**Familia: Barrera o facilitador en el proceso de inclusión laboral de 3 poblaciones en edad productiva**

Marco teórico

# Escalamiento óptimo: Análisis de componentes principales categórico

El objetivo del escalamiento óptimo es asignar cuantificaciones numéricas a las categorías de una variable cualitativa, esto para aplicar los métodos estadísticos tradicionales en la solución de un problema. En este sentido el análisis de componentes principales categórico conocido por el acrónimo CATPCA, del inglés CATegorical Principal Components Analysis, busca la reducción de un conjunto original de variables en un conjunto más pequeño de componentes independientes que representen la mayor parte de la información encontrada en las variables originales, esto a través de la cuantificación óptima de las variables cualitativas bien sean estas de tipo: nominal u ordinal.

El procedimiento opera de la siguiente forma:

* En primera medida se asignan las restricciones correspondientes, dadas por la escala de medición de cada una de las variables en la tabla de datos.
* A continuación, el método determina dos tipos de parámetros: parámetros de estimación del modelo del análisis de componentes principales y parámetros de cuantificación óptima. De este modo los parámetros de cuantificación óptima se inicializan con valores aleatorios.
* Acto seguido el algoritmo ajusta un Análisis de componentes principales clásico (PCA) sobre los parámetros de cuantificación óptima iniciales, los cuales se van actualizando durante el proceso, reemplazando los parámetros viejos por los nuevos, hasta conseguir la convergencia.

El fin de efectuar un CATPCA para la presente investigación consiste en cuantificar de manera apropiada las variables categóricas que componen la base de datos con el propósito de identificar la influencia de las componentes de entorno sobre la inclusión laboral, luego posterior a la cuantificación apropiada de las variables se procede a utilizar estas con el método PLS-PM explicado en el siguiente apartado.

# Partial Least Squares – Path Modeling (Mínimos cuadrados parciales – Modelación de rutas)

Metodología

El proceso metodológico desarrollado para estudiar el efecto que tienen diferentes componentes de entorno sobre la inclusión laboral en los tres grupos poblaciones analizados (población con discapacidad cognitiva, auditiva-sensorial y población sin discapacidad), abarcó las siguientes etapas:

1. Análisis descriptivo de datos.
2. Exploración de relaciones entre las variables (análisis de correlación).
3. Identificación de variables que definan las componentes de entorno orientadas hacia la inclusión labora en la población en estudio. Para este fin se especificaron los siguientes constructos:

* Componente ambiental (constructo de primer orden)
* Componente de apoyos y servicios (constructo de primer orden)
* Componente de calidad de vida (constructo de primer orden)
* Inclusión laboral (constructo de segundo orden)

1. Escalamiento óptimo de variables categóricas
2. Ajuste y evaluación del modelo estructural PLS-PM.

**Fase 1. Análisis descriptivo de datos**

Previo al proceso de escalamiento óptimo de las variables categóricas se realizó un análisis descriptivo de las variables más importantes que caracterizan la composición de la muestra, utilizando estadísticas de resumen como medidas de tendencia central y variabilidad.

**Fase 2. Exploración de relaciones entre variables**

Como primer paso para identificar las posibles relaciones de dependencia entre los grupos de variables, se parte de la realización de un test de independencia Chi-cuadrado entre todos los pares de variables. Posteriormente haciendo uso de los coeficientes de correlación no lineales para variables categóricas se identificó la dirección de las relaciones previamente detectadas por el test de independencia Chi-cuadrado.

