³⁰ A modo de comparación con otras técnicas estadísticas, podemos citar los supuestos de la Regresión Lineal. Está requiere el cumplimiento de normalidad, incorrelación y homocedasticidad en los residuos. Además el tipo de escala de medida de sus variables independientes, restringe de manera considerable su utilización en fenómenos donde las variables explicativas no pueden ser medidas de forma numérica en su mayoría (recordemos que las variables ficticias juegan más bien un papel de contraste), o la misma variable respuesta no puede ser métrica.

decisiones concernientes a las variables a seleccionar y así evitar problemas en la estimación de los parámetros. A continuación se mencionan dichas limitaciones:

- Al utilizar más de una variable independiente numérica (covariable) se puede llegar a crear una tabla de contingencia de gran tamaño, lo que provoca que un número considerable de casillas contengan pocos casos o ninguno. Esto genera problemas en el cálculo de los estadísticos de prueba¹³¹, necesarios en el contraste de hipótesis.
- Al utilizar un gran número de variables independientes se llega a la creación de tablas de probabilidades de casillas extensas, lo que conlleva al mismo inconveniente de utilizar más de una covariable.

iii. Estimación del modelo

Después del planteamiento de un modelo general y citar los supuestos que debe satisfacer el modelo estadístico, el siguiente paso que concierne al procedimiento es la estimación del modelo. En realidad, esta parte se refiere a la estimación de los parámetros β s y α s para presentar el predictor lineal que modele las probabilidades deseadas.

Para ello la técnica de Regresión Ordinal utiliza el método de Máxima Verosimilitud para encontrar las β s y α s adecuadas para la estimación del modelo. A continuación se presenta la teoría del método en cuestión.

Estimación del modelo por Máxima Verosimilitud

Supongamos que la distribución de X depende de un parámetro λ , siendo X y λ vectores. Por simplicidad, asumiremos que λ es un único parámetro. Entonces representamos la función de densidad de X en x_i como $f(x_i|\lambda)$. Si tenemos n observaciones independientes e idénticamente distribuidas, el producto de las densidades individuales es igual al producto de la distribución conjunta.

$$f(x_1, x_2, x_3, \dots, x_m|\lambda) = f(x_1|\lambda)f(x_2|\lambda)f(x_3|\lambda) \dots f(x_m|\lambda)$$

Y definimos

$$\prod_{i=1}^m f(x_i|\lambda) = L(\lambda|x_1, x_2, x_3, \dots, x_m)$$

El último término, $L(\lambda|x_1, x_2, x_3, \dots, x_m)$, es la función de verosimilitud que mide qué tan factible es observar la muestra obtenida¹³². Esta función de verosimilitud tiene los parámetros en función de los datos, a diferencia de la función de densidad que tiene los datos en función de los parámetros.

¹³¹ Por ejemplo, la Ji-Cuadrada generalmente requiere que cada casilla de la tabla de contingencia contenga $n \geq 5$.

¹³² El principio de máxima verosimilitud parte del supuesto de que la muestra observada es la muestra con la mayor probabilidad de ser obtenida.

Estimadores de Máxima Verosimilitud

Ahora que se ha definido la función de máxima verosimilitud, se quiere encontrar el valor del parámetro λ que hace máxima la función $L(\lambda|x_1, x_2, x_3, \dots, x_m)$. Es decir, dados los datos observados se quiere encontrar el parámetro que maximiza la probabilidad de que la muestra haya sido obtenida.

Para cada muestra $(x_1, x_2, x_3, \dots, x_m)$ de X , el estimador de máxima verosimilitud $\hat{\lambda}$ de λ es aquél que hace máxima la función de $L(\lambda|x_1, x_2, x_3, \dots, x_m)$:

$$L(\hat{\lambda}|x_1, x_2, x_3, \dots, x_m) = \max L(\lambda|x_1, x_2, x_3, \dots, x_m)$$

Intuitivamente $\hat{\lambda}$ es el valor del parámetro que hace más verosímil la muestra observada. Los estimadores de máxima verosimilitud son muy buenos estimadores y tienen las siguientes propiedades (Manuel Barrón, 2010): consistencia¹³³, normalidad asintótica¹³⁴, invarianza y eficiencia asintótica¹³⁵.

Cálculo del estimador máximo verosímil

Como se había mencionado, λ puede ser un vector y no sólo un único parámetro, por lo que $\lambda = (\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_h)$. Cabe mencionar que en la práctica el cálculo del estimador máximo verosímil generalmente se trabaja con el logaritmo de la función de verosimilitud¹³⁶, el cual será llamado función de log-verosimilitud.

$$\ln L(\lambda|x_1, x_2, x_3, \dots, x_m) = \ln \left(\prod_{i=1}^m f(x_i|\lambda) \right) = \sum_{i=1}^m \ln f(x_i|\lambda)$$

Entonces, si la función $\ln L(\lambda|x)$ es derivable en λ_j , las soluciones de las ecuaciones que generan las derivadas de primer orden,

$$\frac{\partial}{\partial \lambda_j} \ln L(\lambda|x) = 0, \quad j = 1, 2, \dots, h$$

Son candidatas a ser el estimador máximo verosímil. Los ceros de las primeras derivadas no son necesariamente máximos de la función de verosimilitud ya que pueden ser máximos locales o globales, mínimos locales o globales o puntos de inflexión. Para asegurar que la solución encontrada

¹³³ El valor esperado de la función de verosimilitud es maximizada en el valor verdadero de los parámetros.

¹³⁴ En muestras grandes ($n \rightarrow \infty$), el estimador de Máxima Verosimilitud sigue una distribución normal.

¹³⁵ Dentro de la clase de estimadores consistentes, los estimadores de Máxima Verosimilitud tienen la menor varianza.

¹³⁶ Dado que el logaritmo natural es una función estrictamente creciente, el valor máximo de $L(\lambda|x)$ si existe, estará en el mismo punto que el máximo de la función $\ln L(\lambda|x)$. Además, esto permite transformar los productos en suma, lo que resulta en una función más manejable.

es un máximo (local o global) hay que comprobar que las segundas derivadas sean definidas negativas.

El método de máxima verosimilitud que se utiliza para la obtención de $\beta = (\beta_1, \beta_2, \beta_3, \dots, \beta_m)$ y $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$ y, busca obtener un estimador máximo verosímil de un parámetro multidimensional, y se denota de la siguiente forma:

$$\begin{aligned} L(\alpha, \beta | Y, X) &= \prod_{i=1}^m \prod_{j=2}^{n-1} [P(Y \leq Y_1 | X_i)]^{\delta_{i1}} [P(Y \leq Y_j | X_i) - P(Y \leq Y_{j-1} | X_i)]^{\delta_{ij}} \\ &= \prod_{i=1}^m \prod_{j=2}^{n-1} [P(Y \leq Y_1 | X_i)]^{\delta_{i1}} [P(Y = Y_j | X_i)]^{\delta_{ij}} \\ &= \prod_{i=1}^m \prod_{j=2}^{n-1} \left[\frac{1}{1+e^{-(\alpha_1 + \beta' X_i)}} \right]^{\delta_{i1}} \left[\frac{1}{1+e^{-(\alpha_j + \beta' X_i)}} - \frac{1}{1+e^{-(\alpha_{j-1} + \beta' X_i)}} \right]^{\delta_{ij}} \end{aligned}$$

Donde $\delta_{ij} = \begin{cases} 1, & \text{si la } i\text{-ésima observación de la muestra presenta } Y = Y_j \\ 0, & \text{en otro caso} \end{cases}$

Prueba de Wald

Para evaluar la significancia estadística de los parámetros estimados, y con ello decidir si forman parte del modelo de regresión ordinal, se utiliza la Prueba de Wald que consiste en dividir el cuadrado de $\widehat{\beta}_i$ por el cuadrado del error estándar de $\widehat{\beta}_i$. Y viene dado por la siguiente expresión:

$$Wald = \frac{(\widehat{\beta}_i)^2}{(EE \widehat{\beta}_i)^2}$$

El estadístico Wald se distribuye según una $\chi^2_{(1)}$. Y las hipótesis son:

$$H_0: \beta_i = 0 \quad vs \quad H_1: \beta_i \neq 0$$

Es decir, si el valor de Wald es tal que la probabilidad de obtenerlo por azar, bajo los supuestos de la hipótesis nula, es menor que el nivel de significancia fijado en el contraste, se rechaza la hipótesis nula y se concluye que el valor de β_i es estadísticamente significativo diferente de cero. En caso contrario, se determina que $\beta_i = 0$.

Es importante precisar que la Prueba de Wald anteriormente definida es adecuada para el contraste de un modelo ordinal con una sola variable independiente, o bien el contraste para cada β_i de manera individual.

En caso de presentarse un modelo de Regresión Ordinal Múltiple, debe emplear la Prueba de Wald que se define de la siguiente forma:

$$Wald = \hat{\beta}' [Var(\hat{\beta})]^{-1} \hat{\beta} \quad \text{donde } \hat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2, \dots, \widehat{\beta}_m)'$$

El estadístico se distribuye $\chi^2_{(k)}$ con k es el número de contrastes. A continuación se presentan las hipótesis:

$$H_0: \beta_1 = 0, \beta_2 = 0, \dots, \beta_m = 0 \quad vs \quad H_1: \text{al menos una } \beta_i \neq 0$$

Si se rechaza la hipótesis nula, entonces el valor de $\beta = (\beta_1, \beta_2, \dots, \beta_m)$ es estadísticamente diferente de cero, es decir, al menos una β_i es diferente de cero.

iv. Bondad de ajuste del modelo

Una prueba de bondad de ajuste se emplea para decidir cuándo un conjunto de datos se apega a una distribución de probabilidad dada, es decir, se interesa en ver hasta qué punto es razonable suponer que los datos provienen de una cierta distribución.

Dada una muestra X_1, X_2, \dots, X_m de v.i.i.d.¹³⁷ con distribución F , un problema básico en estadística es encontrar un modelo para los datos, y llamaremos a la distribución propuesta como F_0 . Por lo que se definen las siguientes hipótesis:

$$H_0: F(x) = F_0(x) \quad vs \quad H_a: F(x) \neq F_0(x)$$

Para llevar a cabo la contrastación de las hipótesis anteriormente definidas se pueden utilizar diferentes pruebas, una que se ocupa en la Regresión Ordinal es:

Prueba χ^2 de Pearson¹³⁸

Es una prueba considerada como no paramétrica que mide la discrepancia entre una distribución observada y otra teórica. Indica en qué medida las diferencias entre ambas, si es que existen, se deben al azar en el contraste de hipótesis. Entonces se tienen N observaciones que se dividen en n celdas.

A partir de ello se propone el estadístico:

$$S = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

Donde:

n = número de celdas en la tabla

O_i = frecuencia observada en la i -ésima celda. Donde $\sum_{i=1}^n O_i = N$

E_i = frecuencia esperada (teórica) en la i -ésima celda.

¹³⁷ Variables independientes idénticamente distribuidas.

¹³⁸ La primera prueba de bondad de ajuste fue propuesta por Karl Pearson en el año de 1900. Pearson propuso evaluar el ajuste de una función de distribución F_0 a una muestra de v.i.i.d mediante el uso de un estadístico de tipo cuadrático. Este planteamiento constituye la primera evaluación rigurosa de la calidad del ajuste a una distribución.

La estadística S , tiene una distribución χ^2 con $(n - p)$ grados de libertad, donde $p = s + 1$; y s es el número de parámetros utilizados en el ajuste de la distribución. El criterio de decisión es que no se rechaza la hipótesis nula si $S < \chi^2_{(n-p)}$. Si existe una concordancia perfecta entre las frecuencias que se observan y las que se esperaban, la estadística tendrá un valor igual a cero. Por otro lado, si existe gran discrepancia entre las frecuencias observadas y las esperadas, el estadístico tomará un valor muy grande.

Devianza

En estadística, la devianza es una cualidad del estadístico de ajuste de un modelo que se utiliza frecuentemente. Por lo que se refiere a Devianza, como una medida de bondad de ajuste que tiene el siguiente estadístico de prueba:

$$D = 2 \sum_{i=1}^n O_i \left[\ln \left(\frac{O_i}{E_i} \right) \right]$$

Donde:

n = número de celdas en la tabla

O_i = frecuencia observada en la i -ésima celda. Donde $\sum_{i=1}^n O_i = N$

E_i = frecuencia esperada (teórica) en la i -ésima celda.

La cual tiene una χ^2 con $(n - p)$ grados de libertad, con los mismos criterios para p que en la prueba anterior "Prueba χ^2 de Pearson". Y de forma análoga se determina que no se rechaza la hipótesis $H_0: F(x) = F_0(x)$ si $D < \chi^2_{(n-p)}$.

Entonces, la devianza es una prueba que mide hasta qué punto un modelo se ajusta bien a los datos y cuanto más pequeño sea el valor obtenido para D mejor el ajuste.

Cociente de Verosimilitudes

Para probar la significancia del modelo se realiza la Prueba de Máxima Verosimilitud, también llamado la Prueba del Cociente de Verosimilitudes. Las hipótesis para realizar la prueba son:

$$H_0: \beta_1 = 0, \beta_2 = 0, \dots, \beta_m = 0 \quad vs \quad H_a: \text{al menos una } \beta_j \neq 0$$

Es decir, la hipótesis nula establece que los coeficientes son iguales a cero, mientras que la hipótesis alternativa dice que al menos uno de los coeficientes es diferente de cero.

Para llevar a cabo la prueba se utiliza el siguiente estadístico de prueba:

$$C = -2 \log \left[\frac{\text{Verosimilitud Modelo Nulo}^{139}}{\text{Verosimilitud Modelo Propuesto}^{140}} \right]$$

¹³⁹ El Modelo Nulo se refiere a aquel donde no están incluidas las variables predictoras (β). $f(\theta_j(X)) = \alpha_j$

¹⁴⁰ El Modelo Propuesto se refiere a aquel donde participan las variables predictoras. $f(\theta_j(X)) = \alpha_j + \beta'X$

El estadístico C permite contrastar dos modelos, por ejemplo: dos modelos en el que en uno de ellos incluye una variable adicional (o varias, según sea el caso) con respecto al primer modelo. Las diferencias en la función de verosimilitud se alteran arbitrariamente con escala de medida, por lo que la forma adecuada de comparar es mediante cocientes.

Para la prueba se tiene una χ^2 con $(s - p)$ grados de libertad, donde

$s =$ suma de las categorías de las variables independientes y

$p =$ número de variables independientes.

Si $C > \chi^2_{(s-p)}$, se rechaza la hipótesis nula y se concluye que el modelo es significativo al encontrarse una mejora en la verosimilitud del modelo. Es importante el rechazo de la hipótesis nula para el modelo, porque justifica que alguna de las variables explicativas es importante para explicar el fenómeno en estudio.

Contraste de Máxima Verosimilitud.

Cuando se desea probar si la introducción de una o más variables mejora el modelo, se puede utilizar un método que compare la verosimilitud de los modelos para determinar si la incorporación de esta(s) variable(s) mejora sensiblemente la verosimilitud y por lo tanto el nuevo modelo ajusta mejor.

El estadístico de prueba se define como:

$$G = (-2\text{Log}[\text{Verosimilitud Modelo Anterior}]) - (-2\text{Log}[\text{Verosimilitud Modelo Nuevo}])$$

Donde:

Modelo Anterior= El modelo "sin nuevas variables predictivas"

Modelo Nuevo= El modelo "con nuevas variables predictivas"

Que se distribuye χ^2 con (r) grados de libertad donde $r =$ número de variables nuevas.

Entonces se rechaza la hipótesis nula si $G > \chi^2_{(r)}$ y se establece que la inclusión de la nuevas variable(s) mejora significativamente la verosimilitud; entonces el Modelo Nuevo ajusta mejor. En caso de que no sea rechazada la hipótesis, se determina que el Modelo Nuevo carece de utilidad y se queda el Modelo Anterior.

Entonces, gracias a este contraste de máxima verosimilitud se puede distinguir entre una mejora real y una aportación trivial, cuando son incorporadas nuevas variables al modelo.

Residuos de Pearson

Cuando se han calculado las probabilidades estimadas, se puede corroborar la bondad de ajuste del modelo con las frecuencias esperadas y observadas mediante el cálculo de los Residuos de Pearson.

El Residuo Pearson representa una diferencia estandarizada entre las frecuencias observadas y esperadas y se define como:

$$\text{Residuo de Pearson} = \frac{O_{ij} - E_{ij}}{\sqrt{n_i \widehat{P}_{ij} (1 - \widehat{P}_{ij})}}$$

Donde:

O_{ij} = Frecuencias observadas en la celda de la fila i y la columna j

E_{ij} = Frecuencias esperadas en la celda de la fila i y la columna j

n_i = Frecuencias de la fila i

\widehat{P}_{ij} = Probabilidad estimada en la celda de la fila i y la columna j

Si el *Residuo Pearson* = 0, el ajuste del modelo es perfecto. Entre más cercano a cero sea el Residuo de Pearson mejor ajuste presenta el modelo.

Cuando *Residuo Pearson* > 0 significa que las frecuencias esperadas resultaron menores que las observadas y de forma inversa, cuando *Residuo Pearson* < 0 significa que las frecuencias esperadas fueron mayores a las observadas.

P-seudo R^2

En este tipo de modelos no se emplea la R^2 para mostrar la bondad del ajuste, sino que se utilizan las conocidas como *pseudo R^2* . Reciben este nombre, porque no van a tener el significado que tiene la R^2 en regresión lineal, sin embargo, las *pseudo- R^2* son estadísticos que pretenden simular la función que tiene R^2 .

Estas *pseudo R^2* miden el ajuste usando la función de verosimilitud, para ello calculan el incremento de la verosimilitud. En otras palabras, se compara el logaritmo de la verosimilitud del modelo nulo (sin las variables) con el del modelo propuesto (con todas las variables).

Dado que el empleo de dichos coeficientes es un asunto controversial, resulta necesario complementar la evaluación del modelo de regresión ordinal con otras pruebas como las ya mencionadas Pearson, Devianza, Cociente de verosimilitudes y Contraste de máxima verosimilitud para poder interpretarlo correctamente.

Los siguientes son métodos que se utilizan para realizar una estimación del coeficiente.

El primero de ellos es la *R^2 de Cox y Snell*. Es un coeficiente de determinación generalizado, esta Pseudo- R^2 se basa en la comparación del log de la verosimilitud para el modelo propuesto respecto al log de la verosimilitud para un modelo nulo.

$$R_{CS}^2 = 1 - \left[\frac{L(B^{(0)})}{L(\hat{B})} \right]^{\frac{2}{N}}$$

Donde:

$L(B^{(0)})$ = Log de la verosimilitud para el modelo nulo

$L(\hat{B})$ = Log de la verosimilitud para el modelo propuesto

N = número de casos

Este coeficiente esta acotado de la siguiente forma: $0 \leq R_{CS}^2 < 1$

Por otra parte está el R^2 de Nagelkerke se define como una versión mejorada del R^2 de Cox y Snell, ya que corrige la escala del estadístico para cubrir el rango completo de 0 a 1.

$$R_N^2 = \frac{1 - \left[\frac{L(B^{(0)})}{L(\hat{B})} \right]^{\frac{2}{N}}}{R_{Max}^2}$$

Donde:

$L(B^{(0)})$ = Log de la verosimilitud para el modelo nulo

$L(\hat{B})$ = Log de la verosimilitud para el modelo propuesto

N = número de casos

Donde $R_{Max}^2 = 1 - \left[L(B^{(0)}) \right]^{\frac{2}{N}}$ para así poder alcanzar el valor 1.

Por último el coeficiente R^2 de McFadden es el siguiente:

$$R_{MF}^2 = 1 - \left[\frac{L(\hat{B})}{L(B^{(0)})} \right]$$

Donde:

$L(B^{(0)})$ = Log de la verosimilitud para el modelo sin nulo

$L(\hat{B})$ = Log de la verosimilitud para el modelo propuesto

N = número de casos

Suele considerarse de buena calidad el ajuste cuando $0.2 \leq R_{MF}^2 \leq 0.4$ ¹⁴¹ y se considera excelente cuando $R_{MF}^2 > 0.4$, ya que raramente se aproxima a 1.

La Pseudo – R^2 de interpretación más directa es la R^2 de Nagelkerke, que puede tomar valores entre 0 y 1. En primer lugar, debe recordarse que la R^2 de Nagelkerke es una versión

¹⁴¹ Fernández Pando, Martín Fernández(2004). *Regresión Logística Multinomial*, [PDF] España.

mejorada de la R^2 de *Cox and Snell*, porque permite que el intervalo llegue a 1 al corregirse la escala del estadístico. Con respecto a la R^2 de *McFadden*, se considera mejor la R^2 de *Nagelkerke* (al menos para fines de esta tesis) porque toma en cuenta el número de casos que involucra el modelo, mientras que la R^2 de *McFadden* no. Entonces se tomará la R^2 de *Nagelkerke* como la medida formal del ajuste del modelo (usando la función de verosimilitud).

Los factores que constituyen un “buen” valor de *pseudo* R^2 varían entre las distintas áreas de aplicación. Mientras que estos estadísticos pueden ser indicativos por sí solos, son más útiles para comparar modelos que compiten con los mismos datos, es decir, para seleccionar el modelo más adecuado.

c) Modelo de regresión ordinal aplicado al desempeño del PASDBP

Con el fin de facilitar la interpretación del procedimiento de la construcción del modelo de regresión ordinal para el desempeño del PASDBP, a continuación, se describen el procedimiento que se siguió y las pruebas que se aplicaron¹⁴².

- **PRIMERO.** Se define la variable dependiente como $Y = \text{Desempeño del PASDBP}$ con $n = 4$ categorías y se obtiene el siguiente orden en sus categorías (orden ascendente):

$$Y_1 = \text{bajo} < Y_2 = \text{medio} < Y_3 = \text{alto} < Y_4 = \text{excelente}$$

También se define X como el conjunto de m variables independientes, como un vector:

$$X = (X_1, X_2, X_3, \dots, X_m)$$

El predictor lineal que buscamos obtener es:

$$f(\theta_j(X)) = f(P(Y \leq y_i|X)) = \alpha_j + \beta'X$$

Con $j = 1,2,3,4$ Donde α_j y $\beta = (\beta_1, \beta_2, \dots, \beta_m)'$ son los parámetros a estimar.

¹⁴² De las pruebas presentadas en el apartado anterior, se decidió no incluir en el procedimiento de análisis a los Residuales de Pearson, por considerarse de poca ayuda en el análisis estadístico que se desea realizar en esta tesis. Esto debido a que no tiene un criterio concreto de cuándo se está bien ajustado o no (ya que no proporciona un valor específico para determinar el ajuste más que el perfecto, si Residual de Pearson = 0), sino más bien proporciona una noción de discrepancia entre las frecuencias esperadas y observadas. Por otro lado, al tener un número grande de variables independientes (y por lo tanto de categorías) el tamaño de la tabla donde se presenta la información representa un problema.

- **SEGUNDO.** Se realizará la *Prueba de Líneas Paralelas*, para verificar el cumplimiento del supuesto más importante sobre la estructura de los datos. Si no se rechaza la hipótesis nula $H_0: \beta^1 = \beta^2 = \dots = \beta^{n-1}$, se encuentra que los datos son adecuados para llevar a cabo este método estadístico. (En caso contrario, de rechazar la hipótesis nula, se considera que no es conveniente optar por este método para modelar los datos)
- **TERCERO.** En seguida se procederá a aplicar dos pruebas para determinar si tiene sentido el modelo y realizar comparaciones con modelos anteriores propuestos para observar si mejora la verosimilitud. Primero se realiza la *Prueba de Máxima Verosimilitud*, donde se establece si tiene validez la estimación de los parámetros para construir el modelo de regresión ordinal. Si se rechaza la hipótesis $H_0: \beta_1 = 0, \beta_2 = 0, \dots, \beta_m = 0$, se concluye que el modelo tiene sentido.
- **CUARTO.** Si se tiene evidencia que es conveniente realizar el cálculo del modelo, se aplica el *Contraste de Máxima Verosimilitud* para comparar con modelos anteriores (con menos variables independientes) y elegir el más verosímil para calcularlo. Si se rechaza la hipótesis nula, quiere decir que la inclusión de nuevas variables mejora el modelo y por tanto éste es el que posteriormente debe calcularse.
- **QUINTO.** Ya que se ha decidido qué modelo es mejor, se estima por el método de Máxima Verosimilitud los valores correspondientes a los coeficientes de cada una de las categorías de las variables independientes y de los umbrales (constantes) de la variable dependiente. Pero antes de incorporarlas al modelo, se realiza el *Prueba de Wald* para determinar si son o no significantes para el modelo; en caso de resultar ser significativos se incorporan al modelo, de lo contrario se les da el valor de cero.
- **SEXTO.** El siguiente paso es calcular los predictores lineales para cada una de las categorías de la variable respuesta, posteriormente se calculan las probabilidades condicionales acumuladas y a partir de ellas se obtienen las probabilidades condicionales absolutas. En caso de ser un modelo con muy pocas variables (una o dos tal vez) se presenta un cuadro con las probabilidades obtenidas, en el cual no sólo se resume la información sino que provee de una herramienta de análisis muy sencilla y clara (principalmente se quiere dar interpretación a las probabilidades absolutas haciendo referencia a las probabilidades condicionales más altas o bajas para analizar cómo las caracterizan las variables independientes).
- **SÉPTIMO.** A continuación, se realizan dos pruebas para comprobar que el modelo tenga un buen ajuste.
- **OCTAVO.** Por último, se calcula la *Pseudo – R² de Nagelkerke* para poder valorar el ajuste del modelo usando la función de verosimilitud.

Con ello concluye el marco teórico del procedimiento de Regresión Ordinal y la descripción de los procedimientos para su aplicación a la problemática de esta tesis.

CAPÍTULO 4. Análisis estadístico múltiple de los determinantes del desempeño del PASDBP

Finalmente, después de explicar el marco teórico y contextual en que se fundamenta esta tesis, se presenta la construcción del modelo que identifique los principales factores que afectan el desempeño del Programa de Acceso a Servicios Digitales en Bibliotecas Públicas.

A continuación se presenta la construcción de los modelos así como su análisis para entender cuáles son los factores determinantes de cada uno de los grupos analizados (características de la localidad, del MSD, del encargado de módulo y de difusión), así como el análisis de cómo se relacionan en conjunto con el desempeño del programa. Y por último se construye el modelo completo del desempeño del PASDBP, donde se toman en cuenta las variables de todos los grupos.

a) Construcción y análisis de los modelos de regresión ordinal para cada uno de los grupos

En este apartado se presenta la construcción de modelos grupales con variables predictoras pertenecientes a cada uno de ellos. Mediante el método de “Backward Stepwise”, también llamado de “Eliminación progresiva”, en el cual se parte de un modelo de regresión con todas las variables y en cada paso se elimina la variable menos influyente¹⁴³ hasta encontrar el modelo final que contenga todas las variables significativas.

Para la modelación del desempeño del PASDBP para cada grupo se define la variable dependiente como:

$Y = \text{Desempeño del PASDBP}$ con $n = 4$ categorías,

y el siguiente orden en sus categorías (orden ascendente):

$$Y_1 = \text{bajo} < Y_2 = \text{medio} < Y_3 = \text{alto} < Y_4 = \text{excelente}$$

X como el conjunto de m variables independientes, como un vector: $X = (X_1, X_2, X_3, \dots, X_m)$

El predictor lineal que buscamos obtener es: $f(\theta_j(X)) = f(P(Y \leq y_j | X)) = \alpha_j + \beta'X$

Con $j = 1, 2, 3$ Donde α_j y $\beta = (\beta_1, \beta_2, \dots, \beta_m)'$ son los parámetros a estimar.

