

RESUMEN EJECUTIVO

Este informe presenta los hallazgos clave y conclusiones derivadas del análisis para proponer un Índice de Marginación Alternativo basado en el método de Análisis de Componentes Principales (PCA). El objetivo es proporcionar una visión simplificada pero informativa sobre la marginación en los municipios, respaldada por una metodología robusta y fácil de comprender.

A. ANÁLISIS DE PCA DE LOS INDICADORES DE CONAPO

Se realizó un Análisis de Componentes Principales (PCA) sobre los 9 indicadores de marginación proporcionados por CONAPO. Los resultados destacan que el primer componente principal explica la mayoría de la variabilidad en los datos.

A continuación presento una tabla que muestra el porcentaje de varianza explicada por cada componente principal, es decir, estamos trabajando con un conjunto de datos que tiene muchas dimensiones, es decir, muchas variables diferentes que describen diferentes aspectos. La varianza es una medida de cuánto varían estos datos. Cuando reducimos la dimensionalidad de nuestros datos utilizando un método llamado Análisis de Componentes Principales (PCA), obtenemos nuevas variables llamadas componentes principales. Cada uno de estos componentes principales captura una parte de la variabilidad total de los datos.

En esta curva, se representa cada uno de estos componentes principales. La altura de la curva indica cuánta variabilidad de los datos captura ese componente en particular. Si sumamos todas las alturas de las barras, obtenemos el total de la variabilidad original de los datos.

En resumen, esta gráfica nos ayuda a entender cuánta información estamos reteniendo al reducir la complejidad de nuestros datos. Es una forma de simplificar la información sin perder demasiada variabilidad, lo que facilita su interpretación y análisis.

En el contexto de la información proporcionada sobre el análisis de componentes principales (PCA), es importante destacar que el PCA se utiliza para reducir la dimensionalidad de los datos, pero no proporciona directamente información sobre fenómenos sociales como la marginación en un país. Sin embargo, podemos hacer algunas observaciones generales basadas en las cargas de las variables en las componentes principales.

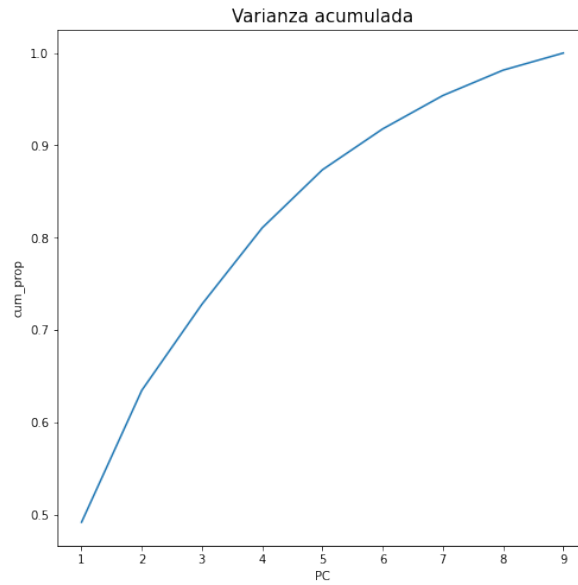


Figura 0.1: Varianza acumulada del análisis de PCA

Al momento de realizar este análisis obtenemos una tabla como la que se muestra a continuación:

	pc1	pc2	pc3	pc4	pc5	pc6	pc7	pc8	pc9
ANALF	0.41	-0.11	-0.09	-0.18	0.01	0.55	-0.09	-0.19	-0.65
SBASC	0.39	-0.29	0.11	0.02	0.07	0.51	0.18	0.01	0.67
OVSDE	0.24	0.51	0.50	-0.17	-0.51	0.08	-0.18	0.33	0.04
OVSEE	0.27	0.56	0.23	0.17	0.47	-0.09	0.45	-0.32	-0.05
OVSAE	0.24	0.28	-0.58	0.62	-0.37	0.06	-0.03	-0.00	0.06
OVPT	0.37	0.12	-0.27	-0.14	0.54	-0.17	-0.39	0.53	0.05
VHAC	0.36	0.02	-0.25	-0.50	-0.23	-0.41	-0.11	-0.52	0.22
PL5000	0.30	-0.33	0.46	0.50	0.05	-0.29	-0.45	-0.23	-0.07
PO2SM	0.37	-0.36	0.03	0.01	-0.21	-0.38	0.59	0.39	-0.23