**Fase 3. Identificación de variables por componentes de entorno**

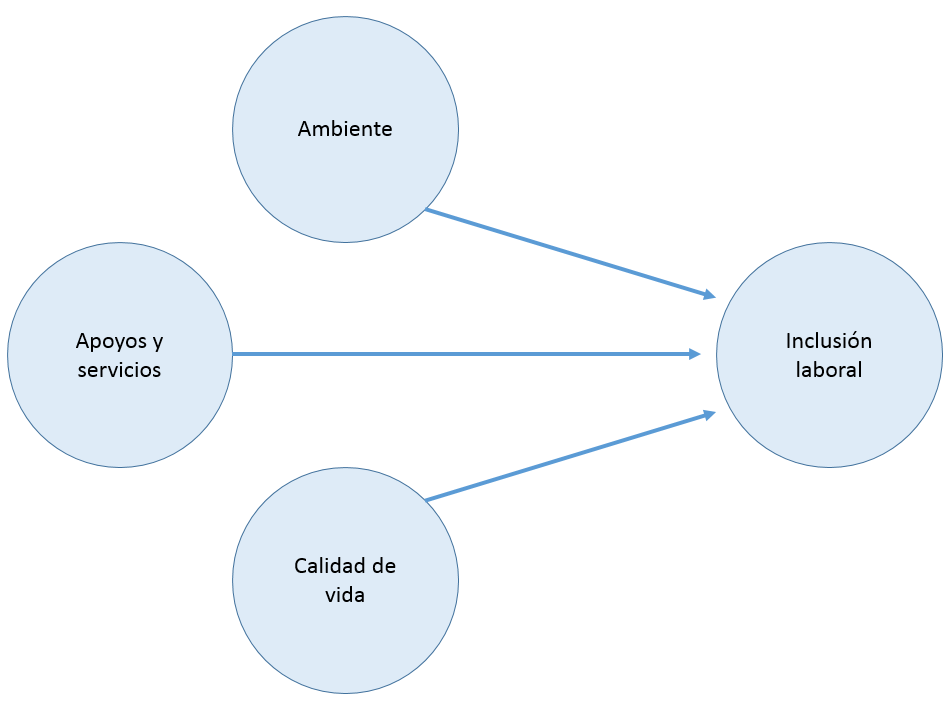
Posterior al análisis de correlación donde se identificó la relación existente entre las variables medidas, se realiza una selección de las variables que presentan mayor importancia en la explicación de cada componente de entorno, esto para depurar el modelo de ítems que no midan lo que realmente se necesita medir.

**Fase 4. Escalamiento óptimo de variables categóricas**

A partir de la utilización del Análisis de componentes principales categórico se cuantificó de manera óptima las variables categóricas de la base de datos. Esto por cada componente de entorno definida. Las variables cuantificadas son las que sirven de insumo para el modelo PLS-PM donde se busca estimar las componentes de entorno, entendiéndose estas como factores latentes no observables.

**Fase 5. Ajuste y evaluación del modelo PLS-PM**

Acto seguido se ajustó el modelo PLS-PM, el cual en primera medida estima las componentes de entorno a partir de las variables con escalamiento óptimo y posteriormente explora la relación entre las componentes de entorno y la inclusión laboral como se muestra en el siguiente diagrama de rutas.

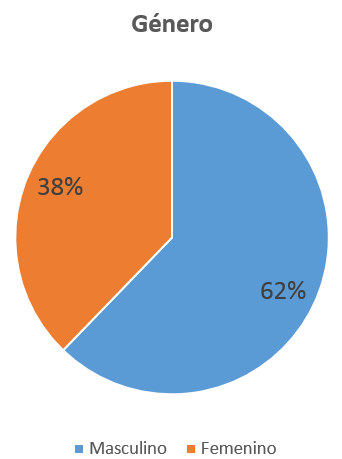


La inclusión laboral se consideró como un constructo de segundo orden, es decir, que no posee ninguna variable observada que lo defina y se determinó a partir de las componentes de entorno orientadas de forma adecuada.