A continuación el procedimiento que se utiliza en cada uno de los grupos de las variables independientes para obtener el modelo de regresión ordinal.

¹⁴³ Este procedimiento de “Eliminación progresiva” tiene los inconvenientes de necesitar muchos cálculos si se tiene un gran número de variables independientes, pero tiene la ventaja de no eliminar variables significativas para el modelo.

i. Modelo con las características de localidad

El primer grupo es el de Características de Localidad, con dos variables: Tipo de Localidad y Nivel de Marginación.

Paso 1. Modelo Completo

Para la construcción del modelo con ambas variables, primero se verifica el cumplimiento del supuesto de Líneas paralelas para determinar si la regresión ordinal es el método adecuado para el análisis de los datos. Para la prueba se plantean las siguientes hipótesis:

$$H_o: \beta^1 = \beta^2 = \beta^3_{144}$$

$$H_a: \beta^i \neq \beta^j \quad \text{para } i \neq j \quad \text{con } i, j = 1, 2, 3$$

En el siguiente cuadro se resume la información para realizar la prueba.

Prueba de Líneas Paralelas

Modelo	-2 Log Likelihood	Ji-Cuadrada	G.L.	P-value
Hipótesis Nula ¹⁴⁵	79.122			
General	73.199	5.923	6	0.432

Función liga: Logit.

A partir de la información anterior se obtiene:

$$\chi^2_{(6), \alpha=0.05} = 12.5916 > 5.923, \text{ la cual tiene asociada una } p - \text{value} = 0.432 > \alpha = 0.05$$

Por lo que no hay evidencia suficiente para rechazar la hipótesis nula y se cumple el supuesto de Líneas paralelas.

Ahora se procede a realizar la Prueba de Máxima Verosimilitud, que básicamente determina si el modelo tiene sentido, y con ello el cálculo de las probabilidades, contrastando la verosimilitud del modelo nulo (modelos sin variables) vs la verosimilitud del modelo (considera las dos variables: tipo de localidad y nivel de marginación).

¹⁴⁴ En este caso $n-1=3$, ya que la variable dependientes tienen cuatro niveles de desempeño del PASDBP.

¹⁴⁵ La hipótesis nula establece que todos los coeficientes dependientes son los mismos en todas las categorías de respuesta.

Bajo las hipótesis:

$$H_0: \beta_1 = 0, \beta_2 = 0, \beta_3 = 0$$

$$H_1: \text{al menos una } \beta_j \neq 0^{146}$$

Se busca rechazar la hipótesis nula, es decir, que la estimación de las probabilidades para el desempeño es mejor con un modelo que incluye las variables predictivas que uno sin ellas.

Prueba de Verosimilitudes

Modelo	-2 Log Likelihood	Ji-Cuadrada	G.L.	P-value
Nulo	100.915			
Final¹⁴⁷	79.122	21.793	3	0.000

Función liga: Logit.

Con información del cuadro anterior se tiene que

$$C^{148} = 21.793 > \chi^2_{(5-2)}, \alpha=0.05 = 7.8147$$

Con $P - value = 0.000 < \alpha = 0.05$, por lo que se rechaza H_0 . Entonces se determina que al menos una de estas variables contribuyen a explicar a la variable respuesta.

A continuación se realiza la estimación del modelo con el método de máxima verosimilitud. En la siguiente tabla se muestran los valores estimados para β 's y α 's

¹⁴⁶ En total se tienen 5 categorías: 2 categorías de la variable tipo de localidad y 3 categorías de la variable nivel de marginación. Recuerde que se tienen un número de β 's = # de categorías de las variables independientes - # de variables independientes. En este caso $3 = 5 - 2$

¹⁴⁷ Modelo que incluye las variables predictoras especificadas

¹⁴⁸ C es el estadístico de prueba de la Prueba de Máxima Verosimilitud.

Estimación de los parámetros.

		Coefficientes	Std. Error	Wald	G.L.	P-value
α's	Bajo	-0.493	0.237	4.334	1	0.037
	Mediano	0.673	0.238	7.981	1	0.005
	Alto	1.811	0.251	52.224	1	0.000
Tipo de Localidad	Urbano	0.917	0.224	16.722	1	0.000
	Rural	0 ^a	.	.	0	.
Nivel de Marginación	Bajo	0.008	0.258	0.001	1	0.974
	Medio	-0.193	0.29	0.445	1	0.505
	Alto	0 ^a	.	.	0	.

Función enlace: Logit.

a. Categoría de comparación.

Para determinar qué valores tomarán los parámetros en los modelos, se utiliza la información de la tabla anterior de la Prueba de Wald, que se distribuye $\chi^2_{(1)}, \alpha=0.05=3.8415$ con las hipótesis:

$$H_0: \beta_i = 0 \quad vs \quad H_1: \text{al menos una } \beta_i \neq 0$$

Con $i = 1, 2, 3$ dónde:

β_1 = Coeficiente de la variable Tipo de Localidad de la categoría Urbano

β_2 = Coeficiente de la variable Nivel de Marginación de la categoría Bajo

β_3 = Coeficiente de la variable Nivel de Marginación de la categoría Medio

Como se puede observar en la tabla anterior, se obtienen los siguientes resultados en la Prueba de Wald:

- Son estadísticamente significativos, y diferentes de cero, los coeficientes de:
 - Tipo de Localidad Urbano.
- No son estadísticamente significativos los coeficientes de:
 - Nivel de Marginación Bajo
 - Nivel de Marginación Medio

Y para determinar el valor de los umbrales¹⁴⁹ α 's (del desempeño del PASDBP) también se realiza la Prueba de Wald, que se distribuye $\chi^2_{(1)}, \alpha=0.05=3.8415$, con las hipótesis:

¹⁴⁹ Los umbrales α 's se refieren a las constantes que se asignan a cada una de las categorías de respuesta en el modelo de regresión ordinal.

$$H_0: \alpha_i = 0 \quad vs \quad H_1: \text{al menos una } \alpha_i \neq 0$$

Con $i = 1, 2, 3$ dónde:

α_1 = Umbral (constante) de Nivel de Desempeño Bajo

α_2 = Umbral (constante) de Nivel de Desempeño Mediano

α_3 = Umbral (constante) de Nivel de Desempeño Alto

- Como se puede observar en la tabla, resultan estadísticamente significativas los tres umbrales de los niveles de desempeño del PASDBP.

Paso 2. Eliminando la variable Nivel de Marginación

Con el fin de llegar el modelo más parsimonioso¹⁵⁰ y siguiendo el procedimiento de eliminación progresiva, se debe eliminar la variable menos significativa del modelo y evaluar si el modelo resultante es mejor que el anterior.

El criterio que se considera para determinar qué variable es la menos significativa, es a partir de la significancia de los coeficientes β 's de cada variable independiente mediante la prueba de Wald que se realizó para la estimación de β 's y α 's del modelo anterior.

Con respecto a la tabla de "Estimación de parámetros" del Paso 1, tenemos que:

- la variable Tipo de localidad está formada por dos categorías. El cálculo del coeficiente β_1 se realiza únicamente para la primera categoría, donde el $P - value = 0.000 < \alpha = 0.05$ para tipo de localidad Urbana.
- la variable Nivel de marginación tiene tres categorías y se tiene un $P - value = 0.974 > \alpha = 0.05$ para la categoría Bajo y $P - value = 0.505 > \alpha = 0.05$ para la categoría Medio.

Por lo tanto, la variable predictiva con valores más altos de $P - value$, que se considera debe salir del modelo por no resultar significativa (en ninguna de sus categorías), es el Nivel de Marginación.¹⁵¹

Lo primero que se tiene que hacer es evaluar si el modelo, sin variable nivel de marginación, cumple con el supuesto de Momios Proporcionales. Con el cual evaluamos si los datos son adecuados para aplicar la regresión ordinal.

¹⁵⁰ Existe un criterio de optar, entre varios modelos igualmente aceptables, por el más sencillo; esto es, el que contenga menos términos y es llamado Criterio de Parsimonia.

¹⁵¹ En el caso de la variable Nivel de Marginación las dos categorías, sometidas a la Prueba de Wald, resultaron no significativas para el modelo. Sin embargo, pudo haberse dado el caso en que una de las categorías sí resultara significativa y la otra no. En ése caso es recomendable hacer estimaciones con la variable y sin la variable para comparar resultados y evaluar el aporte de la misma.

Prueba de Líneas Paralelas

Modelo	-2 Log Likelihood	Ji-Cuadrada	G. L.	P-value
Hipótesis Nula	31.133			
General	30.928	0.205	2	0.903

Función liga: Logit.

Con información del cuadro anterior se realiza la Prueba de Líneas paralelas. Se tiene:

$$\chi^2_{(2), \alpha=0.05} = 5.9915 > 0.205, \text{ con } P - \text{value} = 0.903 > \alpha = 0.05 ;$$

Entonces se concluye que no existe evidencia suficiente para rechazar la hipótesis nula y se cumple el supuesto de Momios proporcionales, con lo cual se puede aplicar el método de regresión ordinal.

En seguida, se efectúa la prueba de Cociente de verosimilitudes para determinar si la construcción del nuevo modelo, con únicamente la variable de Tipo de Localidad, tiene sentido.

Cociente de Verosimilitudes y Contraste de Máxima Verosimilitud

Modelo	-2 Log Likelihood	Ji-Cuadrada	G.L.	P-value
Nulo	52.023			
Final	31.133	20.89	1	0.000

Función enlace: Logit.

Entonces se tiene que $C = 20.89 > \chi^2_{(2-1), \alpha=0.05} = 3.8415$, lo cual se confirma con el $P - \text{value} = 0.000 < \alpha = 0.05$

Por lo tanto, el modelo con la variable independiente tiene mayor significancia que el modelo sin la variable.

Ahora que se probó que tiene sentido el modelo para su posterior construcción, se debe comparar la verosimilitud del modelo con ambas variables de las características de localidad con el nuevo modelo que tiene sólo la variable tipo de localidad, mediante la Prueba de Contraste de Máxima Verosimilitud. Con la información del cuadro de Cociente de Verosimilitudes del Paso 1 y 2, se calcula el estadístico para realizar la prueba

$$G = (-2\text{Log}[Verosimilitud Modelo Anterior]) - (-2\text{Log}[Verosimilitud Modelo Nuevo])$$

Que se distribuye como $\chi^2_{(r), \alpha=0.05}$ donde $r = \text{número de variables nuevas}$.

$$\text{Se obtiene que } G = 79.122 - 31.133 = 47.983 > 3.8415 = \chi^2_{(1), \alpha=0.05} .$$

Con lo que se rechaza la hipótesis nula y se concluye que para el bloque de Características de la Localidad es mejor el nuevo modelo con únicamente la variable Tipo de Localidad.

A continuación se presenta un cuadro que resume los datos de la estimación de los parámetros β 's y α 's por el método de Máxima verosimilitud.

Estimación de los Parámetros

		Coefficientes	Std. Error	Wald	G.L.	P-value
α's	Bajo	-0.442	0.185	5.746	1	0.017
	Mediano	0.723	0.187	14.976	1	0.000
	Alto	1.861	0.203	84.015	1	0.000
Tipo de Localidad	Urbano	0.934	0.207	20.363	1	0.000
	Rural	0 ^a	.	.	0	.

Función enlace: Logit.

a. Categoría de comparación.

Repitiendo el procedimiento para la construcción del modelo, se realiza la Prueba de Wald para determinar que coeficientes son estadísticamente significativos, diferentes de cero.

Se utiliza la información de la tabla anterior para realizar la Prueba de Wald, que se distribuye $\chi^2_{(1)}$, $\alpha=0.05$, con las siguientes hipótesis:

$$H_0: \beta_1 = 0 \quad vs \quad H_1: \beta_1 \neq 0$$

Con β_1 = Coeficiente de la variable Tipo de Localidad de la categoría Urbano

Como resultado en la Prueba de Wald se obtiene que:

- Es estadísticamente significativo, y diferente de cero el coeficiente de:
 - Tipo de Localidad Urbano.

Y para determinar el valor de los umbrales α 's (del desempeño del PASDBP) también se realiza la Prueba de Wald, que se distribuye $\chi^2_{(1)}$, $\alpha=0.05$, con las hipótesis: con las siguientes hipótesis:

$$H_0: \alpha_i = 0 \quad vs \quad H_1: \text{al menos una } \alpha_i \neq 0$$

Con $i = 1, 2, 3$ dónde:

- α_1 = Umbral (constante) de Nivel de Desempeño Bajo
- α_2 = Umbral (constante) de Nivel de Desempeño Mediano
- α_3 = Umbral (constante) de Nivel de Desempeño Alto

- Como se puede observar en la tabla, resultan estadísticamente significativas los tres umbrales de los niveles de desempeño del PASDBP.

Nótese que todos los coeficientes, tanto de los umbrales del nivel de desempeño como del Tipo de Localidad, resultaron significativos, por lo que todos formaran parte del modelo. A continuación se presentan las probabilidades condicionales que son definidas como los predictores lineales para cada una de las categorías del desempeño del PASDBP.

$$\begin{aligned} f(\theta_1(X)) &= f(P(Y \leq Y_1 | X_{TL})) = \alpha_1 - \beta_{TL}X_{TL} = -0.442 - 0.934X_{TL} \\ f(\theta_2(X)) &= f(P(Y \leq Y_2 | X_{TL})) = \alpha_2 - \beta_{TL}X_{TL} = 0.723 - 0.934X_{TL} \\ f(\theta_3(X)) &= f(P(Y \leq Y_3 | X_{TL})) = \alpha_3 - \beta_{TL}X_{TL} = 1.861 - 0.934X_{TL} \end{aligned}$$

Donde:

α_1 = Umbral (constante) de Nivel de Desempeño Bajo

α_2 = Umbral (constante) de Nivel de Desempeño Mediano

α_3 = Umbral (constante) de Nivel de Desempeño Alto

β_{TL} = Coeficiente de la variable Tipo de Localidad de la categoría Urbano

X_{TL} = Variable Tipo de Localidad de la categoría Urbano

En donde $X_{TL} = 0,1$ dependiendo el tipo de localidad al que se refiera. $X_{TL} = 0$ si el tipo de localidad es rural y $X_{TL} = 1$ si el tipo de localidad es urbano.

A continuación se presentan las probabilidades para cada una de las categorías del desempeño del PASDBP $Y_1 = \text{Bajo}$, $Y_2 = \text{Mediano}$, $Y_3 = \text{Alto}$ y $Y_4 = \text{Excelente}$ que se construyen con los respectivos umbrales α 's y los parámetros β 's en cada uno de ellos:

$$P(Y = Y_1 | X_{TL}) = \left[\frac{1}{1 + e^{-(\alpha_1 - \beta_{TL}X_{TL})}} \right] = \left[\frac{1}{1 + e^{-(-0.442 - 0.934X_{TL})}} \right]$$

$$\begin{aligned} P(Y = Y_2 | X_{TL}) &= \left[\frac{1}{1 + e^{-(\alpha_2 - \beta_{TL}X_{TL})}} \right] - P(Y = Y_1 | X_{TL}) \\ &= \left[\frac{1}{1 + e^{-(0.723 - 0.934X_{TL})}} \right] - \left[\frac{1}{1 + e^{-(-0.442 - 0.934X_{TL})}} \right] \end{aligned}$$

$$\begin{aligned} P(Y = Y_3 | X_{TL}) &= \left[\frac{1}{1 + e^{-(\alpha_3 - \beta_{TL}X_{TL})}} \right] - P(Y = Y_2 | X_{TL}) \\ &= \left[\frac{1}{1 + e^{-(1.861 - 0.934X_{TL})}} \right] - \left[\frac{1}{1 + e^{-(0.723 - 0.934X_{TL})}} \right] \end{aligned}$$

$$P(Y = Y_4 | X_{TL}) = 1 - P(Y = Y_3 | X_{TL}) = 1 - \left[\frac{1}{1 + e^{-(1.861 - 0.934X_{TL})}} \right]$$

Entonces, si se desea conocer la probabilidad condicional de obtener un alto desempeño dado que el tipo de localidad es rural, se calcula:

$$P(Y = Y_4 | X_{TL} = 0) = 1 - \left[\frac{1}{1 + e^{-(1.861 - 0.934(0))}} \right] = 0.135$$

Recuerde que en realidad, el modelo de regresión ordinal nos da las probabilidades acumuladas, y mediante unos simples cálculos se obtiene las probabilidades absolutas. Por lo que si se desea conocer la probabilidad condicional de obtener un alto desempeño o menos (es decir, obtener un bajo, medio o alto desempeño), dado que el tipo de localidad es urbano se calcula de la siguiente forma:

$$P(Y \leq Y_3 | X_{TL} = 1) = \left[\frac{1}{1 + e^{-(1.861 - 0.934(1))}} \right] = 0.717$$

Después se calculan todas las probabilidades condicionales resultantes de las $(2 * 4 = 8)$ combinaciones posibles del vector (Y_j, X_{TL}) . A continuación el cuadro presenta la información obtenida de estos cálculos:

Probabilidades Estimadas

Tipo de Localidad	Desempeño del PASDBP			
	Bajo	Mediano	Alto	Excelente
Urbano n= 328	0.202	0.246	0.269	0.283
Rural n= 104	0.391	0.282	0.192	0.135

Función liga: Logit.

El cuadro se interpreta de la siguiente forma: la probabilidad de obtener un desempeño i (bajo, mediano, alto y excelente) dado que se tiene un determinado tipo de localidad j (Rural=0, Urbano=1) es $P(Y = Y_i | X_{TL} = j)$.

Para ejemplificar la lectura de la tabla anterior, obtenemos que la probabilidad de obtener un bajo desempeño en tipo de localidad urbana es de 0.202. Es decir, con esa probabilidad se espera que 66 MSD urbanos presenten bajo desempeño; lo cual es bastante acertado ya que en realidad son 67 los MSD observados en localidad urbana con bajo desempeño.

Haciendo el mismo ejercicio con las demás localidades encontramos que no excede en 2 la diferencia del número MSD esperados con los observados en cada una de las combinaciones.

Es importante resaltar que entre las localidades rurales el desempeño bajo tiene la probabilidad más alta de ocurrencia. Por el contrario, para el tipo de localidad Urbana se obtiene la probabilidad más alta en la categoría de excelente desempeño y va decreciendo hasta la categoría de bajo desempeño.

Se puede concluir que el tipo de localidad donde son instalados los módulos influye en forma significativa en el desempeño del programa; y que la probabilidad de obtener un mejor desempeño es más alta en comunidades urbanas que en rurales.

A continuación, se realizan las pruebas de χ^2 de Pearson, Devianza y se calcula la Pseudo – R^2 de Nagelkerke para evaluar la bondad del modelo, es decir, para saber cuánta de la varianza del desempeño del programa es explicada por el Tipo de Localidad.

La Prueba de Devianza y la Prueba de χ^2 de Pearson comparten las hipótesis de $H_0: F(x) = F_0(x)$ vs $H_a: F(x) \neq F_0(x)$

En el siguiente cuadro se presenta la información respecto a los estadísticos: D para la Prueba de Devianza y S para la de Pearson, con los que se realizan las pruebas.

Pearson y Devianza

	Ji-Cuadrada	G.L.	P-value
Pearson	0.206	2	0.902
Devianza	0.205	2	0.903

Función enlace: Logit.

Para la Prueba de Devianza se tiene $\chi^2_{(2), \alpha=0.05} = 5.9915 > 0.206 = D$, donde el $P - value = 0.902 > \alpha = 0.05$ en consecuencia no se tiene evidencia suficiente para rechazar la hipótesis nula.

En la Prueba χ^2 de Pearson se tiene que $\chi^2_{(2), \alpha=0.05} = 5.9915 > 0.205 = S$, con $P - value = 0.642 > \alpha = 0.05$, por lo que se tiene el mismo resultado de no rechazo de la hipótesis nula que en la Prueba de Devianza.

Finalmente se llega a la Pseudo – R^2 de Nagelkerke:

Pseudo R-Cuadrada

Nagelkerke	0.050
-------------------	-------

Función enlace: Logit.

Se determina que el coeficiente obtenido de R^2 de Nagelkerke es bajo, sin embargo, hasta este momento el modelo considera sólo uno de los múltiples factores que se propusieron como determinantes del desempeño del PASDBP.

En resumen, el modelo de regresión ordinal para el Desempeño del PASDBP que considera únicamente las variables del grupo de Características de Localidad establece únicamente como variable significativa al Tipo de Localidad y desecha el Nivel de Marginación.

ii. Modelo con las características del MSD

El siguiente grupo es el de Características del MSD con las variables Experiencia/ Tiempo en el PASDBP, Antigüedad de la Biblioteca, Tamaño de la Biblioteca, Número de computadoras asignadas, Financiamiento y Sostenibilidad.

Para desarrollar el modelo de regresión ordinal correspondiente de este grupo de variables se siguió el mismo procedimiento, “Backward Stepwise”.

Paso 1. Modelo Completo

Se parte de la construcción de un modelo que incluye todas las variables del grupo, en este caso, contiene seis variables independientes. Es por ello, que rápidamente el número de celdas que debería contener la tabla de contingencia se convierte en el primer inconveniente.¹⁵²

Sin embargo, más allá de la advertencia sobre las tablas de contingencia con las que se trabajará se entiende la importancia de poder estudiar todas las variables del grupo simultáneamente y por consiguiente la significancia del modelo que se obtendría a partir de ello. Por lo que se prosigue a evaluar el cumplimiento del supuesto de Líneas paralelas.

Prueba de Líneas Paralelas

Modelo	-2 Log Likelihood	Ji-Cuadrada	G.L.	Sig.
Hipótesis Nula	641.807			
General	618.896	22.912	22	0.407

Función liga: Logit.

Se tiene que una $\chi^2_{(22)}, \alpha=0.05 = 33.9245 > 22.912$,

Con un $P - value = 0.222 > \alpha = 0.05$

Se concluye que no se tiene evidencia suficiente para rechazar la hipótesis nula, por lo tanto los datos cumplen el supuesto de Líneas paralelas.

Continuamos con la aplicación de la Prueba de Verosimilitudes, también llamada la Prueba de Cociente de Máxima Verosimilitud. La cual se realiza bajo las hipótesis:

$$H_0: \beta_1 = 0, \beta_2 = 0, \dots, \beta_{11} = 0 \quad vs \quad H_1: \text{al menos una } \beta_j \neq 0$$

Básicamente desea probar que la estimación del modelo es estadísticamente significativa con coeficientes diferentes de cero.

Prueba de Verosimilitudes

Modelo	-2 Log Likelihood	Ji-Cuadrada	G.L.	P-value
Nulo	702.081			
Final¹⁵³	641.807	60.274	11	0.000

Función liga: Logit.

¹⁵² Al trabajar con todas ellas en una tablas múltiples extensas se advierte que es considerable el número de celdas que resultan con cero frecuencias. En este caso, se encuentra que 440 celdas (62.5152%) de la tabla no presentan ninguna frecuencia, es decir, que existen 440 combinaciones del vector $(Y_j, X_E, X_{AB}, X_{TB}, X_{NC}, X_F, X_S)$ con frecuencias nulas.

¹⁵³ Modelo que incluye las variables predictoras especificadas.

Con la información del cuadro anterior se obtiene:

$$C = 60.274 > \chi^2_{(17-6)}, \alpha=0.05 = 19.5762$$

Con un $P - value = 0.000 < \alpha = 0.05$

Por lo que se rechaza la hipótesis nula y se determina que el modelo tiene sentido, ya que al menos una $\beta_j \neq 0$. Las estimaciones de los parámetros se muestran en el siguiente cuadro:

Estimación de los Parámetros

		Coeficientes	Std. Error	Wald	G.L.	P-value
α's	Bajo	-2.3556	0.475	24.618	1	0.000
	Mediano	-1.092	0.464	5.529	1	0.019
	Alto	0.232	0.459	0.256	1	0.613
Experiencia/Tiempo en el PASDBP	Muy Poca	-0.992	0.3266	9.219	1	0.002
	Poca	-0.883	0.291	9.179	1	0.002
	Mediana	-0.981	0.306	10.254	1	0.001
	Mucha	0	.	.	0	.
Antigüedad de la Biblioteca	Hasta 14 años	0.158	0.345	0.209	1	0.648
	Más 14 y menos 20 años	-0.348	0.315	1.227	1	0.268
	Más 20 y menos 23 años	0.065	0.299	0.0475	1	0.828
	23 años y más	0 ^a	.	.	0	.
Tamaño de la Biblioteca	Adecuado para el MSD	0.104	0.216	0.233	1	0.629
	Inadecuado para el MSD	0 ^a	.	.	0	.
Computadoras Asignadas	2 -3 computadoras	-1.625	0.369	19.383	1	0.000
	4- 6 computadoras	-0.857	0.339	6.394	1	0.011
	9- 13 computadoras	0 ^a	.	.	0	.
Sostenibilidad	Cuenta con plan para continuar	0.431	0.212	4.127	1	0.042
	No Cuenta con plan para continuar	0 ^a	.	.	0	.
Financiamiento	Si	0.682	0.236	8.317	1	0.004
	No	0 ^a	.	.	0	.

Función liga: Logit

a. Categoría de comparación.

Para determinar qué valores tomarán los parámetros en los modelos, se utiliza la Prueba de Wald, con las siguientes hipótesis:

$$H_0: \beta_i = 0 \quad vs \quad H_1: \beta_i \neq 0, \text{ para algún } i$$

Con $i = 1, 2, \dots, 11$ donde:

β_1 = Coeficiente de la variable Computadoras asignadas por el PASDBP de la categoría 2-3 computadoras

β_2 = Coeficiente de la variable Computadoras asignadas por el PASDBP de la categoría 4-6 computadoras

β_3 = Coeficiente de la variable Experiencia en el PASDBP de la categoría Muy Poca

β_4 = Coeficiente de la variable Experiencia en el PASDBP de la categoría Poca

β_5 = Coeficiente de la variable Experiencia en el PASDBP de la categoría Mediana

β_6 = Coeficiente de la variable Antigüedad de la Biblioteca de la categoría Hasta 14 años

β_7 = Coeficiente de la variable Antigüedad de la Biblioteca de la categoría Más 14 y hasta 20 años

β_8 = Coeficiente de la variable Antigüedad de la Biblioteca de la categoría Más 20 y hasta 23 años

β_9 = Coeficiente de la variable Tamaño de la Biblioteca de la categoría Adecuado para el módulo

β_{10} = Coeficiente de la variable Financiamiento de la categoría Si

β_{11} = Coeficiente de la variable Sostenibilidad por la categoría Cuenta con Plan para continuar el programa

Como se puede observar en la tabla anterior, los siguientes resultados obtenidos en la Prueba de Wald son los siguientes:

- Son estadísticamente significativos, y diferentes de cero, los coeficientes de:
 - Las tres categorías de la variable Experiencia en el PASDBP (muy poca, poca y mediana)
 - Las dos categorías de la variable Computadoras Asignadas por el PASDBP (2-3, 4-6)
 - La categoría Cuenta con plan para continuar le programa de la variable Sostenibilidad
 - La categoría Sí de la variable Financiamiento
- No son estadísticamente significativos los coeficientes de:
 - Las tres categorías de la variable Antigüedad de la Biblioteca (hasta 14 años, Más de 14 y hasta 20 años, Más de 20 y hasta 23 años.)
 - La categoría de Adecuado para el MSD de la variable Tamaño de la Biblioteca

Y para determinar el valor de los umbrales α 's (del desempeño del PASDBP) también se realiza la Prueba de Wald, con las siguientes hipótesis:

$$H_0: \alpha_i = 0 \quad vs \quad H_1: \text{al menos para una } \alpha_i \neq 0 \text{ Con } i = 1, 2, 3 \text{ dónde:}$$

α_1 = Umbral (constante) de Nivel de Desempeño Bajo

α_2 = Umbral (constante) de Nivel de Desempeño Mediano

α_3 = Umbral (constante) de Nivel de Desempeño Alto

- A partir de la información de la tabla se puede observar que resultan estadísticamente significativas los umbrales de los niveles Bajo y Mediano de desempeño del PASDBP.