Cuadro 0.1: Resultado del análisis PCA

Al observar las cargas de las variables en las componentes principales, notamos que algunas variables tienen un impacto más fuerte en la variabilidad total. Por ejemplo, si ".ANALF" tiene una carga alta en varias componentes, podría indicar que la alfabetización está vinculada a diferentes aspectos en tus datos. Las variables relacionadas con la educación (".ANALF") y el acceso a servicios básicos ("SBASC") suelen ser indicadores socioeconómicos. Estas variables son relevantes, ya que podrían señalar áreas donde la población puede enfrentar desafíos, como falta de educación o limitado acceso a servicios esenciales. Es crucial considerar el contexto específico del país. Por ejemplo, si "SBASC" refleja el acceso a servicios de salud, su carga en ciertas componentes podría sugerir disparidades en la atención médica y posiblemente contribuir a una comprensión más profunda de la marginación.

B. CONSTRUCCIÓN DEL ÍNDICE DE MARGINACIÓN ALTERNATIVO

El primer componente principal obtenido del PCA se utilizó para construir un Índice de Marginación Alternativo. Este índice condensa la información de los 9 indicadores en una única medida, preferentemente en el rango $[0, 1]$.

Para la construcción de este índice de marginación tomamos en cuenta los pesos de la primera componente en cada uno de los 9 índices que teníamos disponibles, para así después para hacerlo un número representativo se normalizo

Índice de Marginación Alternativo (IMalternativo):

Utilizando la combinación de variables en PC1, hemos creado un nuevo índice llamado IMalternativo. Este índice intenta representar una medida de marginación, combinando información de educación, acceso a servicios, y otras variables importantes.

Normalización del Índice (IMalternativo):

Normalizar significa ajustar los valores del índice para que estén en un rango específico, en este caso, de 0 a 1. Esto facilita la comparación y comprensión de cómo se distribuyen los valores del índice.

Además, se exploró la variabilidad al considerar el segundo y tercer componente principal. Sin embargo, se concluyó que el primer componente principal es la opción más representativa y útil para la construcción del índice.

INDICE	IMM-CONAPO	IMM-CP1	IMM-CP2	IMM-CP3
1	Benito Juárez	Benito Juárez	San Miguel Santa Flor	San Antonino Castillo
2	San Pedro Garza García	San Pedro Garza García	San Miguel Ahuehuetitlán	Copalillo
3	San Nicolás de los Garza	San Nicolás de los Garza	Santiago Yaitepec	Chanal
4	Cuauhtémoc	Cuauhtémoc	Santos Reyes Tepellijo	Tezoyuca
5	Apodaca	Apodaca	Santiago Texcalcingo	Oxchuc

Cuadro 0.2: Comparación de índices calculados por la CONAPO y por el estudio de PCA.

Como podemos ver en la tabla anterior, nuestro índice calculado está tomando los 5 mejores municipios al igual que el índice calculado por la CONAPO. Esto es un buen indicador pero necesitaremos hacer más pruebas.

También como parte de las primeras vistas, vemos que al tomar en cuenta la segunda y tercer componente no encontramos lógica en lo obtenemos debido a que como se explico antes, no alcanzamos a cubrir esta variabilidad de los datos. Ahora para poder nosotros ver cuál será

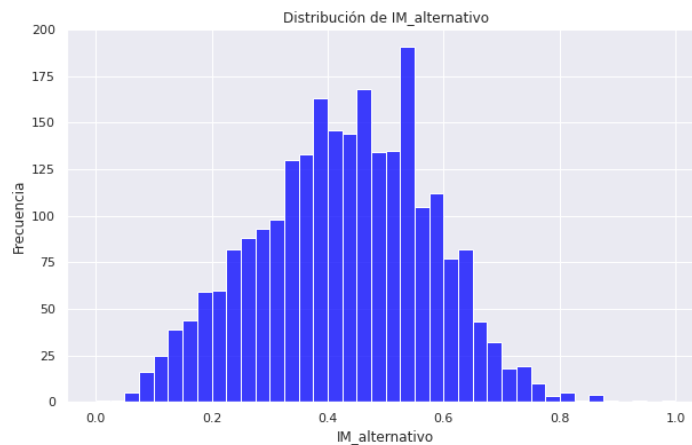


Figura 0.2: Frecuencia de datos en el IMM-Alternativo

la forma en que dividiremos los datos, trabajaremos con la gráfica de barras anterior, que es

una gráfica de frecuencias, lo principal que queremos demostrar aquí es que tantos valores de IM-alternativo tenemos en ese rango, esto con el fin de poder proponer una forma para dividir los datos en las categorías: Muy bajo, bajo, medio, alto y muy alto. Por lo que al definir nuestra metodología, podemos empezar a clasificar cada uno de nuestros municipios y obtenemos lo siguiente:

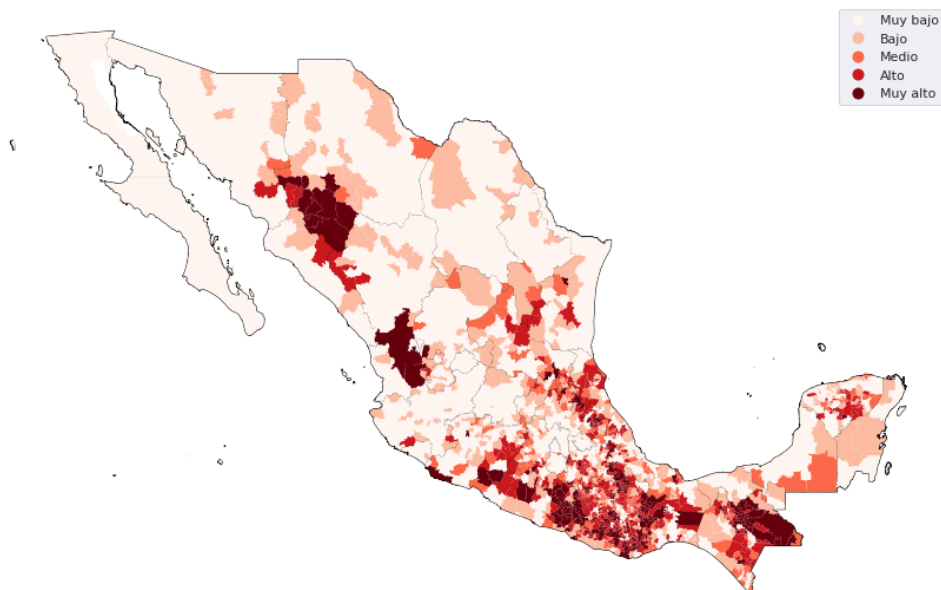


Figura 0.3: Índice de marginación nacional propuesto con la metodología de PCA

CONCLUSIONES

- El análisis de PCA proporciona una herramienta eficaz para simplificar la información de múltiples indicadores de marginación.
- El Índice de Marginación Alternativo, basado en el primer componente principal, ofrece una medida consolidada que puede ser más fácilmente interpretada.
- La inclusión de información adicional puede mejorar la precisión y completitud del índice. Este informe busca proporcionar una base sólida para la toma de decisiones en la Secretaría del Desarrollo Social, permitiendo una comprensión más clara y eficiente de la marginación en los municipios.

ANEXOS TÉCNICOS

1. Se anexa el código para poder hacer el PCA

```

1 from sklearn.decomposition import PCA
2 pca = PCA()
3 # ajustar en los datos (estandarizados)
4 pca.fit(X)
5 dat = {'PC':range(1,10),'std':np.sqrt(pca.explained_variance_), 'var_prop':pca.explained_variance_ratio_,
6 'cum_prop':np.cumsum(pca.explained_variance_ratio_)}
7 stds = pd.DataFrame(data = dat)
8 stds

```

```

9 comps = pd.DataFrame(data=pca.components_.T,columns=['pc1','pc2','pc3','pc4','pc5','pc6','pc7','pc8','pc9'
10 ],
11 index=indicadores.columns)
12 print(comps)

```

Proyecciones:

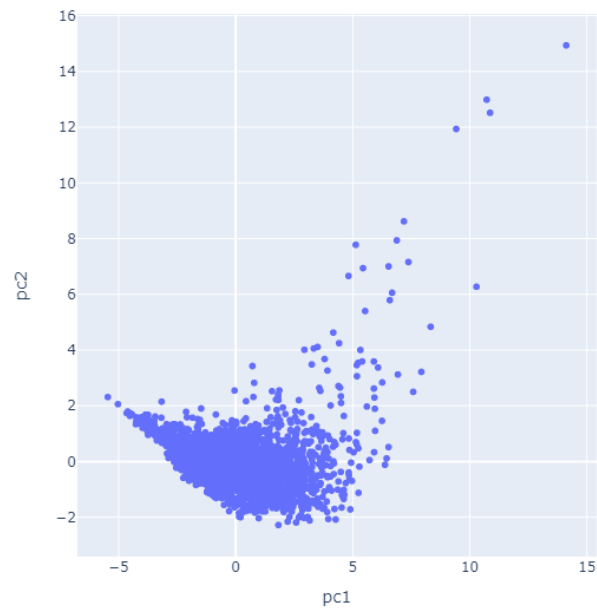


Figura 0.4: Proyección en las dos primeras componentes del PCA

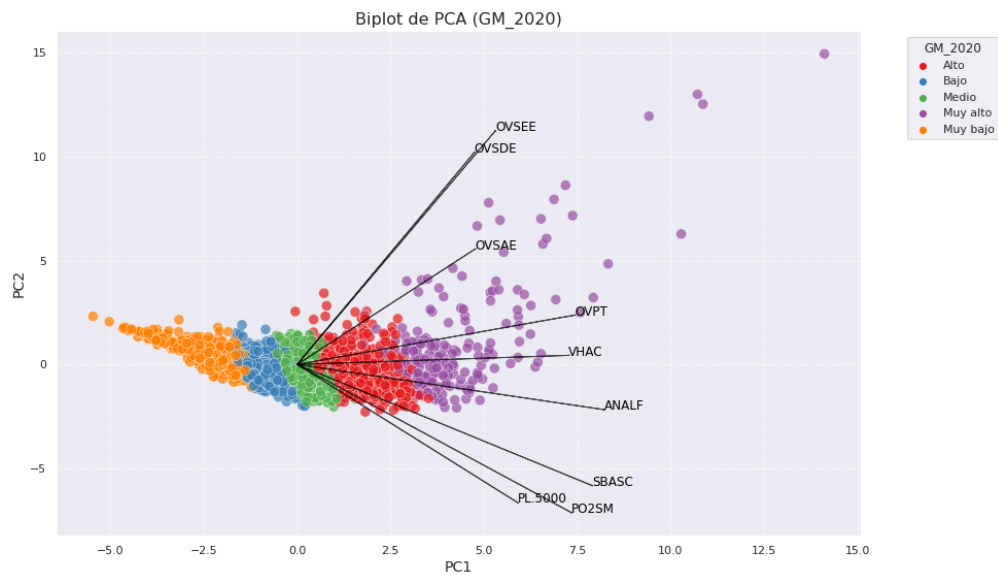


Figura 0.5: Biplot desallorado con el fin de ver las magnitudes y direcciones de los indicadores

2. Construcción del Índice Alternativo

```

1 # Obtener las cargas del primer componente principal
2 PC1 = comps['pc1']
3
4 #IM alternativo
5 IM_alternativo = np.dot(imm[['ANALF', 'SBASC', 'OVSDE', 'OVSEE', 'OVSAE', 'OVPT', 'VHAC', 'PL
    .5000', 'PO2SM']].values, PC1)
6
7 # Normalizar [0, 1]
8 IM_alternativo = (IM_alternativo - np.min(IM_alternativo)) / (np.max(IM_alternativo) - np.min(IM_
    alternativo))
9 imm['IM_alternativo'] = IM_alternativo
10 imm.head()

```

3. Propuestas para incluir en mi índice:

La elaboración de un índice puede beneficiarse de la inclusión de diversas fuentes de información para proporcionar una visión más completa y precisa, por ejemplo

Datos socioeconómicos adicionales:

Tasa de desempleo.

Ingresos medios y distribución de ingresos.

Acceso a servicios de salud y calidad de atención médica.

Acceso a servicios educativos y calidad de la educación.

Datos de vivienda:

Costos de vivienda y tasas de propiedad.

Condiciones de vivienda y hacinamiento.

Acceso a servicios básicos de vivienda.

Datos de movilidad y transporte:

Accesibilidad y calidad de las infraestructuras de transporte.

Disponibilidad y uso de transporte público.

Congestión del tráfico.

4. Confiabilidad de la metodología

De primera vista estoy de acuerdo con la metodología, pero se que podria hacerse aun mucho más para poder mejorar este indice, hablando tal vez desde iniciar con mejorar la forma de calcular los indices y agregar algunos otros más.

Considero que PCA si cumple con los requisitos, solamente habria que buscar estrategias para poder hacer una buena clasificación de acuerdo a la información arrojada.