Resultados

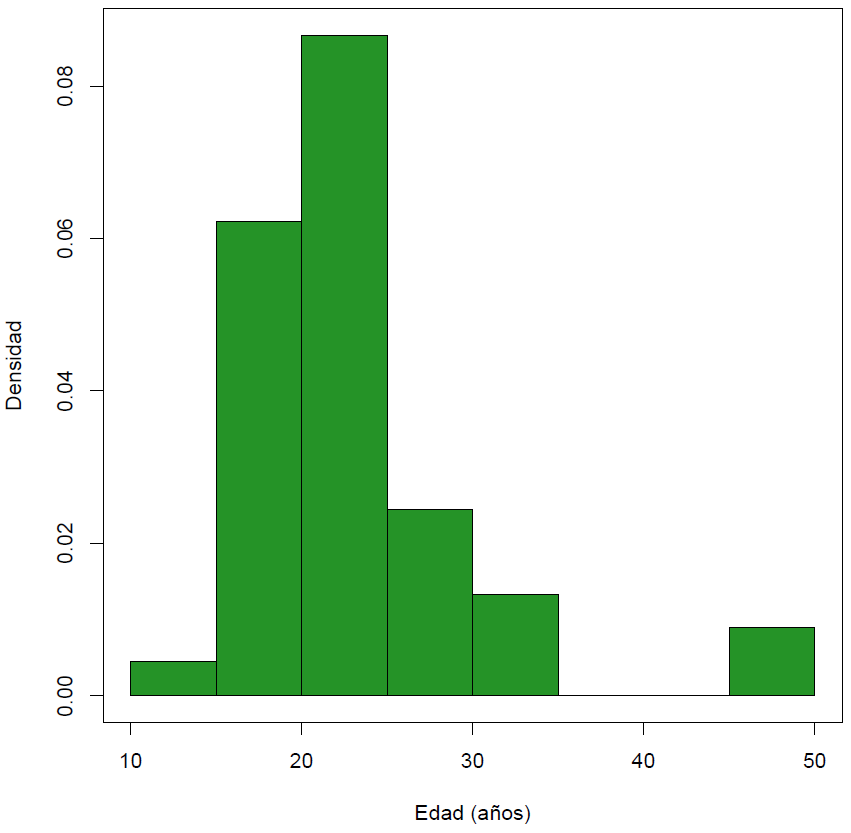
**Análisis descriptivo de datos**

**Distribución porcentual del género**. Conclusión rápida: la mayor parte de la muestra (62%) está compuesta por hombres.



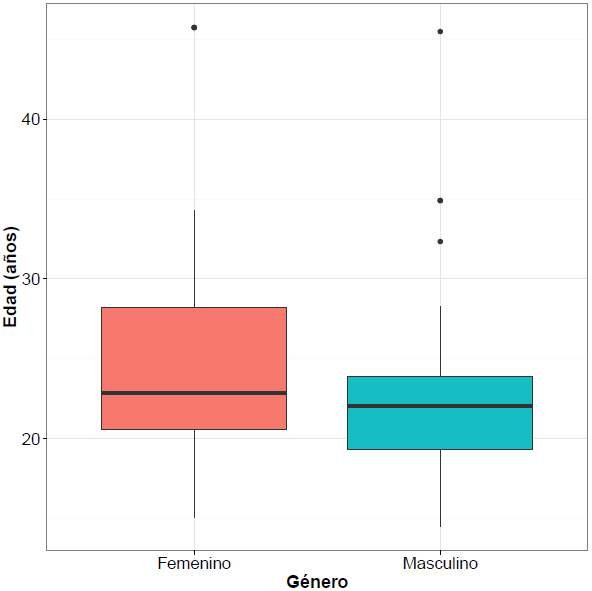
|  |  |  |
| --- | --- | --- |
| Sexo | Conteo | Porcentaje |
| Masculino | 56 | 62% |
| Femenino | 34 | 38% |

**Distribución de la edad**. Conclusión rápida: en base a la tabla de estadísticas descriptivas la edad mínima es de 14 años mientras la máxima es de 45 años, por otro lado el promedio de edades oscila alrededor de los 23.4 años con una desviación estándar de 6.7 años.