Después de analizar mediante la Prueba de Wald que coeficientes son significativos, las probabilidades quedan definidas como los predictores lineales:

$$\begin{aligned}
 f(\theta_1(X)) &= f(P(Y \leq Y_1 | X_{E1}, X_{E2}, X_{E3}, X_{AB1}, X_{AB2}, X_{AB3}, X_{TB}, X_{NC1}, X_{NC2}, X_F, X_S)) \\
 &= \alpha_1 - \beta_{E1}X_{E1} - \beta_{E2}X_{E2} - \beta_{E3}X_{E3} - \beta_{AB1}X_{AB1} - \beta_{AB2}X_{AB2} - \beta_{AB3}X_{AB3} \\
 &\quad - \beta_{TB}X_{TB} - \beta_{NC1}X_{NC1} - \beta_{NC2}X_{NC2} - \beta_S X_S - \beta_F X_F \\
 &= -2.356 + 0.992X_{E1} + 0.883X_{E2} + 0.981X_{E3} - 0X_{AB1} - 0X_{AB2} - 0X_{AB3} \\
 &\quad - 0X_{TB} + 1.625X_{NC1} + 0.857X_{NC2} - 0.431X_S - 0.682X_F
 \end{aligned}$$

$$\begin{aligned}
 f(\theta_2(X)) &= f(P(Y \leq Y_2 | X_{E1}, X_{E2}, X_{E3}, X_{AB1}, X_{AB2}, X_{AB3}, X_{TB}, X_{NC1}, X_{NC2}, X_F, X_S)) \\
 &= \alpha_2 - \beta_{E1}X_{E1} - \beta_{E2}X_{E2} - \beta_{E3}X_{E3} - \beta_{AB1}X_{AB1} - \beta_{AB2}X_{AB2} - \beta_{AB3}X_{AB3} \\
 &\quad - \beta_{TB}X_{TB} - \beta_{NC1}X_{NC1} - \beta_{NC2}X_{NC2} - \beta_S X_S - \beta_F X_F \\
 &= -1.092 + 0.992X_{E1} + 0.883X_{E2} + 0.981X_{E3} - 0X_{AB1} - 0X_{AB2} - 0X_{AB3} \\
 &\quad - 0X_{TB} + 1.625X_{NC1} + 0.857X_{NC2} - 0.431X_S - 0.682X_F
 \end{aligned}$$

$$\begin{aligned}
 f(\theta_3(X)) &= f(P(Y \leq Y_3 | X_{E1}, X_{E2}, X_{E3}, X_{AB1}, X_{AB2}, X_{AB3}, X_{TB}, X_{NC1}, X_{NC2}, X_F, X_S)) \\
 &= \alpha_3 - \beta_{E1}X_{E1} - \beta_{E2}X_{E2} - \beta_{E3}X_{E3} - \beta_{AB1}X_{AB1} - \beta_{AB2}X_{AB2} - \beta_{AB3}X_{AB3} \\
 &\quad - \beta_{TB}X_{TB} - \beta_{NC1}X_{NC1} - \beta_{NC2}X_{NC2} - \beta_S X_S - \beta_F X_F \\
 &= 0 + 0.992X_{E1} + 0.883X_{E2} + 0.981X_{E3} - 0X_{AB1} - 0X_{AB2} - 0X_{AB3} - 0X_{TB} \\
 &\quad + 1.625X_{NC1} + 0.857X_{NC2} - 0.431X_S - 0.682X_F
 \end{aligned}$$

Donde:

α_i = Umbral (constante) de los Niveles de Desempeño con $i = 1,2,3$.

Para 1=Bajo, 2= Mediano y 3=Alto

β_{Ci} = Coeficiente de la variable Computadoras asignadas por el PASDBP de la categoría 2-3 computadoras con $i = 1,2$

Para 1= "2-3 computadoras" y 2= "4-6 computadoras"

β_{Ei} = Coeficiente de la variable Experiencia en el PASDBP con $i = 1,2,3$

Para 1=Muy Poca, 2= Poca y 3=Mediana

β_{ABi} = Coeficiente de la variable Antigüedad de la Biblioteca con $i = 1,2,3$

Para 1= Hasta 14 años, 2= Más de 14 y hasta 20 años y 3= Más de 20 y hasta 23 años

β_{TB} = Coeficiente de la variable Tamaño de la Biblioteca de la categoría Adecuado para el módulo

β_F = Coeficiente de la variable Financiamiento de la categoría Si

β_S = Coeficiente de la variable Sostenibilidad par la categoría Cuenta con Plan para continuar el programa

X_{Ci} = Variable Capacitación por parte de la DGB

X_{Ei} = Variable Experiencia en el PASDBP

X_{ABi} = Variable Antigüedad de la Biblioteca

X_{TB} = Variable Tamaño de la Biblioteca de la categoría Adecuado para el módulo

X_F = Variable Financiamiento de la categoría Si

X_S = Variable Sostenibilidad par la categoría Cuenta con Plan para continuar el programa

Así que $X_{Ei} = 1,0$ con $i = 1,2,3$ dependiendo del Tiempo en el PASDBP al que se refiera. Por ejemplo,

si $X_{E1} = 1, X_{E2} = 0$ y $X_{E3} = 0$ entonces se tiene muy poca experiencia en el PASDBP.

Si $X_{E1} = 0, X_{E2} = 1$ y $X_{E3} = 0$ entonces se tiene poca experiencia en el programa.

Si $X_{E1} = 0, X_{E2} = 0$ y $X_{E3} = 1$ entonces se tiene mediana experiencia en el programa.

Y por último, si $X_{E1} = 0, X_{E2} = 0$ y $X_{E3} = 0$ entonces se tiene mucha experiencia en el programa.

De forma similar la variable independiente Antigüedad de la Biblioteca puede expresar sus cuatro categorías con X_{AB1}, X_{AB2} y X_{AB3} . Donde $X_{ABi} = 1, 0$ con $i = 1, 2, 3$. Por ejemplo,

si $X_{AB1} = 0, X_{AB2} = 1$ y $X_{AB3} = 0$ entonces se tiene una Biblioteca de 14 a 20 años de antigüedad.

También la variable de Computadoras asignadas donde $X_{NCi} = 1, 0$ con $i = 1, 2$ expresa sus tres categorías de forma análoga a las variables anteriores.

La categoría de Adecuada para el MSD en la variable Tamaño de la Biblioteca se expresa como $X_{TB} = 1$ y $X_{TB} = 0$ si resulta inadecuado.

Por su parte, las variables de Financiamiento y Sostenibilidad que también tienen un solo coeficiente se expresan de forma análoga.

Una vez definidas cada una de las categorías de las variables predictoras, se construye el modelo de regresión ordinal para las categorías:

$Y_1 = \text{Bajo}, Y_2 = \text{Medio}, Y_3 = \text{Alto}$ y $Y_4 = \text{Excelente}$ y se definen como:

$$P(Y = Y_1 | X_{E1}, X_{E2}, X_{E3}, X_{AB1}, X_{AB2}, X_{AB3}, X_{TB}, X_{NC1}, X_{NC2}, X_F, X_S) = \left[\frac{1}{1 + e^{-(2.356 + 0.992X_{E1} + 0.883X_{E2} + 0.981X_{E3} - 0X_{AB1} - 0X_{AB2} - 0X_{AB3} - 0X_{TB} + 1.625X_{NC1} + 0.857X_{NC2} - 0.431X_S - 0.682X_F)}} \right]$$

$$P(Y = Y_2 | X_{E1}, X_{E2}, X_{E3}, X_{AB1}, X_{AB2}, X_{AB3}, X_{TB}, X_{NC1}, X_{NC2}, X_F, X_S) = \left[\frac{1}{1 + e^{-(1.092 + 0.992X_{E1} + 0.883X_{E2} + 0.981X_{E3} - 0X_{AB1} - 0X_{AB2} - 0X_{AB3} - 0X_{TB} + 1.625X_{NC1} + 0.857X_{NC2} - 0.431X_S - 0.682X_F)}} \right] - P(Y = Y_1 | X_{E1}, X_{E2}, X_{E3}, X_{AB1}, X_{AB2}, X_{AB3}, X_{TB}, X_{NC1}, X_{NC2}, X_F, X_S)$$

$$P(Y = Y_3 | X_{E1}, X_{E2}, X_{E3}, X_{AB1}, X_{AB2}, X_{AB3}, X_{TB}, X_{NC1}, X_{NC2}, X_F, X_S) = \left[\frac{1}{1 + e^{-(0.092 + 0.992X_{E1} + 0.883X_{E2} + 0.981X_{E3} - 0X_{AB1} - 0X_{AB2} - 0X_{AB3} - 0X_{TB} + 1.625X_{NC1} + 0.857X_{NC2} - 0.431X_S - 0.682X_F)}} \right] - P(Y = Y_2 | X_{E1}, X_{E2}, X_{E3}, X_{AB1}, X_{AB2}, X_{AB3}, X_{TB}, X_{NC1}, X_{NC2}, X_F, X_S)$$

$$P(Y = Y_4 | X_{E1}, X_{E2}, X_{E3}, X_{AB1}, X_{AB2}, X_{AB3}, X_{TB}, X_{NC1}, X_{NC2}, X_F, X_S) = 1 - P(Y = Y_3 | X_{E1}, X_{E2}, X_{E3}, X_{AB1}, X_{AB2}, X_{AB3}, X_{TB}, X_{NC1}, X_{NC2}, X_F, X_S)$$

A continuación se presenta un ejemplo del cálculo de una probabilidad condicional. Se obtiene la probabilidad de tener un bajo desempeño dado que se tiene mucha experiencia en el PASDBP, la biblioteca donde está instalado el módulo tiene más de 23 años de antigüedad, el espacio donde se encuentra el MSD es adecuado al tamaño de la biblioteca, se le asignaron 13 computadoras al módulo y no tiene financiamiento ni sostenibilidad:

$$P(Y = Y_1 | X_{E1} = 0, X_{E2} = 0, X_{E3} = 0, X_{AB1} = 0, X_{AB2} = 0, X_{AB3} = 0, X_{TB} = 1, X_{NC1} = 0, X_{NC2} = 0, X_F = 0, X_S = 0) = \left[\frac{1}{1 + e^{-(-2.356 + 0.992(0) + 0.883(0) + 0.981(0) - 0 - 0 - 0 + 1.625(0) + 0.857(0) - 0.431(0) - 0.682(0))}} \right] = 0.0866$$

Con las probabilidades absolutas, como la que se calculó en el ejemplo anterior, se construyen las tablas con el fin de comparar frecuencias esperadas y observadas. A partir de ellas se realizan diversas pruebas para corroborar la bondad de ajuste del modelo.

Paso 2. Eliminando la variable Tamaño de la Biblioteca.

Como se pudo observar en el cuadro de “Estimación de Parámetros” del Paso 1, hay dos variables que en ninguno de sus coeficientes resultan significativos para el modelo. Estas variables son “Tamaño de la biblioteca” y “Antigüedad de la Biblioteca”. Comparando los resultados se determina que la variable predictiva “Tamaño de biblioteca” será excluida en el nuevo modelo.

Entonces, se plantea un modelo con cinco variables independientes (todas las variables del grupo Características del MSD, excepto la de Tamaño de Biblioteca).

Antes de comenzar la construcción del modelo, se realiza la prueba de Líneas Paralelas, el siguiente cuadro presenta la información.

Prueba de Líneas Paralelas

Modelo	-2 Log Likelihood	Ji-Cuadrada	G.L.	P-value
Hipótesis Nula	537.130			
General	515.222	21.908	20	0.346

Función liga: Logit

Entonces se tiene que $\chi^2_{(20), \alpha=0.05} = 31.4104 > 21.908$,

con un $P - value = 0.346 > \alpha = 0.05$,

Por lo tanto no se considera que exista evidencia suficiente para rechazar la hipótesis nula y los datos se consideran adecuados para ser modelados mediante una regresión ordinal.

Posteriormente, se realizan las Pruebas de Cociente de Verosimilitud y Contraste de Máxima Verosimilitud.

Prueba de Verosimilitudes

Modelo	-2 Log Likelihood	Ji Cuadrada	G.L.	P-value
Nulo	597.178			
Final	537.130	60.048	10	0.000

Función liga: Logit

Se tiene que $C = 60.048 > \chi^2_{(15-5), \alpha=0.05} = 18.3070$

Con $P - value = 0.000 < \alpha = 0.05$, por lo tanto se rechaza la hipótesis nula.

Ahora, se compara mediante la Prueba de Contraste de Máxima Verosimilitud la significancia estadística del modelo con todas las variables del grupo de las Características del Encargado de Módulo vs el modelo con las variables predictivas: Experiencia/Tiempo en el PASDBP, Antigüedad de la Biblioteca, Computadoras Asignadas, Financiamiento y Sostenibilidad, que eliminó la variable Tamaño de la Biblioteca. Se calcula el estadístico:

$$G = (-2\text{Log}[Verosimilitud Modelo Anterior]) - (-2\text{Log}[Verosimilitud Modelo Nuevo])$$

Que se distribuye como $\chi^2_{(r), \alpha=0.05}$ donde r = número de variables nuevas.

$$\text{Entonces } G = 641.807 - 537.30 = 104.507 > 3.8415 = \chi^2_{(1), \alpha=0.05}$$

Por lo tanto se rechaza la hipótesis nula y se concluye que es mejor el nuevo modelo del desempeño del programa. A continuación la estimación de los parámetros β 's y α 's.

Estimación de los Parámetros

		Coeficientes	Std. Error	Wald	G.L.	P-value
α's	Bajo	-2.434	0.447	29.705	1	0.000
	Mediano	-1.170	0.435	7.235	1	0.007
	Alto	0.154	0.429	0.129	1	0.719
Experiencia/Tiempo en el PASDBP	Muy Poca	-0.978	0.326	9.008	1	0.003
	Poca	-0.872	0.291	8.976	1	0.003
	Mediana	-0.977	0.306	10.171	1	0.001
	Mucha	0 ^a	.	.	0	.
Antigüedad de la Biblioteca	Hasta 14 años	0.163	0.345	0.225	1	0.635
	Más 14 y hasta 20 años	-0.362	0.314	1.329	1	0.249
	Más de 20 y hasta 23 años	0.061	0.299	0.042	1	0.838
	23 años y más	0 ^a	.	.	0	.

Computadoras Asignadas	2 -3 computadoras	-1.653	0.366	20.444	1	0.000
	4- 6 computadoras	-0.873	0.338	6.692	1	0.010
	9- 13 computadoras	0 ^a	.	.	0	.
Sostenibilidad	Cuenta con plan para continuar	0.436	0.212	4.222	1	0.040
	No Cuenta con plan para continuar	0 ^a	.	.	0	.
Financiamiento	Si	0.686	0.236	8.418	1	0.004
	No	0 ^a	.	.	0	.

Función liga: Logit

a. Categoría de comparación.

Al observar el cuadro anterior, los coeficientes de la Antigüedad de la Biblioteca siguen siendo irrelevantes para el modelo. Ya que el coeficiente de las categorías: “Hasta 14 años” tiene es $\chi^2_{(1)}, \alpha=0.05 = 3.8415 > 0.225$ con un $P - value = 0.635 > 0.005 = \alpha$,

Para la categoría “De 14 a 20 años” tiene es $\chi^2_{(1)}, \alpha=0.05 = 3.8415 > 1.329$ con un $P - value = 0.249 > 0.05 = \alpha$

Y para la categoría “De 20 a 23 años” tiene $\chi^2_{(1)}, \alpha=0.05 = 3.8415 > 0.042$ con un $P - value = 0.838 > 0.05 = \alpha$.

Lo anterior nos lleva a concluir que en ninguno de ellos se presenta evidencia suficiente para rechazar la hipótesis nula.

Paso 3. Eliminando la variable Antigüedad de la Biblioteca

Se decide plantear un nuevo modelo que contenga las siguientes variables: Experiencia/Tiempo en el PASDBP, Computadoras Asignadas, Financiamiento y Sostenibilidad.

Siguiendo el procedimiento definido se procede a evaluar el cumplimiento del supuesto de Líneas paralelas.

Prueba de Líneas Paralelas

Modelo	-2 Log Likelihood	Ji-Cuadrada	G.L.	P-value
Hipótesis Nula	335.419			
General	314.324	21.095	14	0.099

Función liga: Logit

Entonces se tiene que $\chi^2_{(14)}, \alpha=0.05 = 23.6848 > 21.095$ con un $P - value = 0.099 > \alpha = 0.05$

Por lo que no se tiene evidencia suficiente para rechazar la hipótesis nula y los datos cumplen el supuesto de Líneas Paralelas.

Posteriormente, se realizan las Pruebas de Cociente de Verosimilitud y Contraste de Máxima Verosimilitud.

Prueba de Verosimilitudes

Modelo	-2 Log Likelihood	Ji-Cuadrada	G.L.	P-value
Nulo	385.719			
Final	335.419	50.299	7	0.000

Función liga: Logit

Se tiene que $C = 50.299 > \chi^2_{(11-4), \alpha=0.05} = 14.0671$

Con un $P - value = 0.000 < \alpha = 0.05$,

Por lo tanto se rechaza la hipótesis nula y se afirma que el modelo de regresión ordinal tiene sentido.

Ahora, mediante la Prueba de Contraste de Máxima Verosimilitud se compara la significancia estadística del modelo con las variables predictivas: Experiencia/Tiempo en el PASDBP, Antigüedad de la Biblioteca, Computadoras Asignadas, Financiamiento y sostenibilidad) vs el modelo con las variables predictivas: Experiencia/Tiempo en el PASDBP, Computadoras Asignadas, Financiamiento y sostenibilidad. Gracias a la información del cuadro de Cociente de Verosimilitudes del Paso 2 y 3, se calcula el estadístico:

$$G = (-2\text{Log}[Verosimilitud Modelo Anterior]) - (-2\text{Log}[Verosimilitud Modelo Nuevo])$$

Que se distribuye como $\chi^2_{(r), \alpha=0.05}$ donde $r =$ número de variables nuevas.

$$\text{Entonces } G = 537.130 - 335.419 = 201.711 > 3.8415 = \chi^2_{(1), \alpha=0.05}$$

Por lo tanto se rechaza la hipótesis nula y se concluye que es mejor el nuevo modelo del desempeño del programa.

Mediante el método de Máxima Verosimilitud se estiman los coeficientes $\beta's$ y $\alpha's$ con los que se obtiene el modelo de regresión ordinal. A continuación, se presenta un cuadro con la información para realizar la *Prueba de Wald* y determinar cuáles de ellos son significativos para el modelo.

Estimación de los Parámetros

		Coeficientes	Std. Error	Wald	G.L.	P-value
α's	Bajo	-2.499	0.393	40.479	1	0.000
	Mediano	-1.278	0.382	11.212	1	0.001
	Alto	-.014	0.375	0.001	1	0.970
Experiencia/Tiempo en el PASDBP	Muy Poca	-.924	0.286	10.454	1	0.001
	Poca	-.759	0.262	8.414	1	0.004
	Mediana	-.803	0.273	8.657	1	0.003
	Mucha	0 ^a	.	.	0	.
Computadoras Asignadas	2 -3 computadoras	-1.581	0.335	22.224	1	0.000
	4- 6 computadoras	-.972	0.322	9.115	1	0.003
	9- 13 computadoras	0 ^a	.	.	0	.
Sostenibilidad	Cuenta con plan para continuar	0.381	0.189	4.050	1	0.044
	No Cuenta con plan para continuar	0 ^a	.	.	0	.
Financiamiento	Si	0.440	0.212	4.300	1	0.038
	No	0 ^a	.	.	0	.

Función liga: Logit

a. Categoría de comparación.

Para determinar qué valores tomarán los parámetros en los modelos se utiliza la Prueba de Wald y se obtienen los siguientes resultados en la Prueba de Wald:

- Son estadísticamente significativos, y diferentes de cero, todos los coeficientes de todas las variables.

Para determinar el valor de los umbrales α 's (del desempeño del PASDBP) también se realiza la Prueba de Wald y nuevamente resultan estadísticamente significativos los umbrales de los niveles de desempeño del PASDBP: Bajo y Medio.

Posteriormente, se definen los predictores lineales con los coeficientes que resultaron significativos para cada variable (en este caso todos resultaron significativos) como:

$$\begin{aligned}
 f(\theta_1(X)) &= f(P(Y \leq Y_1 | X_{E1}, X_{E2}, X_{E3}, X_{NC1}, X_{NC2}, X_F, X_S)) \\
 &= -2.499 + 0.924X_{E1} + 0.759X_{E2} + 0.803X_{E3} + 1.581X_{NC1} + 0.972X_{NC2} \\
 &\quad - 0.381X_S - 0.440X_F
 \end{aligned}$$

$$\begin{aligned}
f(\theta_2(X)) &= f(P(Y \leq Y_2 | X_{E1}, X_{E2}, X_{E3}, X_{NC1}, X_{NC2}, X_F, X_S)) \\
&= -1.278 + 0.924X_{E1} + 0.759X_{E2} + 0.803X_{E3} + 1.581X_{NC1} + 0.972X_{NC2} \\
&\quad - 0.381X_S - 0.440X_F
\end{aligned}$$

$$\begin{aligned}
f(\theta_3(X)) &= f(P(Y \leq Y_3 | X_{E1}, X_{E2}, X_{E3}, X_{NC1}, X_{NC2}, X_F, X_S)) \\
&= -0 + 0.924X_{E1} + 0.759X_{E2} + 0.803X_{E3} + 1.581X_{NC1} + 0.972X_{NC2} \\
&\quad - 0.381X_S - 0.440X_F
\end{aligned}$$

Después de explicar en el Paso 1 como se define cada una de las categorías de las variables predictoras, se construye el modelo de regresión ordinal para las categorías $Y_1 = \text{Bajo}$, $Y_2 = \text{Mediano}$, $Y_3 = \text{Alto}$ y $Y_4 = \text{Excelente}$, y se definen como:

$$\begin{aligned}
P(Y = Y_1 | X_{E1}, X_{E2}, X_{E3}, X_{NC1}, X_{NC2}, X_F, X_S) \\
= \left[\frac{1}{1 + e^{-(-2.499 + 0.924X_{E1} + 0.759X_{E2} + 0.803X_{E3} + 1.581X_{NC1} + 0.972X_{NC2} - 0.381X_S - 0.440X_F)}} \right]
\end{aligned}$$

$$\begin{aligned}
P(Y = Y_2 | X_{E1}, X_{E2}, X_{E3}, X_{NC1}, X_{NC2}, X_F, X_S) \\
= \left[\frac{1}{1 + e^{-(-1.278 + 0.924X_{E1} + 0.759X_{E2} + 0.803X_{E3} + 1.581X_{NC1} + 0.972X_{NC2} - 0.381X_S - 0.440X_F)}} \right] \\
- P(Y = Y_1 | X_{E1}, X_{E2}, X_{E3}, X_{NC1}, X_{NC2}, X_F, X_S)
\end{aligned}$$

$$\begin{aligned}
P(Y = Y_3 | X_{E1}, X_{E2}, X_{E3}, X_{NC1}, X_{NC2}, X_F, X_S) \\
= \left[\frac{1}{1 + e^{-(-0 + 0.924X_{E1} + 0.759X_{E2} + 0.803X_{E3} + 1.581X_{NC1} + 0.972X_{NC2} - 0.381X_S - 0.440X_F)}} \right] \\
- P(Y = Y_2 | X_{E1}, X_{E2}, X_{E3}, X_{NC1}, X_{NC2}, X_F, X_S)
\end{aligned}$$

$$\begin{aligned}
P(Y = Y_4 | X_{E1}, X_{E2}, X_{E3}, X_{NC1}, X_{NC2}, X_F, X_S) \\
= 1 - P(Y = Y_3 | X_{E1}, X_{E2}, X_{E3}, X_{NC1}, X_{NC2}, X_F, X_S)
\end{aligned}$$

Para ejemplificar el cálculo de una probabilidad condicional se obtiene la probabilidad de que un MSD tenga un bajo desempeño, dado que se tiene poca experiencia en el PASDBP, la biblioteca donde está instalado el módulo tiene 23 años o más de antigüedad, el espacio donde se encuentra el MSD no es adecuado al tamaño de la biblioteca, se le asignaron dos computadoras al módulo y no tiene financiamiento ni sostenibilidad:

$$\begin{aligned}
P(Y = Y_1 | X_{E1} = 0, X_{E2} = 1, X_{E3} = 0, X_{AB1} = 0, X_{AB2} = 0, X_{AB3} = 0, X_{TB} = 0, X_{NC1} = 1, X_{NC2} = 0, X_F = 0, X_S = 0) \\
= \left[\frac{1}{1 + e^{-(-2.356 + 0.992(0) + 0.883(1) + 0.981(0) - 0 - 0 - 0 + 1.625(1) + 0.857(0) - 0.431(0) - 0.682(0))}} \right] = 0.5379
\end{aligned}$$

A continuación se presenta un cuadro con los valores de las *pseudo – R²'s de Nagelkerke* de los modelos anteriormente planteados, y que fueron desechados en el desarrollo del procedimiento “Backward Stepwise”.

Variabes Predictivas del Modelo	R² de Nagelkerke
Experiencia/ Tiempo en el PASDBP, Antigüedad de la Biblioteca, Tamaño de la Biblioteca, Computadoras Asignadas, Financiamiento y Sostenibilidad	0.181
Experiencia/ Tiempo en el PASDBP, Antigüedad de la Biblioteca, Computadoras Asignadas, Financiamiento y Sostenibilidad	0.180
Experiencia/ Tiempo en el PASDBP, Computadoras Asignadas, Financiamiento y Sostenibilidad	0.128

Como se observa en el cuadro, la medida porcentual que nos da la *R² de Nagelkerke* comenzó en 18.1%, un porcentaje que puede ser considerado como bueno dentro de las escala que maneja el *R² de Nagelkerke*.

Puede observarse que la disminución menos significativa ocurre cuando se excluye a la variable Tamaño de la Biblioteca con un 0.1%. Por otro lado, la variable Antigüedad de la Biblioteca provoca la disminución del porcentaje más perceptible con un 5.2% abajo después de su salida. En realidad, el porcentaje a valorar la variabilidad del modelo explicado por las variables independientes (calculando el incremento de la verosimilitud) descendió porcentualmente en un 29.28% respecto al inicial¹⁵⁴.

Y a pesar de que no se obtienen coeficientes de *R² de Nagelkerke* muy altos, se considera que las características del MSD contribuyen en alguna medida el desempeño del PASDBP y se considera este modelo final como adecuadamente ajustado y justificado por las pruebas correspondientes no sólo a la bondad de ajuste sino a la verosimilitud comparada con los otros modelos propuestos.

iii. Modelo con las características del encargado de módulo

El siguiente modelo a obtener es aquél que involucra como variables predictivas a las variables correspondientes al grupo de Características del Encargado del MSD siguiendo el procedimiento anteriormente descrito.

Paso 1. Modelo Completo

Para la construcción del modelo de Características del encargado de módulo se incluyen las cuatro variables del grupo, las cuales son: Tipo de encargado, Tiempo de trabajar del encargado, Escolaridad del encargado y Capacitación por parte de la DGB.

¹⁵⁴ Se tomó como la variabilidad inicial del modelo 0.181, que desciende hasta un 0.128.

Primero se debe evaluar el cumplimiento del supuesto de Líneas paralelas.

Prueba de Líneas Paralelas

Modelo	-2 Log Likelihood	Ji-Cuadrado	G.L.	Sig.
Hipótesis Nula	401.044			
General	381.073	19.971	16	0.222

Función liga: Logit.

Se tiene que una $\chi^2_{(16)}, \alpha=0.05 = 26.2962 > 19.971$,

con un $P - value = 0.222 > \alpha = 0.05$

Por lo tanto se concluye que no se tiene evidencia suficiente para rechazar la hipótesis nula y se considera que los datos cumplen el supuesto de Líneas paralelas.

De acuerdo a la conclusión anterior, se prosigue con la Prueba de Verosimilitudes.

Prueba de Verosimilitudes

Modelo	-2 Log Likelihood	Ji-Cuadrada	G.L.	P-value
Nulo	467.791			
Final ¹⁵⁵	401.044	66.747	8	0

Función liga: Logit

De acuerdo a la información de la tabla, se tiene que $C = 66.747 > \chi^2_{(12-4)}, \alpha=0.05 = 15.5073$

con un $P - value = 0.000 < \alpha = 0.05$, por lo que se rechaza la hipótesis nula y se determina que el modelo tiene sentido, es decir, justifica que alguna de las variables explicativas es importante para explicar el fenómeno en estudio.

A continuación se presenta un cuadro con la información obtenida mediante del Método de Máxima Verosimilitud.

Estimación de los Parámetros

		Coefficientes	Std. Error	Wald	G.L.	P-value
Desempeño del PASDBP	Bajo	-1.709	0.312	29.916	1	0.000
	Mediano	-0.403	0.302	1.784	1	0.182

¹⁵⁵ Modelo que incluye las variables predictoras especificadas.

	Alto	0.799	0.304	6.912	1	0.009
Tipo de Encargado	EM	0.621	0.226	7.538	1	0.006
	EM-B	0 ^a	.	.	0	.
Tiempo de trabajar del Encargado	Menos de un año	0.049	0.316	0.024	1	0.877
	De 1 y menor a 2.5 años	0.846	0.306	7.634	1	0.006
	De 2.5 y menor a 9 años	0.23	0.302	0.581	1	0.446
	9 años o más	0 ^a	.	.	0	.
Escolaridad del Encargado	Educación Básica	-1.145	0.328	12.177	1	0.000
	Educación Media Superior	-0.838	0.224	14.018	1	0.000
	Educación Superior	0 ^a	.	.	0	.
Capacitación por parte de la DGB	0 - 1 curso	-0.907	0.277	10.755	1	0.001
	2 - 4 cursos	-0.452	0.216	4.384	1	0.036
	5 - 7 cursos	0 ^a	.	.	0	.

Función liga: Logit.

a. Categoría de comparación.

Se utiliza la Prueba de Wald con las siguientes hipótesis:

$$H_0: \beta_i = 0 \quad vs \quad H_1: \text{al menos una } \beta_i \neq 0$$

Con $i = 1, 2, \dots, 8$ donde:

β_1 = Coeficiente de la variable Tipo de Encargado de la categoría EM

β_2 = Coeficiente de la variable Tiempo de trabajar del encargado la categoría Menos de un año

β_3 = Coeficiente de la variable Tiempo de trabajar del encargado la categoría De 1 y menor a 2.5 años

β_4 = Coeficiente de la variable Tiempo de trabajar del encargado la categoría De 2.5 y menor a 9 años

β_5 = Coeficiente de la variable Escolaridad del encargado de la categoría Educación Básica

β_6 = Coeficiente de la variable Escolaridad del encargado de la categoría Educación Media Superior

β_7 = Coeficiente de la variable Capacitación por parte de la DGB de la categoría 0-1 curso

β_8 = Coeficiente de la variable Capacitación por parte de la DGB de la categoría 2-4 cursos

Como se puede observar en la tabla anterior, se obtienen los siguientes resultados en la Prueba de Wald:

- Son estadísticamente significativos, y diferentes de cero, los coeficientes de:

- Las dos categorías de la variable Escolaridad del encargado (Educación Básica y Educación Media Superior)
- Las dos categorías de la variable Capacitación por parte de la DGB (0-1 curso, 2-4 cursos)
- La categoría De 1 y menor a 2.5 años de la variable Tiempo de trabajar del encargado
- La categoría EM de la variable Tipo de Encargado
- No son estadísticamente significativos los coeficientes de:
 - Dos categorías de la variable Tiempo de trabajar del encargado (Menos de 1 año, De 2.5 y menor a 9 años)

Y para determinar el valor de los umbrales α 's (del desempeño del PASDBP) también se realiza la Prueba de Wald, cuyo estadístico se distribuye $\chi^2_{(1)}$, $\alpha=0.05$, con las siguientes hipótesis:

$$H_0: \alpha_i = 0 \quad vs \quad H_1: \alpha_i \neq 0$$

Con $i = 1, 2, 3$ dónde:

α_1 = Umbral (constante) de Nivel de Desempeño Bajo
 α_2 = Umbral (constante) de Nivel de Desempeño Mediano
 α_3 = Umbral (constante) de Nivel de Desempeño Alto

- Como se puede observar en la tabla, resultan estadísticamente significativas los umbrales de los niveles Bajo y Alto de desempeño del PASDBP.

Después de analizar mediante la Prueba de Wald qué coeficientes son significativos, las probabilidades quedan definidas como los siguientes predictores lineales:

$$\begin{aligned} f(\theta_1(X)) &= f(P(Y \leq Y_1 | X_{TE}, X_{TT1}, X_{TT2}, X_{TT3}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})) \\ &= \alpha_1 - \beta_{TE}X_{TE} - \beta_{TT1}X_{TT1} - \beta_{TT2}X_{TT2} - \beta_{TT3}X_{TT3} - \beta_{EE1}X_{EE1} - \beta_{EE2}X_{EE2} \\ &\quad - \beta_{C1}X_{C1} - \beta_{C2}X_{C2} \\ &= -1.709 - 0.621X_{TE} - 0X_{TT1} - 0.846X_{TT2} - 0X_{TT3} + 1.145X_{EE1} \\ &\quad + 0.838X_{EE2} + 0.907X_{C1} + 0.452X_{C2} \end{aligned}$$

$$\begin{aligned} f(\theta_2(X)) &= f(P(Y \leq Y_2 | X_{TE}, X_{TT1}, X_{TT2}, X_{TT3}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})) \\ &= \alpha_2 - \beta_{TE}X_{TE} - \beta_{TT1}X_{TT1} - \beta_{TT2}X_{TT2} - \beta_{TT3}X_{TT3} - \beta_{EE1}X_{EE1} - \beta_{EE2}X_{EE2} \\ &\quad - \beta_{C1}X_{C1} - \beta_{C2}X_{C2} \\ &= -0 - 0.621X_{TE} - 0X_{TT1} - 0.846X_{TT2} - 0X_{TT3} + 1.145X_{EE1} + 0.838X_{EE2} \\ &\quad + 0.907X_{C1} + 0.452X_{C2} \end{aligned}$$

$$\begin{aligned} f(\theta_3(X)) &= f(P(Y \leq Y_3 | X_{TE}, X_{TT1}, X_{TT2}, X_{TT3}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})) \\ &= \alpha_3 - \beta_{TE}X_{TE} - \beta_{TT1}X_{TT1} - \beta_{TT2}X_{TT2} - \beta_{TT3}X_{TT3} - \beta_{EE1}X_{EE1} - \beta_{EE2}X_{EE2} \\ &\quad - \beta_{C1}X_{C1} - \beta_{C2}X_{C2} \\ &= 0.799 - 0.621X_{TE} - 0X_{TT1} - 0.846X_{TT2} - 0X_{TT3} + 1.145X_{EE1} + 0.838X_{EE2} \\ &\quad + 0.907X_{C1} + 0.452X_{C2} \end{aligned}$$

Donde:

α_i = Umbral (constante) de los Niveles de Desempeño con $i = 1,2,3$.

Para 1=Bajo, 2= Mediano y 3=Alto

β_{TE} = Coeficiente de la variable Tipo de encargado de la categoría EM

β_{TTi} = Coeficiente de la variable Tiempo de trabajar del encargado con $i = 1,2,3$.

Para 1= Menos 1 año, 2= De 1 y menor a 2.5 años y 3= De 2.5 y menor de 9 años

β_{EEi} = Coeficiente de la variable Escolaridad del encargado con $i = 1,2$

Para 1= Educación Básica, 2= Educación Media Superior

β_{Ci} = Coeficiente de la variable Capacitación por parte de la DGB con $i = 1,2$

Para 1= 0-1 curso, 2= 2-4 cursos

X_{TE} = Variable Tipo de encargado de la categoría EM

X_{TTi} = Variable Tiempo de trabajar del encargado

X_{EEi} = Variable Escolaridad del encargado

X_{Ci} = Variable Capacitación por parte de la DGB

Así que X_{TE} puede tomar los valores 0 y 1, dependiendo al Tipo de Encargado al que se refiera. Por ejemplo, si $X_{TE} = 1$ entonces el tipo de encargado es EM (exclusivo del MSD) y si $X_{TE} = 0$ entonces el tipo de encargado es EM-B.

De forma similar la variable independiente Tiempo de trabajar del encargado puede expresar sus cuatro categorías con las siguientes combinaciones de X_{TT1} , X_{TT2} y X_{TT3} .

- Si $X_{TT1} = 1, X_{TT2} = 0$ y $X_{TT3} = 0$, entonces se refiere a que el encargado de módulo ha trabajado menos de un año.
- Si $X_{TT1} = 0, X_{TT2} = 1$ y $X_{TT3} = 0$, se refiere a que el encargado de módulo ha trabajado de 1 y hasta menos 2.5 años.
- Si $X_{TT1} = 0, X_{TT2} = 0$ y $X_{TT3} = 1$, entonces se refiere a que el encargado de módulo ha trabajado de 2.5 y menos de 9 años.
- Y por último si $X_{TT1} = 0, X_{TT2} = 0$ y $X_{TT3} = 0$ se refiere a que el encargado de módulo ha trabajado 9 años o más

Con la variable Escolaridad del Encargado se tiene X_{EE1} y X_{EE2} donde $X_{EEi} = 1,0$

con $i = 1,2$.

- Si queremos indicar que el encargado tiene Educación Básica, entonces $X_{EE1} = 1$ y $X_{EE2} = 0$;
- si el encargado tiene Educación Media Superior entonces $X_{EE1} = 0$ y $X_{EE2} = 1$.
- Finalmente si se desea indicar que el encargado tiene Educación Superior entonces $X_{EE1} = 0$ y $X_{EE2} = 0$.

Y para la variable Capacitación se tiene X_{C1} y X_{C2} donde $X_{Ci} = 1,0$ con $i = 1,2$.

- Si queremos indicar que el encargado tomó de 0-1 curso por parte de la DGB, entonces $X_{C1} = 1$ y $X_{C2} = 0$.

- Si el encargado tomó de 2-4 cursos entonces $X_{C1} = 0$ y $X_{C2} = 1$.
- Si se desea indicar que el encargado tomó de 5-7 cursos por parte de la DGB se tiene $X_{C1} = 0$ y $X_{C2} = 0$.

Una vez explicadas cada una de las categorías de las variables, se construye el modelo de regresión ordinal para las categorías $Y_1 = \text{Bajo}$, $Y_2 = \text{Medio}$, $Y_3 = \text{Alto}$ y $Y_4 = \text{Excelente}$ y se definen como las siguientes probabilidades absolutas (obtenidas a partir de las probabilidades acumuladas):

$$P(Y = Y_1 | X_{TE}, X_{TT1}, X_{TT2}, X_{TT3}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2}) = \left[\frac{1}{1 + e^{-(-1.709 - 0.621X_{TE} - 0X_{TT1} - 0.846X_{TT2} - 0X_{TT3} + 1.145X_{EE1} + 0.838X_{EE2} + 0.907X_{C1} + 0.452X_{C2})}} \right]$$

$$P(Y = Y_2 | X_{TE}, X_{TT1}, X_{TT2}, X_{TT3}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2}) = \left[\frac{1}{1 + e^{-(-0.621X_{TE} - 0X_{TT1} - 0.846X_{TT2} - 0X_{TT3} + 1.145X_{EE1} + 0.838X_{EE2} + 0.907X_{C1} + 0.452X_{C2})}} \right] - P(Y = Y_1 | X_{TE}, X_{TT1}, X_{TT2}, X_{TT3}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})$$

$$P(Y = Y_3 | X_{TE}, X_{TT1}, X_{TT2}, X_{TT3}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2}) = \left[\frac{1}{1 + e^{-(0.799 - 0.621X_{TE} - 0X_{TT1} - 0.846X_{TT2} - 0X_{TT3} + 1.145X_{EE1} + 0.838X_{EE2} + 0.907X_{C1} + 0.452X_{C2})}} \right] - P(Y = Y_2 | X_{TE}, X_{TT1}, X_{TT2}, X_{TT3}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})$$

$$P(Y = Y_4 | X_{TE}, X_{TT1}, X_{TT2}, X_{TT3}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2}) = 1 - P(Y = Y_3 | X_{TE}, X_{TT1}, X_{TT2}, X_{TT3}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})$$

Para ejemplificar el cálculo de las probabilidades condicionales, se obtiene la probabilidad de tener un desempeño excelente del PASDBP dado que se tiene un Encargado de módulo exclusivo para el MSD (EM), que lleva menos de un año trabajando, que tiene una escolaridad de Educación Superior y que tomó todos los cursos otorgados por la DGB es:

$$P(Y = Y_4 | X_{TE} = 1, X_{TT1} = 1, X_{TT2} = 0, X_{TT3} = 0, X_{EE1} = 0, X_{EE2} = 0, X_{C1} = 0, X_{C2} = 0) = 1 - \left[\frac{1}{1 + e^{-(0.799 - 0.621(1) - 0(1) - 0.846(0) - 0 + 1.145(0) + 0.838(0) + 0.907(0) + 0.452(0))}} \right] = 0.456$$

Con las probabilidades absolutas, como la que se calculó en el ejemplo anterior, se construyen las tablas de frecuencias esperadas y observadas, con las cuales se realizan diversas pruebas para corroborar la bondad de ajuste del modelo.

Paso 2. Eliminando la variable Tiempo de trabajar del encargado.

Como se muestra en el cuadro anterior, la variable Escolaridad del encargado presenta en los coeficientes de sus dos categorías Educación Básica y Educación Medio Superior un $P - value =$

0.000. Con ello no solamente se corrobora el rechazo de la hipótesis nula, sino que al ser los *P – value's* más pequeños se considera a la Escolaridad del encargado del modulo la variable más significativa en el modelo con las cuatro variables predictoras de las Características del Encargado del MSD.

En último lugar de significancia se tiene a la variable Tiempo de trabajar del encargado, ya que en dos de las tres pruebas de Wald que se realizaron para dicha variable predictiva, dos coeficientes resultaron no ser significativos para incluirse en el modelo.

Ya que se han definido las variables del nuevo modelo: Escolaridad del Encargado, Tipo de Encargado y Capacitación por parte de la DGB; se realiza la prueba de líneas paralelas, para determinar si el método de regresión ordinal es el indicado para analizar los datos. El siguiente cuadro presenta la información necesaria para llevarla a cabo.

Prueba de Líneas Paralelas

Modelo	-2 Log Likelihood	Ji-Cuadrada	G.L.	P-value
Hipótesis Nula	185.176			
General	175.084	10.092	10	0.432

Función liga: Logit

Entonces se tiene que $\chi^2_{(10), \alpha=0.05} = 18.3070 > 10.092,$

con un $P – value = 0.432 > \alpha = 0.05,$

Por lo que se concluye que no se tiene evidencia suficiente para rechazar la hipótesis nula y por lo tanto los datos pueden ser modelados mediante una regresión ordinal.

Por consiguiente, se realiza la Prueba de Verosimilitudes para determinar si estimar el nuevo modelo con las tres variables tiene sentido.

Prueba de Verosimilitudes

Modelo	-2 Log Likelihood	Ji-Cuadrada	G.L.	P-value
Nulo	253.797			
Final	185.176	68.62	5	0.000

Función liga: Logit

Entonces se tiene que $C = 68.62 > \chi^2_{(8-3), \alpha=0.05} = 11.0705$

con un $P – value = 0.000 < \alpha = 0.05,$

Por lo que se rechaza la hipótesis nula y se puede decir que la estimación del modelo de regresión ordinal con las variables: Escolaridad del Encargado, Tipo de Encargado y Capacitación por parte de la DGB tiene sentido.

Ahora, se compara mediante la Prueba de Contraste de Máxima Verosimilitud la significancia estadística del modelo con todas las variables del grupo de las Características del Encargado de Módulo vs el modelo nuevo (con las variables predictivas: Escolaridad del Encargado, Tipo de Encargado y Capacitación por parte de la DGB). Gracias a la información del cuadro de Cociente de Verosimilitudes del Paso 1 y 2, se calcula el estadístico:

$$G = (-2\text{Log}[Verosimilitud Modelo Anterior]) - (-2\text{Log}[Verosimilitud Modelo Nuevo])$$

Que se distribuye como $\chi^2_{(r)}$, $\alpha=0.05$ donde $r = \text{número de variables nuevas}$.

Entonces $G = 401.044 - 185.176 = 215.868 > 3.8415 = \chi^2_{(1)}$, $\alpha=0.05$, por lo tanto se rechaza la hipótesis nula y se concluye que es mejor el modelo del desempeño del programa con las variables de Escolaridad del Encargado, Tipo de Encargado y Capacitación por parte de la DGB¹⁵⁶.

A continuación, se presenta un cuadro con la estimación de los parámetros:

Estimación de los Parámetros

		Coeficientes	Std. Error	Wald	G.L.	P-value
Desempeño del PASDBP	Bajo	-1.769	0.265	44.393	1	0.000
	Mediano	-0.485	0.254	3.658	1	0.056
	Alto	0.727	0.255	8.117	1	0.004
Tipo de Encargado	EM	0.913	0.186	24.175	1	0.000
	EM-B	0 ^a	.	.	0	.
Escolaridad del Encargado	Educación Básica	-1.012	0.3	11.403	1	0.001
	Educación Media Superior	-0.659	0.21	9.859	1	0.002
	Educación Superior	0 ^a	.	.	0	.
Capacitación por parte de la DGB	0 - 1 curso	-1.083	0.243	19.798	1	0
	2 - 4 cursos	-0.517	0.204	6.399	1	0.011
	5 - 7 cursos	0 ^a	.	.	0	.

Función liga: Logit.

¹⁵⁶ La tabla de contingencia para este modelo tienen teóricamente tiene (18 fila * 4 columnas) = 72 celdas. En este caso, la tabla de contingencia presenta 6 celdas (8.3%) con frecuencia igual cero, es decir, existen 6 combinaciones del vector $(Y_j, X_{TE}, X_{EE}, X_C)$ con frecuencia nula. Lo que nos dice que no es un problema el número de celdas vacías al ser baja su proporción.

a. Categoría de comparación.

Repitiendo el procedimiento del Paso 1, se realiza la Prueba de Wald con las siguientes hipótesis:

$$H_0: \beta_i = 0 \quad vs \quad H_1: \text{al menos una } \beta_i \neq 0$$

Con $i = 1, 2, \dots, 5$ donde:

β_1 = Coeficiente de la variable Tipo de Encargado de la categoría EM

β_2 = Coeficiente de la variable Escolaridad del encargado de la categoría Educación Básica

β_3 = Coeficiente de la variable Escolaridad del encargado de la categoría Educación Media Superior

β_4 = Coeficiente de la variable Capacitación por parte de la DGB de la categoría 0-1 curso

β_5 = Coeficiente de la variable Capacitación por parte de la DGB de la categoría 2-4 cursos

Se obtienen los siguientes resultados en la Prueba de Wald:

- Son estadísticamente significativos, y diferentes de cero, todos los coeficientes de todas las variables anteriores.

Y para determinar el valor de los umbrales α 's (del desempeño del PASDBP) también se realiza la Prueba de Wald, cuyo estadístico se distribuye $\chi^2_{(1)}$, $\alpha=0.05$, con las siguientes hipótesis:

$$H_0: \alpha_i = 0 \quad vs \quad H_1: \alpha_i \neq 0$$

Con $i = 1, 2, 3$ dónde:

α_1 = Umbral (constante) de Nivel de Desempeño Bajo

α_2 = Umbral (constante) de Nivel de Desempeño Mediano

α_3 = Umbral (constante) de Nivel de Desempeño Alto

- Como se puede observar en la tabla, resultan estadísticamente significativas los umbrales de los niveles Bajo y Alto de desempeño del PASDBP.

Mediante la Prueba de Wald se determinó qué coeficientes son significativos, por lo que se puede definir las probabilidades condicionales como los siguientes predictores lineales:

$$\begin{aligned} f(\theta_1(X)) &= f(P(Y \leq Y_1 | X_{TE}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})) \\ &= -1.769 - 0.913X_{TE} + 1.012X_{EE1} + 0.659X_{EE2} + 1.083X_{C1} + 0.517X_{C2} \end{aligned}$$

$$\begin{aligned} f(\theta_2(X)) &= f(P(Y \leq Y_2 | X_{TE}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})) \\ &= 0 - 0.913X_{TE} + 1.012X_{EE1} + 0.659X_{EE2} + 1.083X_{C1} + 0.517X_{C2} \end{aligned}$$

$$f(\theta_3(X)) = f(P(Y \leq Y_3 | X_{TE}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2}))$$

$$= 0.727 - 0.913X_{TE} + 1.012X_{EE1} + 0.659X_{EE2} + 1.083X_{C1} + 0.517X_{C2}$$

Ya que se ha explicado a detalle en el Paso anterior cómo es que se define cada una de las categorías de las variables, se construye el modelo de regresión ordinal para los niveles del Desempeño del PASDBP como:

$$P(Y = Y_1 | X_{TE}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})$$

$$= \left[\frac{1}{1 + e^{-(-1.769 - 0.913X_{TE} + 1.012X_{EE1} + 0.659X_{EE2} + 1.083X_{C1} + 0.517X_{C2})}} \right]$$

$$P(Y = Y_2 | X_{TE}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})$$

$$= \left[\frac{1}{1 + e^{-(-0.913X_{TE} + 1.012X_{EE1} + 0.659X_{EE2} + 1.083X_{C1} + 0.517X_{C2})}} \right]$$

$$- P(Y = Y_1 | X_{TE}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})$$

$$P(Y = Y_3 | X_{TE}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})$$

$$= \left[\frac{1}{1 + e^{-(0.727 - 0.913X_{TE} + 1.012X_{EE1} + 0.659X_{EE2} + 1.083X_{C1} + 0.517X_{C2})}} \right]$$

$$- P(Y = Y_2 | X_{TE}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})$$

$$P(Y = Y_4 | X_{TE}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2}) = 1 - P(Y = Y_3 | X_{TE}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})$$

Con estas probabilidades acumuladas y absolutas, se puede calcular cualquier combinación del vector $(Y_j, X_{TE}, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})$. Por ejemplo, se quiere calcular la probabilidad de obtener un desempeño medio o menos dado que el tipo de encargado es EM-B, que cuenta Educación Básica y que ha asistido a todos los cursos por parte de la DGB es:

$$P(Y \leq Y_2 | X_{TE} = 0, X_{EE1} = 1, X_{EE2} = 0, X_{C1} = 0, X_{C2} = 0)$$

$$= \left[\frac{1}{1 + e^{-(-0.913(0) + 1.012(1) + 0.659(0) + 1.083(0) + 0.517(0))}} \right] = 0.733$$

Como un análisis complementario se presenta el siguiente cuadro que muestra los valores de las *pseudo* - R^2 's de los modelos anteriormente planteados.

Variables Predictivas del Modelo	R ² de Nagelkerke
Tipo de Encargado, Tiempo de trabajar del encargado, Escolaridad del Encargado y Capacitación por parte de la DGB.	0.171
Tipo de Encargado, Escolaridad del Encargado y Capacitación por parte de la DGB.	0.159

Se observa una disminución en el porcentaje obtenido de la R^2 de Nagelkerke cuando se excluye la variable Tiempo de trabajar del encargado, del segundo modelo, en un 1.2%.

En resumen, se considera bueno el porcentaje en la escala de la *pseudo* – R^2 's tomando en cuenta el número de variables¹⁵⁷ que forman parte del modelo final para el bloque de Características del Encargado de Módulo. Por lo que se considera adecuadamente ajustado y justificado por las pruebas correspondientes no sólo a la bondad de ajuste sino a la verosimilitud comparada con los otros modelos propuestos.

iv. Modelo con la difusión

Para la construcción del último modelo grupal que corresponde al bloque de Difusión, no será necesario establecer ningún criterio de selección de variables predictivas porque únicamente contiene la variable independiente “Difusión”.

Así que comenzamos la modelación, con la Prueba de Líneas Paralelas.

Prueba de Líneas Paralelas

Modelo	-2 Log Likelihood	Ji-Cuadrada	G.L.	P-value
Hipótesis Nula	30.388			
General	30.017	0.371	2	0.830

Función liga: Logit.

Obsérvese que $\chi^2_{(2)}, \alpha=0.05 = 5.9915 > 0.371$, por lo que la prueba de líneas paralelas muestra que no se rechaza la hipótesis nula con un $P - value = 0.830 > \alpha = 0.05$. Y se concluye que se cumple el supuesto de Líneas Paralelas.

El siguiente paso, es aplicar la Prueba de Máxima Verosimilitud.

Cociente de Máxima Verosimilitud

Modelo	-2 Log Likelihood	Ji-Cuadrada	G.L.	P-value
Nulo	38.603			
Final¹⁵⁸	30.388	8.215	1	0.004

Función liga: Logit.

Conforme a los resultados obtenidos, mostrados en la tabla anterior, tenemos que $\chi^2_{(2-1)}, \alpha=0.05 = 3.8415 < 8.125$,

con un $P - value = 0.004 < \alpha = 0.05$;

Por lo tanto se rechaza la hipótesis nula y se concluye que la estimación del modelo de Regresión Ordinal para el grupo de Difusión tiene sentido.

¹⁵⁷ Al tener solamente tres variables independientes, la variabilidad explicada no puede ser tan grande. Recuerde que siempre que se incorporan variables al modelo, aumenta la variabilidad explicada.

¹⁵⁸ Modelo que incluye las variables predictoras especificadas.

Los cálculos para la estimación de los parámetros β 's y α 's se realizan mediante el método de Máxima verosimilitud. El siguiente cuadro presenta la información necesaria para interpretar cuáles parámetros formarán parte del modelo.

Estimación de los parámetros

		Coefficientes	Std. Error	Wald	G.L.	P-value
Desempeño del PASDBP	Bajo	-0.543	0.228	5.661	1	0.017
	Mediano	0.593	0.229	6.722	1	0.010
	Alto	1.708	0.241	50.39	1	0.000
Difusión	Si	0.688	0.243	8.025	1	0.005
	No	0 ^a	.	.	0	.

Función liga: Logit.

a. Categoría de comparación.

En base a la información obtenida de la tabla anterior, se obtienen los siguientes resultados en la Prueba de Wald:

- La única variable independiente tiene sólo dos categorías por lo que se calcula solo un coeficiente β para el modelo. Entonces para la categoría “Si” (es decir, si tiene difusión el MSD por parte del encargado) de la variable Difusión se tiene $\chi^2_{(1), \alpha=0.05} = 3.8415 < 8.025$ un $P - value = 0.005 < \alpha = 0.05$ por lo que se rechaza la hipótesis nula y se considera significativo el coeficiente para formar parte del modelo, y por lo tanto diferente de cero.