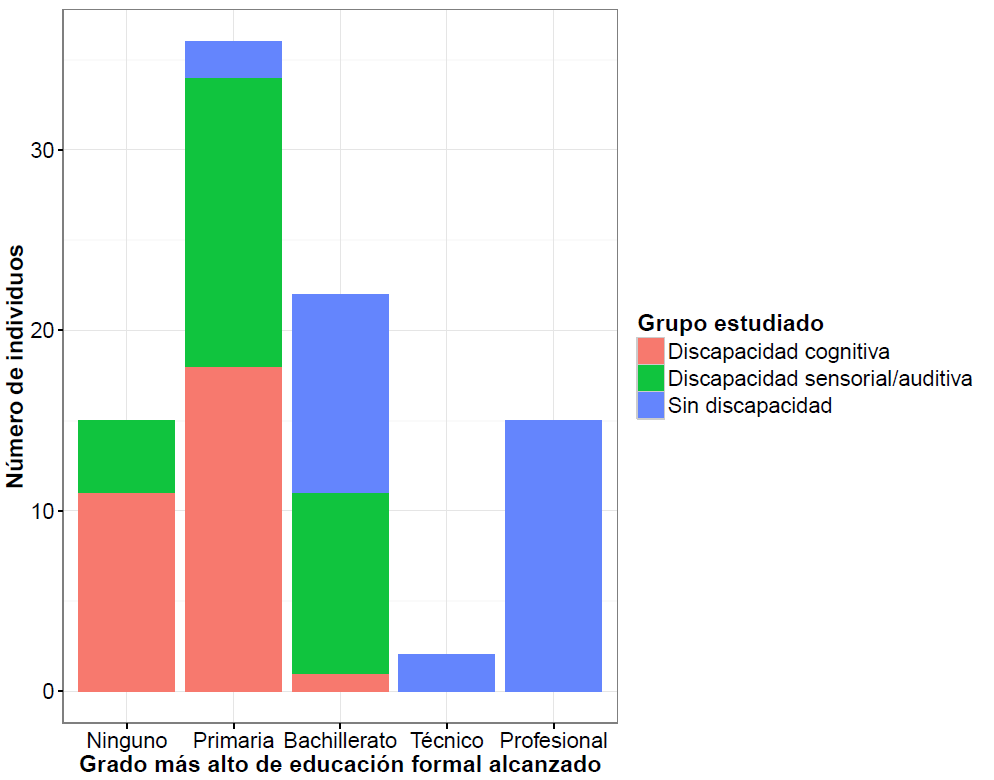
|  |  |  |  |
| --- | --- | --- | --- |
| Mínimo | Media | Desviación estándar | Máximo |
| 14.53 | 23.38 | 6.71 | 45.71 |

**Distribución de la edad según el género**. Conclusión rápida: similaridad en las edades de la población participante en el estudio, aunque las mujeres exhiben una edad ligeramente mayor a la de los hombres. Ver tabla con estadísticas descriptivas.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Género | Mínimo | Media | Desviación estándar | Máximo |
| Masculino | 14.53 | 22.25 | 5.46 | 45.45 |
| Femenino | 15.08 | 25.25 | 8.14 | 45.71 |

**Distribución del nivel educativo más alto alcanzado en función de la condición del encuestado**. Conclusión rápida: grupo sin condición de discapacidad alcanza el nivel educativo más alto (profesional), mientras el grupo con discapacidad cognitiva oscila entre ningún estudio y primaria (se tiene un caso donde alcanzó el bachillerato) y finalmente el grupo con discapacidad sensorial o auditiva tiene una mayor frecuencia en los niveles de primaria y bachillerato.



|  |  |  |  |
| --- | --- | --- | --- |
|  | Discapacidad cognitiva | Discapacidad sensorial/auditiva | Sin discapacidad |
| Ninguno | 11 | 4 | 0 |
| Primaria | 18 | 16 | 2 |
| Bachillerato | 1 | 10 | 11 |
| Técnico | 0 | 0 | 2 |
| Tecnología | 0 | 0 | 0 |
| Profesional | 0 | 0 | 15 |

**Estado civil**. Conclusión rápida: el 96% de las personas nunca ha estado casada



|  |  |  |
| --- | --- | --- |
| Estado civil | Conteo | Porcentaje |
| Nunca ha estado casado | 86 | 96% |
| Viviendo en pareja | 4 | 4% |

Distribución de conteos por componente de entorno. Pendiente

**Análisis de correlación**

Resultados del test Chi-cuadrado de independencia. Conclusión rápida: Del total de variables existe algún tipo de asociación entre el 53% de todos los posibles pares de variables. Lo que indica una fuerte asociación entre ellas, lo cual da pie para los siguientes análisis. El color rojo del heatmap (mapa de calor) indica la existencia de una relación estadísticamente significativa entre el par de variables analizadas. Mientras el color azul indica la independencia entre dos variables.