Para determinar el valor de los umbrales α 's (del desempeño del PASDBP) también se realiza la Prueba de Wald, con las siguientes hipótesis:

$$H_0: \alpha_i = 0 \quad vs \quad H_1: \alpha_i \neq 0$$

Con $i = 1, 2, 3$ dónde:

- α_1 = Umbral (constante) de Nivel de Desempeño Bajo
- α_2 = Umbral (constante) de Nivel de Desempeño Mediano
- α_3 = Umbral (constante) de Nivel de Desempeño Alto

- Como se puede observar en la tabla, resultan estadísticamente significativos todos los umbrales de los niveles de desempeño del PASDBP.

Las probabilidades condicionales pueden ser definidas como los siguientes predictores lineales:

$$f(\theta_1(X)) = f(P(Y \leq Y_1 | X_D)) = \alpha_1 - \beta_D X_D \\ = -0.543 - 0.688X_D$$

$$f(\theta_2(X)) = f(P(Y \leq Y_2 | X_D)) = \alpha_2 - \beta_D X_D = 0.593 - 0.688X_D$$

$$f(\theta_3(X)) = f(P(Y \leq Y_3 | X_D)) = \alpha_3 - \beta_D X_D = 1.708 - 0.688X_D$$

Donde:

α_1 = Umbral (constante) de Nivel de Desempeño Bajo

α_2 = Umbral (constante) de Nivel de Desempeño Mediano

α_3 = Umbral (constante) de Nivel de Desempeño Alto

β_D = Coeficiente de la variable Difusión de la categoría Si

X_D = Variable Difusión de la categoría Si (Si se tiene difusión por parte de los encargados de módulo).

Así que $X_D = 0, 1$ dependiendo si existe o no difusión en el MSD por parte del encargado.

- Si $X_D = 1$ entonces el módulo cuenta con difusión por parte del encargado
- y si $X_D = 0$ entonces no existe difusión del MSD por parte del encargado.

A continuación se presentan los modelos para cada uno de los niveles del desempeño del PASDBP ($Y_1 = \text{Bajo}$, $Y_2 = \text{Medio}$, $Y_3 = \text{Alto}$ y $Y_4 = \text{Excelente}$) que se construyen con los respectivos umbrales α 's y los parámetros β 's en cada uno de ellos:

$$P(Y = Y_1 | X_D) = \left[\frac{1}{1 + e^{-(\alpha_1 - \beta_D X_D)}} \right] \\ = \left[\frac{1}{1 + e^{-(-0.543 - 0.688X_D)}} \right]$$

$$P(Y = Y_2 | X_D) = \left[\frac{1}{1 + e^{-(\alpha_2 - \beta_D X_D)}} \right] - P(Y = Y_1 | X_D) = \\ = \left[\frac{1}{1 + e^{-(0.593 - 0.688X_D)}} \right] - \left[\frac{1}{1 + e^{-(-0.543 - 0.688X_D)}} \right]$$

$$P(Y = Y_3 | X_D) = \left[\frac{1}{1 + e^{-(\alpha_3 - \beta_D X_D)}} \right] - P(Y = Y_2 | X_D) \\ = \left[\frac{1}{1 + e^{-(1.708 - 0.688X_D)}} \right] - \left[\frac{1}{1 + e^{-(0.593 - 0.688X_D)}} \right]$$

$$P(Y = Y_4 | X_D) = [1] - P(Y = Y_3 | X_D) = 1 - \left[\frac{1}{1 + e^{-(1.708 - 0.688X_D)}} \right]$$

El modelo de regresión ordinal proporciona los modelos de probabilidades acumuladas, de los cuáles se obtienen los modelos de las probabilidades absolutas. Para ejemplificar el cálculo de los dos tipos de probabilidades; primero se obtiene la probabilidad condicional de tener un desempeño medio o bajo en el PASDBP dado que no existe difusión del MSD por parte del encargado (probabilidad acumulada):

$$P (Y \leq Y_2 | X_D = 0) = \left[\frac{1}{1 + e^{-(0.593 - 0.688(0))}} \right] = 0.644$$

Por otro lado si se desea conocer la probabilidad condicional de obtener un desempeño medio dado que no existe difusión del modulo por parte de los encargados se calcula como:

$$P (Y = Y_2 | X_D = 0) = \left[\frac{1}{1 + e^{-(0.593 - 0.688X_D)}} \right] - \left[\frac{1}{1 + e^{-(0.543 - 0.688X_D)}} \right] = 0.277$$

Con el modelo de regresión ordinal, se calculan todas las probabilidades condicionales resultantes de las $(2 * 4 = 8)$ combinaciones posibles del vector (Y_j, X_D) . En seguida el cuadro presenta la información:

Probabilidades Condicionales

Difusión	Desempeño del PASDP			
	Bajo	Mediano	Alto	Excelente
Si (n=366)	0.226	0.250	0.259	0.265
No (n=66)	0.367	0.277	0.203	0.153

Función liga: Logit.

Para ejemplificar la lectura de la tabla anterior, obtenemos que la probabilidad de obtener un excelente desempeño dado que exista difusión de parte del Encargado del Módulo es 0.226, es decir, se espera que sean 97 MSD en esa clasificación. Y la diferencia es de 1 MSD, ya que se observa que 98 de los MSD donde se tiene difusión por parte del Encargado del Módulo presentan excelente desempeño.

Haciendo el mismo ejercicio con las demás combinaciones posibles encontramos que no excede en 2 la diferencia entre los MSD esperados con los observados.

Nótese que la probabilidad más baja dentro de la categoría “Si” de la variable Difusión se presenta al obtener un bajo desempeño y va creciendo hasta obtener la mayor probabilidad con un excelente desempeño. De manera inversa, cuando no se tiene difusión por parte del encargado del MSD se obtiene la probabilidad más alta en la categoría bajo desempeño y va decreciendo hasta la categoría de excelente desempeño.

En conclusión, la difusión del MSD influye en forma significativa en el desempeño del programa; y la probabilidad de obtener un mejor desempeño es más alta cuando existe difusión por parte del encargado del módulo.

Ya estimado el modelo de regresión ordinal, se procede a la aplicación de las pruebas estadísticas χ^2 de Pearson y Deviance, así como el cálculo de la Pseudo – R^2 de Nagelkerke para determinar hasta qué punto es razonable suponer que los datos provienen de la distribución propuesta por la regresión ordinal.

Pearson y Devianza¹⁵⁹

	Ji-Cuadrada	G.L.	P-value
Pearson	0.374	2	0.829
Devianza	0.371	2	0.830

Función liga: Logit.

Como se observa en el cuadro anterior, para la Prueba de Devianza se tiene $\chi^2_{(2), \alpha=0.05} = 5.9915 > 0.371 = D$ con un $P - value = 0.83 > \alpha = 0.05$ por lo tanto se determina que no se tiene evidencia suficiente para rechazar la hipótesis nula $H_0: F(x) = F_0(x)$. En la Prueba χ^2 de Pearson se tiene que $\chi^2_{(2), \alpha=0.05} = 5.9915 > 0.374 = S$, con un $P - value = 0.829 > \alpha = 0.05$, por consiguiente se obtiene el mismo resultado de no rechazar la hipótesis nula que en la Prueba de Devianza.

Nótese que el valor de $D = 0.371$ y $S = 0.374$ son pequeños, lo que indica que se tiene un buen ajuste, debido a que entre más pequeño sea el valor obtenido e los estadísticos D y S mejor será el ajuste del modelo.

Por último se muestra la información de la *pseudo – R^2 de Nagelkerke*.

Pseudo R-Cuadrada

Nagelkerke	0.02
-------------------	-------------

Función liga: Logit.

El porcentaje obtenido es el más bajo de nuestros modelos grupales y sigue siendo bajo interpretado en la escala de las *pseudo – R^2 's*. Sin embargo, la razón parece ser que el hecho de que al ser un modelo de regresión ordinal univariado influya en la medida debido a que al incrementarse el número de variables predictivas en un modelo, también se incrementa su verosimilitud.

¹⁵⁹ Estos estadísticos pueden ser correctamente utilizados porque los valores en la celda son grandes ($n_{ij} \geq 5$)

b) Construcción del modelo de regresión ordinal final del desempeño del PASDBP

Después de construir los modelos con las variables correspondientes a cada uno de los grupos, se ha obtenido un bosquejo de cuáles son las variables más significativas en cada uno de ellos, así como cuáles son aquéllas que parecen no aportar mucho a la explicación del comportamiento del desempeño del PASDBP.

Finalmente en este apartado se realiza la construcción del modelo de regresión múltiple completo del Desempeño del Programa de Acceso a Servicios Digitales en Bibliotecas Públicas, compuesto por los factores que lo determinan en mayor medida. Previamente se ha analizado de manera grupal las variables explicativas que parecen afectar el programa. Por lo que se tienen expectativas basadas en los resultados obtenidos, no sólo en los modelos grupales, sino también con los del análisis univariado y bivariado y con los aspectos teóricos que lo preceden y contextualizan.

Este ejercicio estadístico múltiple, que combina los cuatro grupos de variables explicativas, podría generar resultados similares, diferentes o hasta contradictorios a los esperados porque al incluir todas las variables, se analizan nuevas relaciones y eliminamos información redundante. Por ello la importancia de analizar todas las variables de manera simultánea para descubrir patrones y características esenciales de la información que contienen, con el fin de revelar la estructura de los datos.

Por medio de la aplicación de adecuadas pruebas estadísticas se comprueban supuestos (*Prueba de Líneas Paralelas*), se estiman parámetros para conformar los modelos (*Prueba de Wald*), se mide la bondad de ajuste y la variabilidad explicada por el conjunto de variables independientes que contiene el modelo (χ^2 de Pearson, *Devianza*, *Prueba de Máxima Verosimilitud*, *Contraste de Verosimilitudes*, *Pseudo-R²*).

Para determinar cuáles variables componen el modelo final se utilizará el mismo procedimiento de “Backward Stepwise”, empleado en la construcción de los modelos grupales. En refuerzo a la selección de éste método, se tiene la referencia el análisis para la selección de las variables independiente, el análisis bivariado y la misma construcción de los modelos grupales.

Debido a que se incrementa considerablemente el número de variables propuestas para considerarse en el modelo (trece variables explicativas), se utilizará un formato¹⁶⁰ diferente para presentar la información de las pruebas y datos concernientes a los modelos que son propuestos en los pasos de “Backward Stepwise”.

¹⁶⁰ Este formato presenta los resultados de diferentes modelos en dos tablas para la lectura de cada modelo. En la primera tabla se exhibe la información de la Prueba de Líneas Paralelas, Cociente de MV, Contraste de MV y se presenta la variable de cada modelo con mayor y menos significancia. En la segunda tabla, que se presenta la información de Devianza, Pearson y de la P-seudo R² de Nagelkerke.

Como se ha declarado, este modelo toma en cuenta todas las variables de Características de la Localidad (2 variables), Características del MSD (6 variables), Características del encargado (4 variables) y Difusión (1 variable), como posibles variables a incluir en el modelo final.

A continuación se presentan los resultados de las pruebas del modelo 1, el cual considera las variables de todos los grupos y el modelo 2 con una variable menos que el modelo 1 (la variable que se elimina es la de menor significancia, en el modelo anteriormente propuesto).

Modelo	Prueba de Líneas Paralelas		Cociente de Verosimilitudes		Contraste de M.V. ($\alpha=0.05$)		Variables del modelo con	
	<i>P-value</i>	<i>Ho</i>	<i>P-value</i>	<i>Ho</i>	<i>G</i>	<i>Ho</i>	<i>Mayor Significancia</i>	<i>Menor Significancia</i>
1. (Con todas las variables)	0.256	NR	0.000	R	-		Experiencia y Capacitación	Difusión
2. (Eliminando difusión)	0.199	NR	0.000	R	0.014	R	Experiencia y Capacitación	Sostenibilidad

R=rechazo de la hipótesis nula, NR=no se rechaza la hipótesis nula

Modelo	Devianza		Pearson		P-seudo R ² de Nagelkerke	# Variables
	<i>P-value</i>	<i>Ho</i>	<i>P-value</i>	<i>Ho.</i>		
1	1.0	NR	0.196	NR	0.331	13
2	1.0	NR	0.198	NR	0.331	12

Como se puede observar, el Modelo 2 que se propuso sin la variable Difusión, tiene en la Prueba de Contraste de Máxima Verosimilitud:

$$G = (-2\text{Log}[\text{Verosimilitud Modelo anterior}]) - (-2\text{Log}[\text{Verosimilitud Modelo nuevo}])$$

$$G = 707.659 - 707.645 = 0.014 < 3.8415 = \chi^2_{(1)}, \alpha=0.05 ;$$

Por lo tanto se rechaza la hipótesis nula y se concluye que el nuevo modelo sin la variable Difusión es mejor.

Continuando con el procedimiento de “Eliminación Progresiva”, a partir del modelo 2, se tienen los siguientes resultados:

Modelo	Prueba de Líneas Paralelas		Cociente de Verosimilitudes		Contraste de M.V. ($\alpha=0.15$)		Variables del modelo con	
	<i>P-value</i>	<i>Ho</i>	<i>P-value</i>	<i>Ho</i>	<i>G</i>	<i>Ho</i>	<i>Mayor Significancia</i>	<i>Menor Significancia</i>
1	0.256	NR	0.000	R	-		Experiencia y Capacitación	Difusión
2	0.199	NR	0.000	R	0.014	R	Experiencia y Capacitación	Sostenibilidad
3	0.093	NR	0.000	R	4.119	R	Experiencia y Capacitación	Tamaño de la Biblioteca

R=rechazo de la hipótesis nula, NR=no se rechaza la hipótesis nula

* Segunda variable menos significativa. **Tercera variable menos significativa...

Modelo	Devianza		Pearson		P-seudo R ² de Nagelkerke	# Variables
	<i>P-value</i>	<i>Ho</i>	<i>P-value</i>	<i>Ho.</i>		
1	1.0	NR	0.196	NR	0.331	13
2	1.0	NR	0.198	NR	0.331	12
3	1.0	NR	0.158	NR	0.331	11

De manera análoga, continuamos con el método de “Eliminación Progresiva” y en los siguientes cuadros encontramos la información de los modelos que se van eliminando.

Modelo	Prueba de Líneas Paralelas		Cociente de Verosimilitudes		Contraste de M.V. ($\alpha=0.15$)		Variables del modelo con	
	<i>P-value</i>	<i>Ho</i>	<i>P-value</i>	<i>Ho</i>	<i>G</i>	<i>Ho</i>	<i>Mayor Significancia</i>	<i>Menor Significancia</i>
1	0.256	NR	0.000	R	-		Experiencia y Capacitación	Difusión
2	0.199	NR	0.000	R	0.014	R	Experiencia y Capacitación	Sostenibilidad
3	0.093	NR	0.000	R	4.119	R	Experiencia y Capacitación	Tamaño de la Biblioteca
4	0.102	NR	0.000	R	6.521	R	Experiencia, Capacitación y Financiamiento	Antigüedad y Nivel de Marginación*
5	0.448	NR	0.000	R	120.75	NR	Experiencia y Escolaridad	Nivel de Marginación
5.1	0.184	NR	0.000	R	17.303	R	Experiencia, Capacitación y Financiamiento	Antigüedad

R=rechazo de la hipótesis nula, NR=no se rechaza la hipótesis nula

* Segunda variable menos significativa. **Tercera variable menos significativa

Modelo	Devianza		Pearson		P-seudo R ² de Nagelkerke	# Variables
	P-value	Ho	P-value	Ho.		
1	1.0	NR	0.196	NR	0.331	13
2	1.0	NR	0.198	NR	0.331	12
3	1.0	NR	0.158	NR	0.331	11
4	1.0	NR	0.130	NR	0.330	10
5	0.999	NR	0.228	NR	0.262	9
5.1	1.0	NR	0.151	NR	0.328	9

Se puede apreciar que hasta el Paso 5, los modelos propuestos por el método de “Eliminación progresiva” presentan siempre como variables más significativas a las variables del grupo Características del MSD “Experiencia/Tiempo en el PASDBP” y a las variables “Capacitación por parte de la DGB” del grupo Características del Encargado.

Otro detalle que se puede apreciar, es que la disminución en el porcentaje obtenido mediante la P-seudo R² ha sido muy pequeña del Paso 1 al 5. De hecho, se ha disminuido de un 33.1% desde el modelo inicial (Modelo 1) a un 32.8% en el Modelo 5.1 (que contiene 9 variables explicativas). Es decir, se disminuyó en un 0.3% con la exclusión de 4 variables predictivas. Esto indica que el Modelo 5 cumple con el propósito de ser un modelo más compacto (menor número de variables, y por lo tanto de coeficientes) pero bien ajustado.

A continuación, se presenta un cuadro con los siguientes pasos del método.

Modelo	Prueba de Líneas Paralelas		Cociente de Verosimilitudes		Contraste de M.V. ($\alpha=0.15$)		Variables del modelo con	
	P-value	Ho	P-value	Ho	G	Ho	Mayor Significancia	Menor Significancia
1	0.256	NR	0.000	R	-		Experiencia y Capacitación	Difusión
2	0.199	NR	0.000	R	0.014	R	Experiencia y Capacitación	Sostenibilidad
3	0.093	NR	0.000	R	4.119	R	Experiencia y Capacitación	Tamaño de la Biblioteca
4	0.102	NR	0.000	R	6.521	R	Experiencia, Capacitación y Financiamiento	Antigüedad y Nivel de Marginación*
5.1	0.184	NR	0.000	R	17.303	R	Experiencia, Capacitación y Financiamiento	Antigüedad, Tiempo de trabajar del EM*, Computadoras asignadas**.
6	0.088	NR	0.000	R	102.82	NR	Experiencia	Tiempo de trabajar del EM

6.1	0.415	NR	0.000	R	14.079	NR	Experiencia, Financiamiento y Tipo Encargado	Antigüedad
6.2	0.118	NR	0.000	R	16.905	R	Experiencia y Financiamiento	Antigüedad

R=rechazo de la hipótesis nula, NR=no se rechaza la hipótesis nula

* Segunda variable menos significativa. **Tercera variable menos significativa...

Modelo	Devianza		Pearson		P-seudo R ² de Nagelkerke	# Variables
	P-value	Ho	P-value	Ho		
1	1.0	NR	0.196	NR	0.331	13
2	1.0	NR	0.198	NR	0.331	12
3	1.0	NR	0.158	NR	0.331	11
4	1.0	NR	0.130	NR	0.330	10
5.1	1.0	NR	0.151	NR	0.328	9
6	0.998	NR	0.220	NR	0.258	8
6.1	0.999	NR	0.662	NR	0.319	8
6.2	0.999	NR	0.432	NR	0.307	8

Ahora ha disminuido a ocho el número de variables explicativas y el valor de la R^2 de Nagelkerke es de 0.307, ha disminuido únicamente 2.4%. El cual se considera una pequeña disminución del porcentaje tomando en cuenta que se ha reducido de 13 a 8 el número de variables predictivas y tiene un buen nivel considerando la escala de R^2 de Nagelkerke.

Una observación importante que se puede hacer al Modelo 6, es que cuatro de las ocho variables explicativas que lo conforman provienen del grupo Características del Encargado del MSD. En realidad aún no se ha eliminado ninguna variable del grupo Características del Encargado del MSD, lo que hasta el momento nos hace suponer que es el grupo de mayor importancia en el modelo.

Continuando con el proceso de eliminar variables para simplificar el modelo se obtienen los siguientes cuadros:

Modelo	Prueba de Líneas Paralelas		Cociente de Verosimilitudes		Contraste de M.V. ($\alpha=0.15$)		Variables del modelo con	
	P-value	Ho	P-value	Ho	G	Ho	Mayor Significancia	Menor Significancia
1	0.256	NR	0.000	R	-		Experiencia y Capacitación	Difusión
2	0.199	NR	0.000	R	0.014	R	Experiencia y Capacitación	Sostenibilidad
3	0.093	NR	0.000	R	4.119	R	Experiencia y Capacitación	Tamaño de la Biblioteca

4	0.102	NR	0.000	R	6.521	R	Experiencia, Capacitación y Financiamiento	Antigüedad y Nivel de Marginación*
5.1	0.184	NR	0.000	R	17.303	R	Experiencia, Capacitación y Financiamiento	Antigüedad, Tiempo de trabajar del EM*, Computadoras asignadas**.
6.2	0.118	NR	0.000	R	16.905	R	Experiencia y Financiamiento	Antigüedad y Tiempo de trabajar del EM*
7	0.015	R	-	-	-	-	-	-
7.1	0.488	NR	0.000	R	18.119	R	Experiencia y Financiamiento	Antigüedad

R=rechazo de la hipótesis nula, NR=no se rechaza la hipótesis nula

* Segunda variable menos significativa. **Tercera variable menos significativa...

Modelo	Devianza		Pearson		P-seudo R ² de Nagelkerke	# Variables
	P-value	Ho	P-value	Ho.		
1	1.0	NR	0.196	NR	0.331	13
2	1.0	NR	0.198	NR	0.331	12
3	1.0	NR	0.158	NR	0.331	11
4	1.0	NR	0.130	NR	0.330	10
5.1	1.0	NR	0.151	NR	0.328	9
6.2	0.999	NR	0.432	NR	0.307	8
7	-	-	-	-	-	7
7.1	0.989	NR	0.874	NR	0.305	7

Es importante señalar que la estimación del Modelo 7, conformado por Tipo de Localidad, Experiencia/Tiempo en el PASDBP, Financiamiento por parte de las autoridades municipales, Tipo de Encargado, Escolaridad del Encargado, Tiempo de Trabajar del encargado y Capacitación recibida por DGB, resultó ser irrelevante al considerarse que los datos no son adecuados para ser modelados mediante un método de regresión ordinal debido al incumplimiento de la Prueba de Líneas Paralelas.

Nótese que de las siete variables contenidas en el Modelo 7.1: una pertenece al grupo Características de la Localidad (variable Tipo de Localidad), tres pertenecen al grupo Características del MSD (Experiencia/Tiempo en el PASDBP, Antigüedad de la Biblioteca y Financiamiento) y las tres restantes al grupo al grupo Características del Encargado de Módulo (Tipo de Encargado, Escolaridad del Encargado y Capacitación por parte de la DGB).

Siguiendo con la modelación de los determinantes del Desempeño del PASDBP, se presentan los siguientes cuadros con la información de los pasos restantes en la obtención del modelo Final, mediante el procedimiento de eliminación de variables explicativas.

Modelo	Prueba de Líneas Paralelas		Cociente de Verosimilitudes		Contraste de M.V. ($\alpha=0.15$)		Variables del modelo con	
	<i>P-value</i>	<i>Ho</i>	<i>P-value</i>	<i>Ho</i>	<i>G</i>	<i>Ho</i>	<i>Mayor Significancia</i>	<i>Menor Significancia</i>
1	0.256	NR	0.000	R	-		Experiencia y Capacitación	Sostenibilidad*
2.1	0.144	NR	0.000	R	2.729	R	Experiencia y Capacitación	Tamaño de la Biblioteca **
3.2	0.093	NR	0.000	R	3.772	R	Experiencia y Capacitación	Difusión
4	0.102	NR	0.000	R	4.115	R	Experiencia, Capacitación y Financiamiento	Nivel de Marginación*
5.1	0.183	NR	0.000	R	17.303	R	Experiencia y Capacitación	Computadoras asignadas**
6.2	0.118	NR	0.000	R	16.905	R	Experiencia y Financiamiento	Tiempo de trabajar del EM*
7.1	0.488	R	0.000	R	18.119	R	Experiencia y Financiamiento	Antigüedad
8	0.330	R	0.000	R	26.388	R	Experiencia y Tipo de Encargado	Capacitación por parte de la DGB

R=rechazo de la hipótesis nula, NR=no se rechaza la hipótesis nula
* Segunda variable menos significativa. **Tercera variable menos significativa

Modelo	Devianza		Pearson		P-seudo R ² de Nagelkerke	# Variables
	<i>P-value</i>	<i>Ho</i>	<i>P-value</i>	<i>Ho.</i>		
1	1.0	NR	0.196	NR	0.331	13
2.1	1.0	NR	0.181	NR	0.331	12
3.2	1.0	NR	0.160	NR	0.330	11
4	1.0	NR	0.130	NR	0.330	10
5.1	1.0	NR	0.151	NR	0.328	9
6.2	0.999	NR	0.432	NR	0.307	8
7.1	0.989	NR	0.874	NR	0.305	7
8	0.754	NR	0.782	NR	0.232	6

Finalmente se concluye la búsqueda del modelo con los determinantes del desempeño del programa al obtener un modelo con todas las variables estadísticamente significativas, el cual obtiene R² de Nagelkerke de 0.232, que nos sirve como una medida alternativa para valorar el ajuste del modelo mediante la verosimilitud. Si bien este porcentaje puede resultar bajo, es importante

mencionar que para la R^2 ajustada los valores que se pueden considerar aceptables se encuentran en valores moderados e incluso bajos¹⁶¹.

Como se observa en el cuadro anterior, el mayor decrecimiento del porcentaje de la P-seudo R^2 de Nagelkerke del modelo se presenta del Modelo 7.1 al Modelo 8 con un 7.3%. Lo cual supone que tal vez la variable Antigüedad de la Biblioteca ha sido erróneamente excluida del modelo 8, como parte de una metodología y que posiblemente sea mejor dejarla formar parte del modelo final.

Sin embargo, para valorar cuál de estos dos modelos es el indicado para ser el modelo final que explique los determinantes de desempeño del PASDBP, se decide tomar en cuenta dos aspectos principalmente: el porcentaje obtenido en la R^2 de Nagelkerke y el ajuste del modelo. Para ello debemos recordar que al tener una variable más, el modelo 7.1, tiene un mayor porcentaje de R^2 ajustada y una tabla de contingencia más grande.

Como se explicó en el capítulo tres, al incrementar el número de variables explicativas en un modelo de regresión ordinal conlleva a la construcción de tablas donde el número de celdas vacías sea grande. Si bien, es una característica que debe de tomarse en cuenta, el modelo no está limitado a un cierto número de variables o especifica algún porcentaje aceptable de celdas nulas. Por lo que la regresión ordinal es una técnica estadística que permite tener buenas estimaciones, aun cuando se incrementa el número de variables.

Por otra parte, es importante entender que la relación entre un buen porcentaje de la R^2 de Nagelkerke y el ajuste del modelo a partir de un cierto número de variables independientes. Es decir, cuando se tiene un modelo con n variables independientes y se incorpora una variable extra, siempre se conseguirá una mayor coeficiente pero se incrementará el número de celdas nulas lo que nos lleva a que el modelo no se ajuste tan bien (y se reduce la confianza en el modelo).

Por lo anterior, teóricamente se podría escoger el Modelo 7.1 como el modelo final por presentar un P – pseudo R^2 de Nagelkerke más grande. Pero, dado que el fin de esta tesis es analizar mediante una regresión ordinal los determinantes del desempeño del PASDBP y presentar la información para sugerir recomendaciones, se necesita una herramienta de análisis que pueda utilizarse con mayor confianza para entender los factores que determinan en mayor medida el programa. Por lo tanto, se considera más importante tener un modelo mejor ajustado que brinde resultados más precisos y sobre todo reales. Entonces se justifica la elección del Modelo 8 como el final por ser el mejor ajustado. A continuación definimos más a detalle el que será el modelo final del Desempeño del Programa de Acceso a Servicios Digitales en Bibliotecas Públicas.

¹⁶¹ <<Fuentes electrónicas>>

http://books.google.com.mx/books?id=5KdXV7lxHIEC&pg=PA236&lpg=PA236&dq=r2+nagelkerke+valores&source=bl&ots=q_7wJMtW3w&sig=on4mFPJRI_70mSdiUFI_puA-4x8&hl=es-419&sa=X&ei=2KooUKOWNejc2QX95IGABw&ved=0CFYQ6AEwBw#v=onepage&q=r2%20nagelkerke%20valores&f=false (Consulta 30/04/12)

El modelo Final contiene las siguientes variables: Tipo de Localidad, Tipo de Encargado, Experiencia/Tiempo en el PASDBP, Financiamiento, Escolaridad del Encargado y Capacitación por parte de la DGB.

Se continúa con el método de máxima verosimilitud para la estimación de los parámetros β 's y α 's. Con la Prueba de Wald se determina si los coeficientes de cada categoría son estadísticamente significativos para incluirse en el modelo.

En donde, si $P - value < \alpha = 0.05$, se rechaza la hipótesis nula y se concluye que el coeficiente es estadísticamente significativo y diferente de cero.

		Coeficientes	Wald	G.L.	P-value	
Desempeño del PASDBP	Bajo	-1.664	0.429	15.006	0.000	Significativo
	Mediano	-0.300	0.424	0.502	0.479	No significativo
	Alto	1.050	0.425	6.096	0.014	Significativo
Tipo de Localidad	Urbano	0.654	0.230	8.070	0.005	Significativo
	Rural	0 ^a
Tipo de Encargado	EM	0.931	0.196	22.611	0.000	Significativo
	EM-B	0 ^a
Experiencia/Tiempo en el PASDBP	Muy Poca	-1.165	0.291	16.008	0.000	Significativo
	Poca	-0.981	0.272	13.011	0.000	Significativo
	Mediana	-0.918	0.280	10.714	0.001	Significativo
	Mucha	0 ^a
Financiamiento	Si	0.585	0.217	7.288	0.007	Significativo
	No	0 ^a
Escolaridad del Encargado	Educación Básica	-0.882	0.328	7.217	0.007	Significativo
	Educación Media Superior	-0.652	0.228	8.158	0.004	Significativo
	Educación Superior	0 ^a
	Educación Superior	0 ^a
Capacitación por parte de la DGB	0 - 1 curso	-0.999	0.269	13.783	0.000	Significativo
	2 - 4 cursos	-0.441	0.213	4.271	0.039	Significativo
	5 - 7 cursos	0 ^a

Link function: Logit.

a. Categoría de comparación.

Como se puede ver en los resultados la Prueba de Wald del cuadro anterior, todos los coeficientes resultaron significativos, con excepción del umbral de desempeño medio con un $p - value = 0.479 > \alpha = 0.05$ por lo que no se tiene evidencia suficiente para el rechazo de la hipótesis nula. Para el cálculo de las probabilidades se definen los siguientes predictores lineales:

$$\begin{aligned}
f(\theta_1(X)) &= f(P(Y \leq Y_2 | X_{TL}, X_{TE}, X_{E1}, X_{E2}, X_{E3}, X_F, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})) \\
&= \alpha_1 - \beta_{TL}X_{TL} - \beta_{TE}X_{TE} - \beta_{E1}X_{E1} - \beta_{E2}X_{E2} - \beta_{E3}X_{E3} - \beta_F - \beta_{EE1}X_{EE1} \\
&\quad - \beta_{EE2}X_{EE2} - \beta_{C1}X_{C1} - \beta_{C2}X_{C2} \\
&= -1.664 - 0.654X_{TL} - 0.931X_{TE} + 1.165X_{E1} + 0.981X_{E2} + 0.918X_{E3} \\
&\quad - 0.585X_F + 0.882X_{EE1} + 0.652X_{EE2} + 0.999X_{C1} + 0.441X_{C2}
\end{aligned}$$

$$\begin{aligned}
f(\theta_2(X)) &= f(P(Y \leq Y_2 | X_{TL}, X_{TE}, X_{E1}, X_{E2}, X_{E3}, X_F, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})) \\
&= \alpha_2 - \beta_{TL}X_{TL} - \beta_{TE}X_{TE} - \beta_{E1}X_{E1} - \beta_{E2}X_{E2} - \beta_{E3}X_{E3} - \beta_F - \beta_{EE1}X_{EE1} \\
&\quad - \beta_{EE2}X_{EE2} - \beta_{C1}X_{C1} - \beta_{C2}X_{C2} \\
&= -0 - 0.654X_{TL} - 0.931X_{TE} + 1.165X_{E1} + 0.981X_{E2} + 0.918X_{E3} - 0.585X_F \\
&\quad + 0.882X_{EE1} + 0.652X_{EE2} + 0.999X_{C1} + 0.441X_{C2}
\end{aligned}$$

$$\begin{aligned}
f(\theta_3(X)) &= f(P(Y \leq Y_3 | X_{TL}, X_{TE}, X_{E1}, X_{E2}, X_{E3}, X_F, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})) \\
&= \alpha_3 - \beta_{TL}X_{TL} - \beta_{TE}X_{TE} - \beta_{E1}X_{E1} - \beta_{E2}X_{E2} - \beta_{E3}X_{E3} - \beta_F - \beta_{EE1}X_{EE1} \\
&\quad - \beta_{EE2}X_{EE2} - \beta_{C1}X_{C1} - \beta_{C2}X_{C2} \\
&= 1.050 - 0.654X_{TL} - 0.931X_{TE} + 1.165X_{E1} + 0.981X_{E2} + 0.918X_{E3} \\
&\quad - 0.585X_F + 0.882X_{EE1} + 0.652X_{EE2} + 0.999X_{C1} + 0.441X_{C2}
\end{aligned}$$

Donde:

α_i = Umbral (constante) de los Niveles de Desempeño con $i = 1, 2, 3$.

Para 1=Bajo, 2= Mediano y 3=Alto

β_{TL} = Coeficiente de la variable Tipo de Localidad de la categoría Urbano

β_{TE} = Coeficiente de la variable Tipo de encargado de la categoría EM

β_{Ei} = Coeficiente de la variable Experiencia en el PASDBP con $i = 1, 2, 3$

Para 1=Muy Poca, 2= Poca y 3=Mediana

β_F = Coeficiente de la variable Financiamiento de la categoría Si

β_{EEi} = Coeficiente de la variable Escolaridad del encargado con $i = 1, 2$

Para 1= Educación Básica, 2= Educación Media Superior

β_{Ci} =Coeficiente de la variable Capacitación por parte de la DGB con con $i = 1, 2$

Para 1= "0-1 curso" y 2= "2-4 cursos"

X_{TL} = Variable Tipo de Localidad de la categoría Urbano

X_{TE} = Variable Tipo de encargado de la categoría EM

X_{Ei} = Variable Experiencia en el PASDBP

X_F = Variable Financiamiento de la categoría Si

X_{EEi} = Variable Escolaridad del encargado

X_{Ci} = Variable Capacitación por parte de la DGB

En donde, X_{TL} puede tomar los valores 0 y 1, dependiendo el Tipo de Localidad al que se refiera.

- si $X_{TL} = 1$ entonces se refiere al tipo de localidad urbana y

- si $X_{TE} = 0$ entonces el tipo de localidad es rural.

X_{TE} puede tomar los valores 0 y 1, dependiendo el Tipo de Encargado al que se refiera.

- si $X_{TE} = 1$ entonces el tipo de encargado es EM (exclusivo del MSD) y
- si $X_{TE} = 0$ entonces el tipo de encargado es EM-B.

De manera análoga, con la variable Escolaridad del Encargado se tiene X_{EE1} y X_{EE2} donde $X_{EEi} = 1, 0$ con $i = 1, 2$. Para definir el Nivel de Escolaridad del Encargado de Módulo, si queremos indicar que el

- encargado tiene Educación Básica, entonces $X_{EE1} = 1$ y $X_{EE2} = 0$;
- si el encargado tiene Educación Media Superior entonces $X_{EE1} = 0$ y $X_{EE2} = 1$, y
- finalmente si se desea indicar que el encargado tiene Educación Superior entonces $X_{EE1} = 0$ y $X_{EE2} = 0$.

Y para la variable Capacitación se tiene X_{C1} y X_{C2} donde $X_{Ci} = 0, 1$ con $i = 1, 2$.Si queremos indicar que el encargado tomo:

- de 0-1 curso por parte de la DGB, entonces $X_{C1} = 1$ y $X_{C2} = 0$.
- Si el encargado tomó de 2-4 cursos entonces $X_{C1} = 0$ y $X_{C2} = 1$, y
- si se desea indicar que el encargado tomó de 5-7 cursos por parte de la DGB se tiene $X_{C1} = 0$ y $X_{C2} = 0$.

En cuanto a la experiencia en el PASDBP se tiene $X_{Ei} = 1, 0$ con $i = 1, 2, 3$ dependiendo del Tiempo en el PASDBP al que se refiera.

- si $X_{E1} = 1, X_{E2} = 0$ y $X_{E3} = 0$ entonces se tiene muy poca experiencia en el PASDBP.
- Si $X_{E1} = 0, X_{E2} = 1$ y $X_{E3} = 0$ entonces se tiene poca experiencia en el programa.
- Si $X_{E1} = 0, X_{E2} = 0$ y $X_{E3} = 1$ entonces se tiene mediana experiencia en el programa.
- Y por último Si $X_{E1} = 0, X_{E2} = 0$ y $X_{E3} = 0$ entonces se tiene mucha experiencia en el programa.

Una vez definidas las categorías de cada una de las variables predictoras, se construye el modelo de regresión ordinal para las categorías $Y_1 = Bajo, Y_2 = Mediano, Y_3 = Alto$ y $Y_4 = Excelente$ y se definen como:

$$\begin{aligned}
 P(Y = Y_1 | X_{TL}, X_{TE}, X_{E1}, X_{E2}, X_{E3}, X_F, X_{EE1}, X_{EE2}, X_{C1}, X_{C2}) &= \frac{1}{1 + e^{-(\alpha_1 - \beta_{TL}X_{TL} - \beta_{TE}X_{TE} - \beta_{E1}X_{E1} - \beta_{E2}X_{E2} - \beta_{E3}X_{E3} - \beta_F - \beta_{EE1}X_{EE1} - \beta_{EE2}X_{EE2} - \beta_{C1}X_{C1} - \beta_{C2}X_{C2})}} \\
 &= \frac{1}{1 + e^{-(-1.664 - 0.654X_{TL} - 0.931X_{TE} + 1.165X_{E1} + 0.981X_{E2} + 0.918X_{E3} - 0.585X_F + 0.882X_{EE1} + 0.652X_{EE2} + 0.999X_{C1} + 0.441X_{C2})}}
 \end{aligned}$$

$$\begin{aligned}
& P(Y = Y_2 | X_{TL}, X_{TE}, X_{E1}, X_{E2}, X_{E3}, X_F, X_{EE1}, X_{EE2}, X_{C1}, X_{C2}) \\
&= \left[\frac{1}{1 + e^{-(\alpha_2 - \beta_{TL}X_{TL} - \beta_{TE}X_{TE} - \beta_{E1}X_{E1} - \beta_{E2}X_{E2} - \beta_{E3}X_{E3} - \beta_F - \beta_{EE1}X_{EE1} - \beta_{EE2}X_{EE2} - \beta_{C1}X_{C1} - \beta_{C2}X_{C2})}} \right] \\
&- P(Y = Y_1 | X_{TL}, X_{TE}, X_{E1}, X_{E2}, X_{E3}, X_F, X_{EE1}, X_{EE2}, X_{C1}, X_{C2}) \\
&= \left[\frac{1}{1 + e^{-(0.654X_{TL} - 0.931X_{TE} + 1.165X_{E1} + 0.981X_{E2} + 0.918X_{E3} - 0.585X_F + 0.882X_{EE1} + 0.652X_{EE2} + 0.999X_{C1} + 0.441X_{C2})}} \right] \\
&- \left[\frac{1}{1 + e^{-(1.664 - 0.654X_{TL} - 0.931X_{TE} + 1.165X_{E1} + 0.981X_{E2} + 0.918X_{E3} - 0.585X_F + 0.882X_{EE1} + 0.652X_{EE2} + 0.999X_{C1} + 0.441X_{C2})}} \right]
\end{aligned}$$

$$\begin{aligned}
& P(Y = Y_3 | X_{TL}, X_{TE}, X_{E1}, X_{E2}, X_{E3}, X_F, X_{EE1}, X_{EE2}, X_{C1}, X_{C2}) \\
&= \left[\frac{1}{1 + e^{-(\alpha_3 - \beta_{TL}X_{TL} - \beta_{TE}X_{TE} - \beta_{E1}X_{E1} - \beta_{E2}X_{E2} - \beta_{E3}X_{E3} - \beta_F - \beta_{EE1}X_{EE1} - \beta_{EE2}X_{EE2} - \beta_{C1}X_{C1} - \beta_{C2}X_{C2})}} \right] \\
&- P(Y = Y_2 | X_{TL}, X_{TE}, X_{E1}, X_{E2}, X_{E3}, X_F, X_{EE1}, X_{EE2}, X_{C1}, X_{C2}) \\
&= \left[\frac{1}{1 + e^{-(1.050 - 0.654X_{TL} - 0.931X_{TE} + 1.165X_{E1} + 0.981X_{E2} + 0.918X_{E3} - 0.585X_F + 0.882X_{EE1} + 0.652X_{EE2} + 0.999X_{C1} + 0.441X_{C2})}} \right] \\
&- \left[\frac{1}{1 + e^{-(0.654X_{TL} - 0.931X_{TE} + 1.165X_{E1} + 0.981X_{E2} + 0.918X_{E3} - 0.585X_F + 0.882X_{EE1} + 0.652X_{EE2} + 0.999X_{C1} + 0.441X_{C2})}} \right]
\end{aligned}$$

$$\begin{aligned}
& P(Y = Y_4 | X_{TL}, X_{TE}, X_{E1}, X_{E2}, X_{E3}, X_F, X_{EE1}, X_{EE2}, X_{C1}, X_{C2}) \\
&= 1 - P(Y = Y_3 | X_{TL}, X_{TE}, X_{E1}, X_{E2}, X_{E3}, X_F, X_{EE1}, X_{EE2}, X_{C1}, X_{C2}) \\
&= 1 \\
&- \left[\frac{1}{1 + e^{-(1.050 - 0.654X_{TL} - 0.931X_{TE} + 1.165X_{E1} + 0.981X_{E2} + 0.918X_{E3} - 0.585X_F + 0.882X_{EE1} + 0.652X_{EE2} + 0.999X_{C1} + 0.441X_{C2})}} \right]
\end{aligned}$$

Para ejemplificar cómo se obtienen las probabilidades absolutas con el Modelo final de Regresión Ordinal; se calcula la probabilidad de obtener: un excelente desempeño en el PASDBP dado que el tipo de localidad es urbana, se tiene un encargado exclusivo para el módulo (EM), se tiene mucha experiencia en el programa, no se tiene financiamiento por parte de las autoridades municipales, el encargado tiene educación a nivel licenciatura y ha tomado 1 curso de capacitación impartido por la DGB.

$$\begin{aligned}
& P(Y = Y_4 | X_{TL} = 0, X_{TE} = 0, X_{E1} = 0, X_{E2} = 0, X_{E3} = 0, X_F, X_{EE1} = 0, X_{EE2} = 0, X_{C1} = 1, X_{C2} = 0) \\
&= 1 - \left[\frac{1}{1 + e^{-(1.050 - 0.654(1) - 0.931(1) + 1.165(0) + 0.981(0) + 0.918(0) - 0.585(0) + 0.882(0) + 0.652(0) + 0.999(1) + 0.441(0))}} \right] = 1 - 0.614 = 0.386
\end{aligned}$$

Y para ejemplificar cómo se obtienen las probabilidades acumuladas con el Modelo de Regresión Ordinal; se calcula la probabilidad de obtener: un bajo desempeño en el PASDBP dado que el tipo de localidad es rural, se tiene un encargado exclusivo para el módulo (EM), se tiene mucha experiencia en el programa, se tiene financiamiento por parte de las autoridades municipales, el encargado tiene educación a nivel bachillerato y ha tomado 3 cursos de capacitación impartidos por la DGB.

$$\begin{aligned}
& P(Y = Y_1 | X_{TL} = 0, X_{TE} = 1, X_{E1} = 0, X_{E2} = 0, X_{E3} = 0, X_F = 1, X_{EE1} = 0, X_{EE2} = 1, X_{C1} \\
& = 0, X_{C2} = 1) \\
& = \left[\frac{1}{1 + e^{-(-1.664 - 0.654(0) - 0.931(1) + 1.165(0) + 0.981(0) + 0.918(0) - 0.585(1) + 0.882(0) + 0.652(1) + 0.999(0) + 0.441(1))}} \right] \\
& = 0.1104
\end{aligned}$$

Con el modelo de regresión ordinal final de Desempeño del Programa de Acceso a Servicios Digitales en Bibliotecas Públicas se puede calcular cualquier combinación posible¹⁶² del vector

$(Y_j, X_{TL}, X_{TE}, X_{E1}, X_{E2}, X_{E3}, X_F, X_{EE1}, X_{EE2}, X_{C1}, X_{C2})$. Posteriormente se calculan las probabilidades condicionales.

Con la tabla de frecuencias esperadas y observadas se puede realizar las pruebas para comprobar la bondad de ajuste del modelo, en este caso ya se mostraron los resultados de las pruebas en los cuadros anteriores.

Debido a que la tabla de probabilidades estimadas es demasiado extensa, a continuación se presenta un cuadro con las probabilidades estimadas a partir del modelo de regresión ordinal que resumen la información de las variables en bloques para entender mejor cómo las variables pertenecientes a cada grupo influyen en el desempeño del PASDBP. El primero deja fijo las mejores condiciones del MSD (si se tiene financiamiento por parte de las autoridades municipales y se tiene mucha experiencia) y el segundo cuadro deja fijo las peores condiciones del MSD (no se tiene financiamiento por parte de las autoridades municipales y se tiene muy poca experiencia). Esto con el fin de comparar precisamente las probabilidades de tener un excelente desempeño, bajo esas condiciones.

Probabilidad de tener un desempeño excelente dado que:

MEJORES CONDICIONES DE MÓDULO (Se tiene financiamiento por parte de las autoridades municipales y mucha experiencia en el PASDBP)	Tipo de Localidad Rural	Tipo de Localidad Urbano
Peores condiciones del Encargado de Módulo (Contar con un EM-B ¹⁶³ para la atención del MSD que tenga una escolaridad de nivel básico y que tenga 0-1 curso de la DGB)	0.0873	0.1555
Mejores condiciones del Encargado de Módulo (contar con un EM ¹⁶⁴ para la atención del MSD que tenga una escolaridad nivel superior y que haya asistido a 5-7 cursos impartidos por la DGB)	0.6144	0.7539

¹⁶² Combinaciones posibles se refiere a que exista una respuesta para la combinación. Por ejemplo $(X_{C1} = 1, X_{C2} = 1)$ no puede ser una combinación.

¹⁶³ Se refiere a que el Encargado se encarga de atender el MSD y la biblioteca.

¹⁶⁴ Se refiere a que el Encargado se encarga de atender exclusivamente al MSD.

Como se puede observar en el cuadro anterior, las probabilidades de tener un desempeño excelente dado que se tienen las mejores condiciones en todas las variables que componen el modelo es muy alta y se expresa de la siguiente forma:

$$[P(Y = Y_4 | X_{TL} = 1, X_{TE} = 1, X_{E1} = 0, X_{E2} = 0, X_{E3} = 0, X_F = 1, X_{EE1} = 0, X_{EE2} = 0, X_{C1} = 0, X_{C2} = 0) = 0.7539].$$

En otras palabras, si se tiene un MSD en tipo de localidad urbana con financiamiento por parte de las autoridades municipales, tiene mucha experiencia en el PASDBP, cuenta con un Encargado de Módulo (es decir, se dedica exclusivamente a la atención del MSD) que haya asistido de 5-7 cursos impartidos por la DGB y que tenga una escolaridad nivel superior; se tiene una probabilidad de 0.7539 de tener un excelente desempeño.

Por otra parte, un MSD ubicado en una localidad rural que cuente con financiamiento por parte de las autoridades municipales, con mucha experiencia en el PASDBP, que cuente con un Encargado de Módulo-Bibliotecario que tenga un nivel de escolaridad básico y que haya acudido a lo más a 1 curso impartido de la DGB tiene una probabilidad de apenas 0.0873¹⁶⁵ de obtener un excelente desempeño.

Si además de las características del MSD y del Encargado citadas en el ejemplo del párrafo anterior, consideramos que el módulo se encuentre en una localidad urbana, la probabilidad de obtener un excelente desempeño aumenta únicamente en 0.682¹⁶⁶

Pero si consideramos las mejores características del MSD en un tipo de localidad rural, pero consideramos que se cuente con un Encargado de MSD exclusivo para el módulo, que tenga una escolaridad de nivel superior y que además haya asistido a la mayor cantidad de cursos (5-7) por parte de la DGB; se tiene una probabilidad de 0.6144¹⁶⁷ de tener un excelente desempeño.

Entonces se encuentra que el conjunto de factores del grupo Características del Encargado del Módulo resulta más influyente en el desempeño del programa que el Tipo de Localidad donde fue instalado. Esto es un punto muy favorable para mejorar el funcionamiento del PASDBP, ya que el Tipo de Localidad no puede cambiarse pero se pueden modificar las condiciones del Encargado del Módulo si se hacen las medidas adecuadas.

Y para concluir el capítulo se reitera que el modelo es adecuado para modelar los datos, debido a que cumple los supuestos correspondientes (*Líneas Paralelas*), demuestra bondad de ajuste mediante las pruebas correspondientes.

¹⁶⁵ $P(Y = Y_4 | X_{TL} = 0, X_{TE} = 0, X_{E1} = 0, X_{E2} = 0, X_{E3} = 0, X_F = 1, X_{EE1} = 1, X_{EE2} = 0, X_{C1} = 1, X_{C2} = 0) = 0.0873$

¹⁶⁶ $P(Y = Y_4 | X_{TL} = 1, X_{TE} = 0, X_{E1} = 0, X_{E2} = 0, X_{E3} = 0, X_F = 1, X_{EE1} = 1, X_{EE2} = 0, X_{C1} = 1, X_{C2} = 0) = 0.1555$

¹⁶⁷ $P(Y = Y_4 | X_{TL} = 0, X_{TE} = 1, X_{E1} = 0, X_{E2} = 0, X_{E3} = 0, X_F = 1, X_{EE1} = 0, X_{EE2} = 0, X_{C1} = 0, X_{C2} = 0) = 0.6144$

Conclusiones y Recomendaciones

En México, el problema de la brecha digital se ha vuelto una preocupación a resolver debido a que actualmente la pobreza no solamente se encuentra medida en términos alimenticios y de vivienda, sino también en la tecnología. Sin embargo, debido a que la complejidad de su naturaleza va más allá de la carencia o no de acceso a las Tecnologías de la Información y Comunicación (TIC), es fundamental la necesidad de estimular tanto el acceso como su uso y apropiación. Es difícil que un país se pueda desarrollar si no cuenta con la infraestructura tecnológica adecuada para su progreso, combatiendo así otras brechas subyacentes a la digital.

Los programas gubernamentales destinados a disminuir la brecha digital no han sido capaces de atacar de facto la problemática y por ende los esfuerzos realizados no han rendido los resultados esperados. Como se vio a lo largo del capítulo uno aún no se ha podido establecer una estrategia apropiada que unifique y potencialice los programas establecidos por el gobierno. Pareciera que, por el contrario, están gravemente desvinculados entre sí; lo que provoca que entorpezcan su operación y se desaproveche la ayuda otorgada.

En este contexto, el Programa de Acceso a Servicios Digitales en Bibliotecas Públicas (PASDBP) cobra importancia al presentarse como una opción diferente en muchos sentidos. En primera instancia, fue la primera colaboración de una institución privada (Fundación Bill y Melinda Gates) con el gobierno. Que a su vez, se convirtió en una alianza tripartita donde el espacio cultural de la Biblioteca fue una parte fundamental, como punto de reconocimiento al programa. El PASDBP planteó, en primer lugar, resolver el problema de acceso, dotando de computadoras e internet a las comunidades beneficiarias, permitiendo el acceso y uso gratuito a toda la población. Y a su vez, atajó el problema del uso y apropiación de las TIC agregando un actor, el encargado de módulo.

Por lo tanto, la iniciativa de esta tesis surgió como una manera de aportar elementos que contribuyan al buen resultado y mejoramiento del PASDBP o de cualquier otro programa con estructura similar que se desee implementar. La magnitud y metas del PASDBP, pero sobre todo el impacto positivo y tan anhelado de reducir la brecha digital que se podría obtener a partir de él, fueron los incentivos fundamentales para el desarrollo de este trabajo, desde una perspectiva diferente, y a la vez complementaria, a las evaluaciones realizadas anteriormente al programa. De esta manera, el objetivo principal de la tesis fue identificar las principales características de las bibliotecas y por lo tanto de los módulos, que fomentaran el mejor funcionamiento del programa.

La oportunidad de desarrollar el análisis estadístico múltiple de los componentes que determinan en mayor medida el desempeño del programa, arrojó información sustancial sobre defectos o aciertos en su operación y diseño. Y se vuelve una herramienta que podría ser útil, no sólo en nuestro país, sino en otros lugares en donde se ha implementado el programa, o se tiene considerado hacerlo. Los primeros resultados al analizar los determinantes de desempeño según las características de la localidad, se identifican diferencias importantes entre las localidades urbanas y rurales cuando se instala un módulo en alguna comunidad. Se encuentra que la probabilidad de obtener un excelente nivel de desempeño cuando se instala el MSD en una localidad de tipo urbana

es del 0.283, mientras que comunidades rurales es considerablemente más baja, 0.135. Estas diferencias se acentúan cuando se analiza la probabilidad de tener un desempeño bajo. La probabilidad de obtener un bajo desempeño cuando se tiene una comunidad rural es 0.391, y para una localidad urbana es del 0.202

Para el grupo de Características del Módulo de Servicios Digitales, se encontraron resultados interesantes ya que en el análisis no resultaron significativas características de la biblioteca como tamaño y antigüedad de la misma. Esto nos refiere a que, a diferencia de lo que se esperaba, el espacio donde fue instalado el módulo parece estar en segundo término, comparado con otros factores. Por su parte, los elementos referentes a los planes para continuar el programa (Sostenibilidad) y apoyo de recursos económicos por parte de autoridades municipales (Financiamiento) se encuentran como factores significativos para el modelo. En especial, los que aparecen dentro de este grupo de características como los sobresalientes, son los elementos que hacen referencia directamente al programa: la experiencia o tiempo transcurrido desde la instalación del MSD y las computadoras asignadas a la biblioteca. En realidad, encontrar éstas como las variables más significativas del modelo grupal de las Características del módulo da cuenta de la importancia que se debe atribuir al dotar de infraestructura adecuada y suficiente a los módulos y dar tiempo al desarrollo del programa para esperar mejores resultados; también incentivar el diseño de planes y acuerdos para la continuidad del módulo.

La probabilidad de obtener un bajo desempeño con bajos niveles en los factores anteriores, es decir: tener poca experiencia en el PASDBP, contar con el mínimo de computadoras, no tener un plan para continuar el programa y no contar con financiamiento de parte de las autoridades municipales es del 0.501, considerablemente alto al compararlo con el que tendría un módulo con las mejores características de estos factores 0.035; es decir, si se cuenta con mucha experiencia en el programa, se le asigna un número máximo de computadoras, se tiene financiamiento y se tiene un plan para continuar, y a su vez, si se cuenta con las características anteriores, la probabilidad de un excelente desempeño es de 0.694

Por otra parte, dentro del grupo de las Características del Encargado encontramos que la única variable que no se encontró significativa, contrario a lo que se podría pensar, es el tiempo de trabajar del encargado. Se había definido en evaluaciones anteriores¹⁶⁸, que la rotación por parte de los encargados era un factor que parecía tener una conexión visiblemente negativa al desempeño del PASDBP al no permitir al encargado completar su ciclo en el módulo por causas ajenas a él. Por ejemplo, que al cambiar de gobierno a nivel municipal o estatal también se cambiaba de personal por razones “políticas” o bien el incumplimiento del pago por parte de las autoridades, lo que dejaba en desventaja al módulo al romper con los lazos establecidos con los usuarios y la experiencia de hacerse cargo del espacio tecnológico. Por el contrario, en el análisis realizado para la tesis, este factor no resultó significativo, con lo que se le quita un peso de encima a la operación del programa ya que estaba fuera de su alcance hacerse cargo de esta situación de rotación del personal.

¹⁶⁸ Evaluación realizada por el Instituto de Investigaciones Sociales, UNAM.

Las variables que sí resultaron significativas en el modelo fueron tipo de encargado, escolaridad del encargado y capacitación recibida por la DGB. Es así como la probabilidad de obtener un excelente desempeño, si se cuenta con un encargado exclusivo del módulo, el cual tenga una preparación nivel superior y haya asistido al mayor número de cursos posibles de capacitación es de 0.546, mientras que la probabilidad de obtener el mismo nivel de desempeño contando con un encargado de tipo Encargado de Módulo-Bibliotecario, con escolaridad básica y con un curso o ninguno de capacitación de la DGB es de tan sólo el 0.056. Ahora comparando las probabilidades de “bajo desempeño” del programa, tenemos que una biblioteca con un encargado exclusivo para el módulo, con nivel de educación superior y con la mayor asistencia a los cursos de capacitación es de 0.064. Lo que nos indica que si el MSD cuenta con el encargado adecuado, la probabilidad de “bajo desempeño” del programa es muy baja, lo cual subsana el problema de movilidad.

Por su parte, la difusión juega un papel importante en la visibilidad del programa ya que de ello depende que se incorporen nuevos usuarios al sistema tecnológico. El modelo fue construido con sólo una variable difusión, la cual mostró que la probabilidad de obtener un excelente desempeño en el programa, si se tiene difusión por parte del encargado es del 0.265, mientras que si se desea obtener el mismo desempeño sin difusión disminuye a 0.153. Aunque la diferencia más notoria aparece en las probabilidades de bajo desempeño. Cuando no se tiene difusión es de 0.367 y de tener bajo desempeño, si se tiene difusión disminuye a 0.226. En realidad la diferencia de tener un excelente o bajo desempeño si se tiene difusión tan sólo es de 0.039, lo cual no parece decir mucho, pero al menos sí se sabe que se incrementa la probabilidad de un bajo desempeño si no se tiene difusión por parte del encargado.

De acuerdo con los resultados del análisis estadístico multivariado derivados del modelo final de regresión ordinal, en el que se consideraron todos los factores de los diferentes grupos, de manera conjunta se encontró que los factores más importantes para determinar la probabilidad de que una biblioteca tenga un bajo o un excelente desempeño hacen referencia principalmente a las capacidades y habilidades del actor encargado de módulo, así como a la asignación de financiamiento, experiencia en el programa y al tipo de localidad.

Para ilustrar mejor lo anterior, se tiene que la probabilidad de obtener un excelente desempeño cuando se tiene un tipo de localidad urbana, con un MSD con mucha experiencia y que cuenta con financiamiento por parte de las autoridades municipales, que además tiene un encargado exclusivo para el módulo que cuenta con educación nivel superior y que ha recibido 6 cursos de capacitación por parte de la DGB es de 0.754, mientras que la probabilidad de obtener un bajo desempeño en un MSD con esas mismas condiciones es de 0.021.

En cuanto al grupo de factores que se analizaron, se confirma el hecho de que una localidad sea urbana o rural influye de manera diferenciada en el desempeño del PASDBP. Las comunidades rurales tienen menor probabilidad de alcanzar niveles de desempeño excelentes, comparado con las localidades urbanas lo cual puede que sea el reflejo de las mejores condiciones en las que laboran los módulos en las localidades urbanas, es decir, encargados mejor capacitados, mayor asistencia a la biblioteca y al módulo, mejores instalaciones, etc.

De igual manera, en el modelo final se confirma que ni el tamaño de la biblioteca para el MSD, ni la antigüedad de la misma resultaron significativas. Más bien la experiencia del módulo en el programa y el contar con el financiamiento por parte de las autoridades, es lo que contribuye en mayor medida a tener mejor desempeño. Este hecho cobra mucho sentido, ya que el financiamiento determina en gran parte que el MSD siga funcionando debido a que el programa otorga únicamente el equipo de cómputo, con la esperanza de que se cumplan los acuerdos establecidos respecto a su financiamiento por parte del municipio o delegación del que depende su mantenimiento y al mismo tiempo la participación del encargado de módulo ya que también el municipio es el que paga su salario.

Es de resaltar que el grupo de Características del Encargado de Módulo es el que se encontró con mayor influencia para el desempeño del programa, ya que en el modelo final se mantuvieron de manera significativa tres variables: escolaridad, tipo y capacitación del encargado; ya que la única que no resulto significativa fue el tiempo de trabajar del encargado. Lo que da luces acerca de algunas medidas para hacer acciones que mejoren el desempeño del PASDBP, e incluso en programas similares.

Por lo tanto, una de las conclusiones más importantes obtenida a partir de este análisis estadístico es que el desempeño del PASDBP puede mejorarse gracias a las variables referentes al actor encargado, ya que a diferencia de otras variables como tipo de localidad, éstas sí pueden ser modificadas. Por lo tanto, si se tiene un encargado de módulo en las mejores condiciones, es decir, con un alto nivel académico, dedicado exclusivamente para el MSD y estar adecuadamente capacitado, se puede incrementar considerablemente las probabilidades de un desempeño excelente del PASDBP.

Como un complemento a la aseveración anterior, podemos encontrar que en la construcción del modelo final se encontró que la Experiencia/Tiempo transcurrido en el PASDBP y la Capacitación Recibida por la DGB son los factores que resultaron influir en mayor medida. Por lo que se debe tomar en cuenta que los resultados del desarrollo del programa se darán de acuerdo al tiempo que lleva el MSD funcionando, y conforme pase el tiempo se esperarán mejores resultados. Por otra parte, siendo la Capacitación de los encargados de módulo por parte de la DGB la otra variable más influyente, y en especial, aquélla sobre la que se tiene control. Se recomienda ampliamente reforzar la capacitación del actor a cargo del MSD. Además establecerse algún criterio de carácter obligatorio de asistencia a los cursos por parte de la DGB, se podrían diseñar cursos alternativos y de actualización, que permitieran otorgar mejores herramientas de trabajo para el buen funcionamiento del programa.

También se encontró que los encargados de módulo que se dedicaban exclusivamente a él, brindaban mejores resultados debido al tiempo que podían dedicarle a los usuarios y al mantenimiento del equipo o cualquier otra actividad realizada en el espacio. Por lo que se aconseja priorizar la exclusividad de atención del módulo. Otro aspecto respecto al actor encargado del espacio del PASDBP, es su escolaridad ya que, como se muestra, entre mayor sea su nivel de educación, mayor nivel de desempeño propiciará.

Entonces encontramos que tres de las seis variables del modelo final corresponden a características del encargado de módulo. Este hallazgo genera un gran optimismo respecto a los alcances del modelo, ya que se encuentran áreas de oportunidad en su diseño y ayuda a tomar medidas para una mejor operación.

Gracias al análisis de regresión ordinal aplicado, se pueden visualizar de manera más clara algunos elementos que al cambiar podrían hacer diferencias importantes en el desempeño del programa.

Anexo 1. Construcción de la base de datos para el análisis

El primer criterio que se tomó para la construcción de la base de datos es dejar fuera a aquellas bibliotecas que no estuvieran equipadas al momento de la evaluación. Esta decisión reduce el número de 652 a 576 bibliotecas¹⁶⁹.

La siguiente maniobra en la construcción de la base, fue crear una “variable de seguimiento” para identificar el tipo de seguimiento que tuvieran aquellas bibliotecas con doble o triple participación. En el siguiente cuadro se presenta la información obtenida de esta variable.

Tipo de Seguimiento	Número de Bibliotecas
Primera y Tercera Encuesta	15
Segunda y Tercera Encuesta	36
Primera y Segunda Encuesta	35
Todas las Encuestas	9
Solo Primera Encuesta	48
Solo Segunda Encuesta	93
Solo Tercera Encuesta	236
Total	472

Partiendo de esta clasificación, se puede establecer que existen 472 bibliotecas efectivas¹⁷⁰ para el análisis. De las cuales 377 bibliotecas participan en una encuesta, 86 participaron en dos encuestas y únicamente 9 bibliotecas han sido evaluadas en las 3 encuestas. Es decir, se cuenta con 472 bibliotecas de las cuales 95 tienen información repetida.

Selección de actores para análisis.

A pesar de que una de las estrategias para el diseño de los cuestionarios, era llevar un seguimiento de las bibliotecas para poder comparar efectos a corto y largo plazo¹⁷¹, no todos los actores participan en las tres encuestas. Así que se toma como primer criterio selección de los cuestionarios, tomar únicamente aquellos con seguimiento, es decir, que se haya encuestado el mismo actor en los tres ejercicios de evaluación.

Después, se debe analizar cuáles actores ejercen una mayor influencia en el desarrollo del programa. En otras palabras, tiene que haber una fuerte relación del desempeño con estos actores. Para ello, decidimos elegir aquéllos que cumplan al menos una de las siguientes condiciones: ser

¹⁶⁹ No hay que perder de vista que nos referimos a un número real de bibliotecas menor a 576 debido a que hay bibliotecas repetidas en ese grupo.

¹⁷⁰ Llamaremos bibliotecas efectivas, a aquellas que cuenten con al menos una participación en alguna de las evaluaciones.

¹⁷¹ Reconocer los efectos a corto y largo plazo, era uno de las metas de la evaluación propuesta por la Fundación Bill y Melinda Gates; como uno de los componentes del mismo programa.

beneficiario del PASDBP y/o prestar servicios al programa. Al final, los actores seleccionados son usuarios y encargados de módulo.

Estructura de los datos

Existen diferentes tipos de variables¹⁷² y escalas de medición¹⁷³ en la base de datos que hemos creado. A continuación se presenta la clasificación de los tipos de variables que están contenidos en nuestra base

Tipo de Escala	
Nominal (Cualitativa)	Son variables numéricas cuyos valores representan una categoría o identifican un grupo de pertenencia. Este tipo de variables sólo nos permite establecer relaciones de igualdad/desigualdad entre los elementos de la variable. La asignación de los valores se realiza en forma aleatoria por lo que no cuenta con un orden lógico.
Ordinal (Cualitativa)	Son variables numéricas cuyos valores representan una categoría o identifican un grupo de pertenencia contando con un orden lógico. Este tipo de variables nos permite establecer relaciones de igualdad/desigualdad y a su vez, podemos identificar si una categoría es mayor o menor que otra. En las variables ordinales no se puede determinar la distancia entre sus categorías, ya que no es cuantificable o medible.
Escala (Cuantitativa)	Las variables de escala son todas aquellas variables cuyos valores representan magnitudes, ya sea que cuenten con un cero (0) absoluto o no. Con este tipo de variables podemos realizar comparaciones de igualdad/desigualdad, establecer un orden dentro de sus valores y medir la distancia existente entre cada valor de la escala.

¹⁷² Las variables pueden ser clasificadas como cuantitativas o cualitativas.

¹⁷³ La variable es medida utilizando una escala de medición.

Anexo 2. Detalles de la construcción Índice “Desempeño del PASDBP”

Elaboración de los indicadores de Servicios, de las Habilidades tecnológicas de los encargados y de Asistencia

SERVICIOS

Dentro de Servicios, las herramientas tecnológicas¹⁷⁴ tienen el peso más importante (50%) porque dentro de ellas encontramos las computadoras e internet, las cuales son herramientas base de este paquete tecnológico y los motores propuestos para impulsar el funcionamiento del programa.

En segundo lugar y con la misma ponderación (25%) tenemos la infraestructura¹⁷⁵ y la asesoría y capacitación, debido a que ambos juegan un papel igualmente importante en el desempeño. Ya que las condiciones en las que se encuentra el módulo refleja el desempeño los servicios, por ser el recinto físico e instrumental extra del Módulo de Servicios Digitales y asesoría y capacitación que brindan los encargados del MSD, son el factor crucial para que los anteriores sean aprovechados y se desempeñe de forma correcta el programa.

Subdimensión	Porcentaje
Herramientas Tecnológicas	50%
Infraestructura	25 %
Asesoría y Capacitación	25 %
Total (Servicios)	100%

Dentro de la dimensión de Herramientas Tecnológicas se encuentra: internet y paquetería. Y nuevamente se le asigna un peso basada en los cuatro niveles.

La asignación de ponderaciones de la dimensión “Herramientas Tecnológicas se realizó como ejemplo en el apartado de “Construcción del Índice de Desempeño PASDBP. Y sus porcentajes con respecto al subíndice Servicios es el siguiente:

Subdimensión	Indicador	Porcentaje	Porcentaje
Internet	Acceso a Internet	28.125%	37.5%
	Velocidad de Internet	9.375%	
Paquetería	Paquetería		12.5%
Total (Servicios)			50%

¹⁷⁴ Son todos aquellos artefactos o instrumentos de concepción y/o aplicación tecnológica.

¹⁷⁵ Se refiere no solamente al espacio físico donde se encuentra el MSD, sino también a los servicios dentro de él mismo.

Dentro de la dimensión Infraestructura tenemos dos subdimensiones: condiciones físicas del MSD y otros servicios.

Indicador	Porcentaje (Infraestructura)	Porcentaje (Servicios)
Condiciones Físicas del MSD	50%	25%
Otros Servicios	50 %	25%
Total	100 %	50%

La subdimensión Condiciones Físicas del MSD es igual de importante que el de Otros Servicios así que a ambos se les otorgó el mismo peso (50%) dentro de la dimensión Infraestructura. Es importante para el desempeño del programa que el espacio físico cuente con las condiciones necesarias no sólo para que el equipo funcione y se conserve correctamente sino porque es importante que el espacio sea adecuado también para las actividades que realicen los usuarios. El indicador Otros Servicios¹⁷⁶, tiene la misma consideración porque el MSD también puede brindar otros servicios que complementen el uso de computadoras e internet, provocando un mejor desempeño del PASDBP.

Dentro de la dimensión de **Asesoría y Capacitación** tenemos: cursos y asesorías. Donde ambas de sus subdimensiones comparten el mismo peso (50%). Esto se debe a que ambos indicadores son referentes de “atención a los usuarios” pero de formas diferentes.

La existencia de cursos implica una asesoría personalizada en la capacitación de diversos programas, algunos muy básicos y otros más complejos. Sin embargo, se debe hacer énfasis en que este servicio no es forzosamente la actividad de los encargados (aunque existe una alta probabilidad de que ellos los impartan) debido a la posibilidad de contratar personal especial para impartirlos. Por otra parte, la asesoría se refiere a la calidad en la atención que prestan los encargados a los usuarios¹⁷⁷ en las actividades cotidianas dentro del MSD.

Indicador	Porcentaje (Asesoría y Capacitación)	Porcentaje (Servicios)
Cursos	50%	25%
Asesorías	50%	25%
Total	100 %	50 %

HABILIDADES TECNOLÓGICAS DE LOS ENCARGADOS DEL MÓDULO

Dentro del subíndice Habilidades tecnológicas en los encargados de Módulo se tiene únicamente la dimensión de Conocimiento/experiencia en el uso de programas por lo que representa todo el peso

¹⁷⁶ Este indicador se refiere específicamente a los servicios de: Impresión de textos, acceso a programas en CD y quemado de discos.

¹⁷⁷ Un problema que se presentó en este punto, fue que algunos usuarios contestaron no necesitar ayuda de los Encargados de Módulo, esto nos impidió conocer la calidad de la asesoría. Al final, se decidió tomar esas respuestas como la calificación más baja, es decir, como si no tuvieran asesoría.

del subíndice (100%). Fue elegida por reflejar el desempeño del programa a través del nivel de capacitación de los encargados de módulo, y que tiene un efecto directo en la calidad de asesoría que brinda a los usuarios.

ASISTENCIA

Dentro de la dimensión Asistencia tenemos la subdimensión de Frecuencia de asistencia a la biblioteca y al MSD que tiene la mayor ponderación (75%), debido a que el buen desempeño del programa recae en la respuesta que reciba por parte de la comunidad donde es instalado el MSD.

Y en segundo lugar de importancia con un peso más bajo (25%) tenemos la subdimensión Tipo de usuarios. Éste nos indica la inserción del programa en diferentes grupos de la localidad, lo que implicaría una variedad de usos al MSD y a las herramientas tecnológicas que hay en él.

Indicador	Porcentaje
Frecuencia de asistencia a la biblioteca y al MSD	75 %
Tipo de usuarios	25%
Total (Asistencia)	100 %

Coefficientes de Ponderación en el Índice de “Desempeño del PASDBP”

Para resumir cómo hemos asignado el peso a cada una de las dimensiones e indicadores. Presentamos el siguiente cuadro que resume la información de los coeficientes de ponderación, utilizando los niveles para poder visualizar mejor el orden de cada uno de los indicadores y dimensiones respecto de otro.

Servicios 50%	Herramientas Tecnológicas 50%	Internet 75%	Acceso 75%
			Velocidad 25%
		Paquetería 25%	
	Asesoría y Capacitación 25%	Asesoría 50%	
		Cursos 50%	
	Infraestructura 25%	Condiciones Físicas 50%	
		Otros Servicios 50%	
Habilidades Tecnológicas de los Encargados 25%	Uso/Experiencia en programas 100%		
Asistencia 25%	Frecuencia 75%		
	Tipo de Usuarios 25%		

Preguntas seleccionadas para el “Índice de Desempeño PASDBP”

Posterior al cálculo de los coeficientes de ponderación y previo al cálculo técnico del índice, este paso consiste en buscar, dentro de la base de datos que se creó en el primer apartado del segundo capítulo, las preguntas que puedan ayudarnos a satisfacer la información para medir las dimensiones. Posteriormente son utilizadas en la construcción del índice.

Después de analizar a detalle cada una de las preguntas disponibles en nuestra base de datos, se encontraron las preguntas adecuadas para cada uno de los indicadores. El siguiente cuadro presenta seleccionadas.

INDICADOR	¿Qué mide la Variable?	PREGUNTAS DE CUESTIONARIO ENCARGADO DE MODULO	PREGUNTAS DE CUESTIONARIO USUARIOS
Acceso Internet	Acceso a internet	11./18. ¿Están conectadas a Internet las computadoras del módulo? Sí (1) No (2) p17 NS (8) p17 NC (9) p17	
	Número de computadoras que están conectadas a internet (relativo al número de computadoras presentes en el MSD) ¹⁷⁸	8./15 ¿Cuántas computadoras les otorgó el Programa? NS (98) NC (99) 12./19 ¿Cuántas computadoras están conectadas a Internet? NS (98) NC (99)	
Velocidad de Internet	Calificar la calidad en la velocidad de las computadoras del MSD	16./23. ¿Cómo considera la calidad de la conectividad a Internet? Muy Buena (1) p17 Buena (2) p17 Regular (esp.) (3) Mala (4) Muy Mala (5) NS (8) p17 NC (9) p17	46_1. ¿Cómo evaluaría el equipo de cómputo del módulo en Conexión a Internet...? Bueno (1) Regular (2) Malo (3) No hay internet(4) NS(8) NC (9)
Paquetería	Saber si brinda el uso de Paquetería	18./25. ¿El módulo brinda los siguientes servicios...? Uso de Paquetería Si (1) No (2) NS (8) NC (9)	

¹⁷⁸ Es importante mencionar, que lo ideal hubiera sido obtener la proporción de las computadoras otorgadas por el PASDBP y las computadoras conectadas a internet presentes en el MSD que fueron asignadas por el programa. Sin embargo, no existía una pregunta que hiciera la distinción entre las computadoras conectadas que fueron asignadas por el programa y las que no. Resolví esto, dejando la pregunta de las computadoras conectadas a internet en el MSD, porque no había otra forma de medir esta proporción y porque nos interesa saber con cuántas computadoras se tiene acceso a internet, más que si son las del PASDBP o no. Ejemplificando el problema, podría haber una biblioteca a la que se le asignaron 6 computadoras por parte del programa y tenía 2 computadoras previamente; nos dice que tiene 5 computadoras conectadas a internet (en el MSD) con lo cual obtenemos nuestra proporción de 5/6. Sin embargo no tenemos la certeza de que las 5 computadoras conectadas sean las entregadas por el programa, que es lo que nos interesa, ya que pudieran estar conectadas 3 computadoras obtenidas por el PASDBP y las 2 previas al programa, o bien 4 computadoras del PASDBP y una previa o efectivamente las 5 computadoras entregadas por el PASDBP. Lo que nos diría que la proporción real que nosotros buscábamos medir era de 3/6 , 4/6 o 5/6 respectivamente.

Asesoría	Calificar la asesoría que brinda el Encargado del MSD ¹⁷⁹	34. ¿Cuándo solicita ayuda del encargado del módulo... (Leer cada una de las opciones)? Le resuelve sus dudas (1) Es amable pero no le resuelve sus dudas (2) No ofrece ayuda (3) Ayuda, pero no es amable (4) Nunca ha requerido su ayuda (esp.) (5) NS (8) NC (9)
Si ofrece Cursos	Saber si ofrece o no cursos	53. /67. ¿Ofrece la biblioteca cursos de computación o Internet? Sí (1) No (2) NS (8) NC (9)
Cursos	Número de cursos que ofrece el MSD	54./68. ¿Qué tipo de cursos ofrece (Leer opciones y registrar todas las respuestas)..... Si (1) No (2) NS (8) NC (9) Uso básico de la computadora Clases de cómputo al personal de la biblioteca Paquetería de Office Asesoría en búsquedas en Internet Correo electrónico Diseño páginas Web Otro***no se contará
Condiciones físicas del MSD	Calificar la calidad de las condiciones físicas del MSD	45. ¿Cómo evaluaría el módulo en ...? (Leer cada una de las opciones) Bueno (1) Regular (2) Malo (3) NS (8) NC (9) Mobiliario Espacio Físico Iluminación Ruido Ventilación

¹⁷⁹ En esta pregunta es importante mencionar que los usuarios que contestaban que no requerían ayuda no nos daban una respuesta que calificará a la asesoría brindada por los EM, entonces nos encontramos con el problema de cómo calificar dicha pregunta. Al final se optó por asignarles la calificación como si no les dieran asesoría.

Otros servicios	Número de Servicios que ofrece el MSD	18./25. ¿El módulo brinda los siguientes servicios...? ¹⁸⁰ (Leer cada una de las opciones y registrar respuestas) Sí (1) No (2) NS (8) NC (9) Impresión de texto Acceso a programas en CD Quemado de discos	
Programas que maneja el Encargado del Módulo	Número de programas que el Encargado del Módulo ha usado /manejado	28./38. Usted, ¿Tiene alguna experiencia en el manejo y/o uso del siguiente software? (Leer opciones y registrar respuestas) Sí (1) No (2) NC (9) Procesador de texto (Word) Hojas de cálculo Presentadores (Power Point) Juegos Navegación en Internet Correo electrónico Publisher Access Encarta	
Frecuencia de asistencia	Medir los cambios en la frecuencia de asistencia de los usuarios a partir de que hay computadoras y calificar el nivel de asistencia diaria por cada computadora	8./15. ¿Cuántas computadoras les otorgó el Programa? NS (98) NC (99) 21./¿Me podría decir aproximadamente cuántas personas acuden al mes al módulo? *** en caso de EM-B.. promedio de usuarios al día en el módulo	37. A partir de que hay computadoras en la biblioteca, ¿usted ha cambiado la frecuencia con que la visita? (leer opciones) Acudo con más frecuencia (1) Acudo con la misma frecuencia (2) Acudo con menos frecuencia (3) NS (4) NC (5)
Tipo de usuarios	Tipo de distribución de los usuarios por sus grupos de edad		54/69. ¿Cuántos años cumplidos tiene? ————— NC (98)

Como podemos notar, la mayoría de las preguntas generan una respuesta de tipo cualitativo que es codificada numéricamente. Sin embargo, estos números no son adecuados para que tenga sentido el peso que se le ha asignado a cada variable, mucho menos si nos arroja valores en casos perdidos.

Entonces, se recodifican las respuestas para obtener una nueva variable numérica que represente de manera lógica lo que se desea medir y posteriormente se pondere según haya sido designado el coeficiente de ponderación. Recordemos que la calificación que se le busca dar al

¹⁸⁰ Había más opciones de “Otros servicios” para calificar como la lectura de USB, pero al no estar esta opción en los 3 cuestionarios de Encargado de MSD, tuvimos que cuantificar únicamente los servicios que estaban presentes en los 3 cuestionarios.

desempeño está medida entre 0 y 100. El siguiente cuadro nos presenta la recodificación que se hizo.

PREGUNTAS DE ENCARGADO DE MODULO	PREGUNTAS DE USUARIOS	RECODIFICACIÓN PREGUNTAS DE ENCARGADO DE MODULO	RECODIFICACIÓN PREGUNTAS DE USUARIOS	TOTAL
11./18. ¿Están conectadas a Internet las computadoras del módulo? Sí (1) No (2) p17 NS (8) p17 NC (9) p17		Si contesta : Si= 1 No= 0		Esta pregunta no tendrá valor, sólo nos sirve para clasificar a las bibliotecas por acceso a internet. En caso de no contar con el acceso a internet automáticamente tendrá 0 en todo el apartado de internet
8./15 ¿Cuántas computadoras le otorgó el Programa? NS (98) NC (99) 12./19 ¿Cuántas computadoras están conectadas a Internet? NS (98) NC (99)		Se obtiene la siguiente proporción: # Computadoras conectadas / # Computadoras otorgadas por el programa		Se multiplica por 14.0625
16./23. ¿Cómo considera la calidad de la conectividad a Internet? Muy Buena (1) p17 Buena (2) p17 Regular (esp.) (3) Mala (4) Muy Mala (5) NS (8) p17 NC (9) p17	46_1. ¿Cómo evaluaría el equipo de cómputo del módulo en Conexión a Internet...? Bueno (1) Regular (2) Malo (3) No hay internet(4) NS(8) NC (9)	Si contesta: Muy Buena =5 Buena=4 Regular =3 Mala= 2 Muy Mala= 1 No hay internet =0	Si contesta: Buena =5 Regular =3 Mala=1	Se multiplica lo obtenido del encargado por 0.25 y lo de los usuarios por .075 Después se suman y se multiplica el resultado por 0.9375
18./25. ¿El módulo brinda los siguientes servicios...? Uso de Paquetería Si (1) No (2) NS (8) NC (9)		Si contesta: Si=1 No=0		Se multiplica por 6.25
	34. ¿Cuando solicita ayuda del encargado del módulo... (Leer cada una de las opciones)? Le resuelve sus dudas (1) Es amable pero no le resuelve sus dudas (2) No ofrece ayuda (3) Ayuda, pero no es amable (4) Nunca ha requerido su ayuda (esp.) (5) NS (8) NC (9)		Si contesta: Le resuelve sus dudas=5 Ayuda, pero no es amable=3 Es amable pero no le resuelve sus dudas=1 No ofrece ayuda=0 Nunca ha requerido su ayuda= 0	Se multiplica por 1.25 La respuesta de los usuarios promedio (Cabe mencionar que tendremos una pregunta antes, de si necesita ayuda del encargado, que nos servirá para clasificar en si necesita o no asesoría, en caso de no necesitarla se le asigna automáticamente cero)

<p>53. /67. ¿Ofrece la biblioteca cursos de computación o Internet? Sí (1) No (2) NS (8) NC (9)</p>	<p>Si contesta : Si= 1 No=0</p>	<p>Esta pregunta no tendrá valor, únicamente no servirá para clasificar si las bibliotecas tienen o no cursos, en caso de contestar que "no" se le asigna un CERO al apartado de cursos</p>
<p>54./68. ¿Qué tipo de cursos ofrece (Leer opciones y registrar todas las respuestas)..... Si (1) No (2) NS (8) NC (9) Uso básico de la computadora Clases de cómputo al personal de la biblioteca Paquetería de Office Asesoría en búsquedas en Internet Correo electrónico Diseño páginas Web Otro***no se contará</p>	<p>Si contesta: Si=1 No=0 Para los 5 cursos Excepto para diseño de paginas web que Si=1.25 No=0</p>	<p>Es la suma de los cursos</p>
<p>45. ¿Cómo evaluaría el módulo en ...? (Leer cada una de las opciones) Bueno (1) Regular (2) Malo (3) NS (8) NC (9) Mobiliario Espacio Físico Iluminación Ruido Ventilación</p>	<p>Si contesta: Buena =5 Regular = 3 Mala= 1 Para cada una de los 5 aspectos y se suman</p>	<p>Se multiplica por 1.25</p>
<p>18./25. ¿El módulo brinda los siguientes servicios...? (Leer cada una de las opciones y registrar respuestas) Sí (1) No (2) NS (8) NC (9) Impresión de texto Acceso a programas en CD Quemado de discos</p>	<p>Si contesta Si en: Impresión=2.25 Acceso Cd=2 Quemado Discos=2 No=0 Y se suman</p>	<p>La suma todos los servicios</p>

<p>28./38. Usted, ¿Tiene alguna experiencia en el manejo y/o uso del siguiente software? (Leer opciones y registrar respuestas) Sí (1) No (2) NC (9) Procesador de texto (Word) Hojas de cálculo Presentadores (Power Point) Juegos Navegación en Internet Correo electrónico Publisher Access Encarta</p>		<p>Si contesta: Si=1 No =0 para todos excepto para Juegos y Encarta que en su caso tendrán Si=0.5 No=0</p>	<p>Se multiplica por 3.125</p>
<p>8./15. ¿Cuántas computadoras les otorgó el Programa? NS (98) NC (99) 21./¿Me podría decir aproximadamente cuántas personas acuden al mes al módulo? *** en caso de EM-B.. promedio de usuarios al día en el módulo</p>	<p>37. A partir de que hay computadoras en la biblioteca, ¿usted ha cambiado la frecuencia con que la visita? (leer opciones) Acudo con más frecuencia (1) Acudo con la misma frecuencia (2) Acudo con menos frecuencia (3) NS (4) NC (5)</p>	<p>Se obtiene la siguiente proporción: (#personas que acuden mensualmente al MSD/30)/ # Computadoras otorgadas por el programa Si el número de personas que atiende por día: alto=3 medio=2 bajo=1 * Se definirá después que es alto, medio y bajo .</p>	<p>Si contesta: Acudo con más frecuencia=3 Acudo con la misma frecuencia=1 Acudo con menos frecuencia= 0 Se multiplica el resultado del Encargado por 0.25 y la respuesta de los usuarios por 0.75; después se suma y el resultado se multiplica por 6.25</p>
	<p>54/69. ¿Cuántos años cumplidos tiene? <hr/> NC (98)</p>		<p>Si del grupo de usuarios encuestados: pertenecen al mismo grupo de edad=1 si se encuentran en 2 grupos=3 si se encuentran en los 3 grupos=5 Se multiplica por 1.25</p>

Con la información del cuadro de arriba, se puede calcular el desempeño de las bibliotecas. Sin embargo, la construcción de nuestra variable “Desempeño del PSDBP” a partir del índice nos crea otros problemas que deben ser resueltos antes de ser considerada como la variable dependiente.

Selección de Bibliotecas para el “Índice de Desempeño PASDBP”

Una vez concluido el cálculo del “Índice de Desempeño del PASDBP” para cada una de las bibliotecas, tenemos que examinar si hubo datos ausentes y/o errores¹⁸¹ en alguna de las preguntas utilizadas y también si hubo errores para cada una de las bibliotecas. Es decir, se debe analizar los errores por caso y por variable. Dependiendo del número de errores/datos ausentes que encontremos, decidiremos qué criterio tomaremos para su tratamiento.

¹⁸¹ Para nuestro análisis decidimos tomar las respuestas de No Contestó, No Sabe como errores. Entonces si existía un error o dato faltante se refería como información completa.

El siguiente cuadro presenta la información obtenida en el análisis por cada uno de las dimensiones del “Índice de Desempeño PASDBP”.

	# de Bibliotecas para el análisis con Información Completa	# de Bibliotecas para el análisis con Información Incompleta	Total
Servicios	450	22	472
Asistencia	446	26	472
Habilidades Tecnológicas de los Encargados de Modulo	468	4	472
TOTAL BIBLIOTECAS EFECTIVAS	432	40	472

Entonces, al analizar los errores por casos y por variables, encontramos que existe una baja incidencia de errores en el desarrollo de la construcción del “Índice Desempeño del PASDBP” y que los errores que hay en cada una de las dimensiones no afectan a las mismas bibliotecas. Es decir, hasta el momento, no se han presentado muchos casos problemáticos (es decir, bibliotecas con ausencia de información en la mayoría de sus dimensiones) sino más bien dimensiones problemáticas.

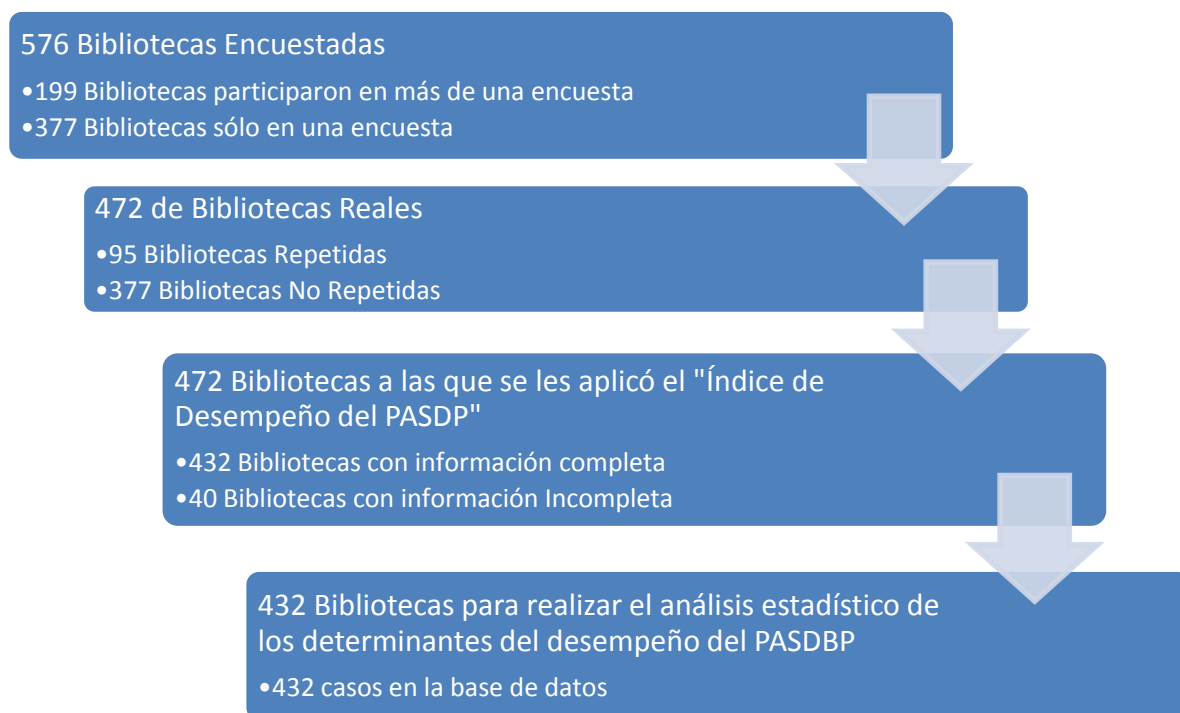
Para estudiar más detalladamente cómo se comportan los errores al interior de las dimensiones, hacemos nuevamente el análisis de errores/datos ausentes en cada uno de los indicadores de la dimensión “Servicios”.

		# de Bibliotecas para el análisis con Información Completa	# de Bibliotecas para el análisis con Información Incompleta	Total	
Servicios	Herramientas Tecnológicas	Internet	460	12	472
		Paquetería	468	4	472
	TOTAL BIBLIOTECAS REAL (en Herramientas Tecnológicas)		458	14	472
	Asesoría y Capacitación	Asesoría	464	8	472
		Cursos	468	4	472
	TOTAL BIBLIOTECAS REAL (en Asesoría y Capacitación)		460	12	472
	Infraestructura	Condiciones Físicas	465	7	472
		Otros Servicios	470	2	472
	TOTAL BIBLIOTECAS REAL (en Infraestructura)		463	9	472
	TOTAL BIBLIOTECAS REAL (en Servicios)		450	22	472

En el siguiente cuadro se examina la distribución de los errores en la dimensión de Asistencia.

Bibliotecas Efectivas		# de Bibliotecas para el análisis con Información Completa	# de Bibliotecas para el análisis con Información Incompleta	Total
Asistencia	Asistencia	446	26	472
	Tipo de Usuario	465	7	472
TOTAL BIBLIOTECAS REAL (en Asistencia)		446	26	472

Se puede concluir que la proporción de incidencia de errores/datos ausentes proviene más de las variables que de los casos. Y se podría considerar como una solución: eliminar los indicadores o dimensiones más problemáticas. Sin embargo, la proporción de errores es pequeña y se decide eliminar los casos (es decir, las bibliotecas) donde exista algún error /dato ausente sin importar que estén distribuidas más en las dimensiones y en indicadores que en las bibliotecas. Entonces, nuevamente nuestra base de datos para el análisis estadístico de los factores de desempeño es modificada. Veamos en el siguiente cuadro como queda definitivamente el número de bibliotecas para el análisis estadístico.



Como se puede observar la variable “Desempeño” presenta muy pocas repeticiones y se encuentra bien distribuida de los valores 18.62 a 95.12. La siguiente tabla muestra cómo es que se distribuyen los valores.

Desempeño			
	Frecuencia	Porcentaje	Porcentaje Acumulado
Hasta 30	5	1.16	1.16
30-40	12	2.78	3.94
40-50	24	5.56	9.49
50-60	71	16.44	25.93
60-70	81	18.75	44.68
70-80	109	25.23	69.91
80-90	113	26.16	96.06
90 y más	17	3.94	100
Total	432	100	

Anexo 3. Métodos de Estratificación

“Primera Aproximación”

El primer método de estratificación que usaremos para categorizar la variable “Desempeño”, es un procedimiento que plantea la estratificación con respecto a un criterio que considera la longitud del intervalo donde están comprendidos los datos. A continuación se presentan los pasos para realizar la estratificación.

- a. Aproximamos los valores mínimo y máximo, como decenas.
- b. Calcular la longitud de los estratos para cada categoría con la siguiente fórmula:

$$L = \frac{vmax - vmin}{Nc}$$

Donde:

vmax = valor máximo

vmin = valor mínimo

L = Longitud del estratos

Nc= Número de Categorías que deseamos obtener¹⁸². ($Nc > 2$)

- c. Calcular los límites de los estratos para definir las categorías.

$$Lim(i) = vmin + (i * L) \quad \text{Sea } i = 1, 2, \dots, Nc - 1$$

Y los límites quedan: $min, Lim(1), Lim(2), \dots, Lim(Nc - 1), max$ ¹⁸³

- d. Construcción de los intervalos con los límites obtenidos.

Estrato	Intervalo
primer Estrato	$[vmin, Lim(1)]$
segundo Estrato	$[Lim(1), Lim(2)]$
:	:
n-ésimo estrato	$[Lim(Nc - 1), vmax]$

¹⁸² En el caso de la variable “Desempeño” se decide que serán 4 categorías.

¹⁸³ Los límites para el “Desempeño” son: 18.6160714, 40, 60, 80, 95.1197917

“Método de Estratificación de Dalenius”

El Método de Estratificación de Dalenius busca que la población se agrupe en estratos según una característica específica¹⁸⁴, de tal manera que la variabilidad de dicha calificación *dentro* de un estrato sea lo más pequeño posible, y a la vez, lo más grande posible *entre* los estratos.

Para poder encontrar los estratos mediante este método se siguen los siguientes pasos:

- Se ordena en forma ascendente las observaciones de la muestra (en este caso las calificaciones del “Índice de Desempeño PASDBP”).
- Se obtienen las frecuencias absolutas¹⁸⁵ de cada uno de los valores.
- Se obtienen las frecuencias relativas¹⁸⁶ y las multiplicamos por 100
- Se obtiene la raíz cuadrada y posteriormente la redondeamos.
- Se obtiene el acumulado de la raíz redondeada.
- Se obtiene la suma de Redondeo de $\sqrt{(f*100)}$,

$$\sum_{i=1}^n \text{Redon.} \sqrt{(f_i * 100)} = S$$

Que será igual al acumulado de la X_n observación Siendo X_i el i -jésimo valor con $i= 1, 2, 3, \dots, n$.
Es decir, $S= acum(n)$

$$\sum_{i=1}^n \text{Redon.} \sqrt{(f_i * 100)} = S^{187} = acum(n)$$

- Para obtener los límites de los estratos se aplica la siguiente fórmula.

$$\text{Límite}(i) = \frac{S (Nc - i)}{S}$$

Donde:

Nc = Número de Categorías¹⁸⁸. ($Nc > 2$)

S = suma de Redondeo de $\sqrt{(f*100)}$

$i = 1, 2, \dots, Nc-1$

- Después se construyen los intervalos, a partir de los límites obtenidos anteriormente.

¹⁸⁴ En este caso la característica específica es la calificación otorgada por el “Índice de Desempeño del PASDBP”

¹⁸⁵ Frecuencia absoluta (n_i) de una variable estadística X_i , es el número de veces que este valor aparece en el estudio. A mayor tamaño de la muestra, aumentará el tamaño de la frecuencia absoluta; es decir, la suma total de todas las frecuencias absolutas debe dar el total de la muestra estudiada (N).

¹⁸⁶ Frecuencia relativa (f_i), es el cociente entre la frecuencia absoluta y el tamaño de la muestra (N). Es decir,

$$f_i = \frac{n_i}{N} = \frac{n_i}{\sum_i n_i} \text{ siendo el } f_i \text{ para todo el conjunto } i.$$

¹⁸⁷ En el caso del “Desempeño” la $S=213.6$

¹⁸⁸ Para la variable “Desempeño” se determinó utilizar 4 categorías.

Intervalos	
Intervalo 1	$[0, \text{Límite (1)}]$
Intervalo 2	$[\text{Límite (1)}, \text{Límite(2)}]$
:	:
Intervalo Nc	$[\text{Límite}(Nc - 1), S]$

- i. Una vez que se tienen los intervalos, se vuelve a las observaciones para obtener los nuevos límites.
- j. Por último se forman los estratos tomando en cuenta los nuevos límites.

Anexo 4. Pruebas de asociación para el análisis bivariado

a) Prueba χ^2 (Ji-Cuadrada).

La Prueba Ji-Cuadrada es una medida de asociación entre dos variables nominales. Permite decir si existe asociación entre dos variables, pero no en qué sentido.

Considera una tabla de contingencia $k \times m$, de las variables X y Y . Es decir, la variable X tiene k categorías y la variable Y tienen m categorías.

Se define como O_{ij} la frecuencia observada de la celda del i -ésimo renglón y la j -ésima columna. Y las frecuencias observadas en el renglón i serán $O_{i.}$ (Es decir, las frecuencias observadas totales en el renglón i). Mientras que las frecuencias observadas en la columna j sería $O_{.j}$. (Es decir, las frecuencias observadas totales en la columna j). El total de variables observadas se puede obtener como $O = \sum_{i=1}^k O_{i.} = \sum_{j=1}^m O_{.j}$

Ahora se dice que la probabilidad de que un valor pertenezca al i -ésimo renglón y la j -ésima columna es P_{ij} . Y que la probabilidad de que pertenezca al i -ésimo renglón es $P_{i.}$, mientras que la probabilidad de que pertenezca a la j -ésima columna es $P_{.j}$.

En la Prueba de Ji-Cuadrada se quiere probar que $P_{ij} = P_{i.}P_{.j}$ y entonces las hipótesis son

$$H_0: P_{ij} = P_{i.}P_{.j} \quad vs \quad H_a: P_{ij} \neq P_{i.}P_{.j}$$

Por otro lado, también debe recordarse que la condición de independencia dice que dos variables X e Y son independientes si y sólo si para cualquier pareja de valores x e y la probabilidad de que X tome el valor x e Y el valor y , simultáneamente, es igual al producto de las probabilidades de que cada una tome el valor correspondiente $X \perp Y \Leftrightarrow f(x,y) = f(x)f(y)$

Para poder realizar la prueba, se definen las probabilidades estimadas como

$$\hat{P}_{i.} = \frac{O_{i.}}{O} \quad y \quad \hat{P}_{.j} = \frac{O_{.j}}{O}$$

Y a partir de ellas se obtienen las frecuencias esperadas o teóricas, que se puede definir como aquella frecuencia que se observaría si ambas variables fueran independientes. La notación es la siguiente E_{ij} la frecuencia esperada para el i -ésimo renglón y la j -ésima columna. Entonces $E_{i.}$ la frecuencia esperada para el i -ésimo renglón y $E_{.j}$ como la frecuencia esperada para la j -ésima columna.

Entonces utilizando lo anteriormente definido, obtenemos la hipótesis nula H_0 como

$$E_{ij} = \frac{O_{i.} O_{.j}}{O} \quad \text{porque} \quad \frac{O_{i.} O_{.j}}{O} = \frac{O_{i.}}{O} \frac{O_{.j}}{O} O = \hat{P}_{i.} \hat{P}_{.j} O$$

Y nuestro estadístico de prueba χ^2 queda como

$$\chi^2_{(k-1)(m-1)} = \sum_{i=1}^k \sum_{j=1}^m \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\chi^2_{(k-1)(m-1)} = \sum_{i=1}^k \sum_{j=1}^m \frac{(O_{ij} - \frac{O_{i.} O_{.j}}{O})^2}{\frac{O_{i.} O_{.j}}{O}}$$

Donde $(k-1)(m-1)$ se refiere a los grados de libertad de la Ji-Cuadrada. Se considera que el estadístico es confiable cuando el 80% de las frecuencias esperadas en cada una de las celdas de la tabla de contingencia es mayor que 5.

Sin embargo, en esta prueba que se debe cumplir $E_{ij} \geq 5$ para $i = 1, 2, \dots, k$ y $j = 1, 2, \dots, m$. En caso de presentar el incumplimiento en alguna de ellas, se deberá de recategorizar la variable para solucionarlo.

Entonces las hipótesis son:

$$H_0: X \text{ es independiente de } Y \quad \text{vs} \quad H_a: X \text{ está asociado con } Y$$

$$H_0: \text{Independencia} \quad \text{vs} \quad H_a: \text{Asociación}$$

Criterio de Decisión: Se rechaza la Hipótesis Nula cuando $\alpha > p - \text{value}$. Aunque generalmente se le asigna α con un valor de 0.05. El valor de α^{189} es establecido por la persona que realiza la prueba, con el fin de establecer una α muy pequeña para realizar la prueba más estrictamente o bien una α más grande para una mayor tolerancia y se consideren asociaciones débiles.

b) Prueba Gamma. (De Goodman y Kruskall)

La Prueba de Gamma de Goodman y Kruskall es una medida de asociación para dos variables ordinales. Para realizar la prueba consideremos un par de observaciones clasificadas según dos variables cualitativas ordinales. En relación al orden de estas dos observaciones en cada una de las variables, el par se puede clasificar como *concordante* o *discordante*. Un par de observaciones es

¹⁸⁹ Alfa (α), también conocido como Error Tipo I, es la probabilidad de rechazar la hipótesis nula cuando es cierta. Es decir, que la prueba muestre significación estadística cuando no es cierto. Especificando un nivel alfa, el investigador fija los márgenes admisible de error, especificando la probabilidad de concluir que la significación existe cuando en realidad no existe.

concordante si aquella que se ordena más alta sobre la variable X también se ordena más alta sobre Y . Un par es discordante si la observación que se ordena más alta sobre X se ordena más baja sobre Y . No todos los $\left(\frac{n}{2}\right)$ pares de observaciones son concordantes o discordantes. Los pares que no son ni concordantes ni discordantes se llaman “ligados”. Un par de observaciones es ligado si ambas tienen la misma clasificación en alguna de las variables o en ambas.

Basándose en las probabilidades de concordancia y discordancia asociadas a una tabla de contingencia, se definen medidas de asociación para variables ordinales, análogas al coeficiente de correlación lineal para variables continuas, que describen el grado de monotonía de la asociación, es decir, si una variable tiende a aumentar cuando lo hace la otra o al revés.

Seleccionados al azar un par de observaciones o de individuos, la probabilidad de concordancia en la tabla es $P_c = 2 \sum_i \sum_j p_{ij} (\sum_{k>i} \sum_{m>j} p_{km})$

y la de discordancia es $P_d = 2 \sum_i \sum_j p_{ij} (\sum_{k>i} \sum_{m<j} p_{km})$

El factor 2 se debe a que la primera observación podría estar en la casilla (i,j) y la segunda en la casilla (k,m) o viceversa. Por otro lado el número total de pares concordantes en una tabla $I \times J$ viene dado por $C = \sum_i \sum_j n_{ij} (\sum_{k>i} \sum_{m>j} n_{km})$

Análogamente, el número total de pares discordantes es $D = \sum_i \sum_j n_{ij} (\sum_{k>i} \sum_{m<j} n_{km})$

Se dirá que existe asociación positiva si $P_c - P_d > 0$, es decir valores bajos de X tienden a ocurrir con valores bajos de Y y valores altos de X con valores altos de Y . Análogamente, existe asociación negativa si $P_c - P_d < 0$ lo que significa que valores altos de X tienden a ocurrir con valores bajos de Y y viceversa.

La medida Gamma fue propuesta por Goodman y Kruskal en 1954 y su valor poblacional se define como $\gamma = \frac{P_c - P_d}{P_c + P_d}$

Que es la diferencia entre las probabilidades de concordancia y discordancia, para aquellos pares no ligados por alguna variable. Su rango de variación es $-1 \leq \gamma \leq 1$. Si las variables son independientes, entonces, $P_c = P_d$ lo que implica que $\gamma = 0$. El recíproco no es cierto. La asociación positiva es tanto mayor cuanto más se aproxima γ a 1 y tanto menor cuanto más se aproxima a cero. A medida que se acerca a -1, aumenta la asociación negativa.

Para realizar la prueba se plantean las siguientes hipótesis:

$$H_0: X \text{ es independiente de } Y \quad \text{vs} \quad H_a: X \text{ está asociado con } Y$$

H_0 : Independencia vs H_a : Asociación

Criterio de Decisión: Se rechaza la Hipótesis Nula cuando $\alpha > p - value$. Aunque generalmente se le asigna α con un valor de 0.05. El valor de α es establecido por la persona que realiza la prueba, con el fin de establecer una α muy pequeña para realizar la prueba más estrictamente o bien una α más grande para una mayor tolerancia y se puedan considerar asociaciones débiles.

c) Coeficiente de Contingencia

Se define como una medida de asociación que permite evaluar el grado de asociación entre dos variables y sus valores oscilan entre $0 \leq C < 1$

Es una medida basada en *Ji-Cuadrada* que intenta corregir el valor del estadístico χ^2 para minimizar el tamaño de la muestra sobre la cuantificación del grado de asociación. $C = \sqrt{\frac{\chi^2}{(\chi^2 + n)}}$

Su valor máximo depende del número de filas y de columnas. Si el número de filas y columnas es k entonces $C = \sqrt{\frac{k-1}{k}}$

Interpretación:

Un coeficiente de 0 indica independencia, mientras que cuando alcanza su valor máximo indica asociación perfecta.

Bibliografía

- AGRESTI, Alan. "Analysis of ordinal categorical data", Ed. John Wiley & Sons, 2010
- AGRESTI, Alan. FINALY, Barbara. "Statistical Methods for the Social Sciences", Ed. Prentice Hall, 3rd Edition., 2008
- ÁLVAREZ CÁCERES, Rafael. "Estadística multivariante y no paramétrica con SPSS, Aplicación a las ciencias de la salud", Ed. Díaz de Santos, 1995.
- COOLEY, William W. LOHNES, Paul R. "Multivariate Data Analysis", Ed. John Wiley & Sons., 1971
- DILLON, Willian R. GOLDSTEIN, Matthew. "Multivariate Analysis, Method and Applications", Ed. John Wiley & Sons., 1985
- Fernández Pando, Martín Fernández(2004). *Regresión Logística Multinomial* , [PDF] España. Consultado el 30 de abril de 2012
- GIBBONS. DICKINSON, Jean. "Non parametric Statistical Inference", Ed. McGraw-Hill., 2010
- HAIR, ANDERSON, TATHAM, BLACK. "Análisis Multivariante", Ed. Prentice Hall, 5ta Edición., 2009
- IBM. Página oficial [en línea] Consultado el 30 de abril de 2012 en <http://www.ibm.com/us/en/>
- José Barrera (2009, 20 de mayo). *Introducción a la Regresión Ordinal*, [en línea] Consultado el 30 de abril de 2012 en http://www.mat.uab.cat/~jbarrera/Teaching_files/Regresion_Ordinal.pdf
- KEELE, Luke. "Semiparametric Regression for Social Sciences", John Wiley & Sons., 2008
- Ordinal Regression Chapter 4, [en línea] USA Consultado el 30 de abril de 2012 en http://www.norusis.com/pdf/ASPC_v13.pdf
- Pérez López César. "Técnicas de Análisis Multivariante de Datos, Aplicaciones con SPSS", Ed. Pearson Prentice Hall, Madrid 2004.
- RINCÓN, Luis. "Curso Intermedio de Probabilidad", Ed. Las prensas de ciencias, 2007.
- RINCÓN MEJÍA, Hugo Alberto. "Álgebra Lineal", Ed. Las prensas de ciencias, 2006
- ROSS, Sheldon. "A First Course of Probability", sixth Edition., 2001
- SANTOS, María Josefa. DE GORTARI, Rebeca. ANGULO, Yvon. "Disminuyendo la brecha digital: El nuevo papel de la biblioteca pública mexicana", CONACULTA, 2012.

SANTOS, María Josefa. DE GORTARI, Rebeca. ANGULO, Yvon. "Acceso tecnológico: una reinterpretación de las bibliotecas públicas", CONACULTA, 2006.

SANTOS, María Josefa. DE GORTARI, Rebeca. "Redefinición del espacio cultural: computadoras e internet en la biblioteca pública mexicana", Ed. IIS-UNAM Pearson.

SPRENT, Peter. SMEETON, Nigel C. "Applied Nonparametric Statistical Methods", Ed. Chapman & Hall/CRC, 4th Edition., 2007

SPSS Inc. (2003). *SPSS Advanced Models 12.0* [en línea] Chicago, USA Consultado el 30 de abril de 2012 en <http://www.docs.is.ed.ac.uk/skills/documents/3639/SPSSAdvancedModels12.0.pdf>

Wackerly, Dennis D. Mendenhall, William III. Scheaffer, Richard L. "Estadística Matemática con aplicaciones", CENCAGE Learning, 7ma Edición, 2010.