# Preprocesamiento de Datos

Para el análisis de la información, se cargaron los archivos mediante el software de R Studio usando las librerías de 'sqldf’ (usa el software SQLite para crear las bases de datos), 'readxl' (para poder cargar los archivos de Excel fuentes de información para las bases de datos y 'stringr' (para hacer la edición de textos) considerando los archivos *Municipios.xlsx* y *Prestadores.xlsx* (se modificó el archivo original el cual tenía extensión *.xls*).

Entre los preprocesamientos de datos, en la base de datos de Municipios se eliminaron caracteres especiales (!&?!’$#%>\*), se reemplazaron los espacios múltiples a espacios sencillos y se eliminaron todos los espacios al inicio o final de los textos. En la base de datos de Prestadores se cambiaron los nombres del campo *depu\_nombre* de los distritos especiales, para que correspondan a sus respectivos departamentos. Se hicieron que todos los nombres de municipios y departamentos estuvieran capitalizados en ambas bases y se verificaron las inconsistencias de nombres entre las bases (Ej.: Sincé – San Luis de Sincé). Finalmente, se observa que solo faltan datos para el municipio de Guachené en Cauca, del que no se tiene población ni índice de ruralidad.

|  |
| --- |
| Figura 1: Porcentaje acumulado prestadores por cada 10.000 personas por municipios  A graph with a line drawn on it  Description automatically generated |

# Análisis de la Información

De forma que se pueda tener un indicador comparable acorde las dimensiones de cada municipio y ciudad considerada, se toma el número de prestadores por cada 10.000 habitantes para los análisis realizados.

|  |
| --- |
| Figura 2: Modelo regresión de número de prestadores por cada 10.000 habitantes VS Población, Índice de ruralidad y Región  A screenshot of a computer  Description automatically generated |

Al analizar la distribución de los prestadores según el número de personas en los diferentes municipios (Figura 1), muestra que el 12.5% de los municipios no tienen prestadores registrados y que el 75% de los municipios tienen menos de 6 prestadores por cada 10.000 habitantes. Solamente el 10% de los municipios registran más de 10 prestadores por cada 10.000 habitantes, indicando una alta concentración de los prestadores en ciertos conjuntos de ciudades.

Para poder determinar cuáles variables son representativas para modelar el número de prestadores por cada 10.000 habitantes, se utilizaron modelos lineales generalizados (Figura 3). Se normalizaron los datos bajo la metodología min-max de cada columna y se convirtió la variable categórica Región en un conjunto de variables binarias, omitiendo la región Pacífico.

Se observa que tanto las variables de Población, Índice de ruralidad y Región son relevantes, dada la posibilidad de rechazar la hipótesis nula dado su p-valor.

|  |
| --- |
| Figura 2: Comparación de Números de Prestadores por 10.000 habitantes según el índice de ruralidad por Región  A graph with different colored dots  Description automatically generated |

Los coeficientes con mayor magnitud corresponden a aquellos correspondientes a la Población (0.408), Índice de Ruralidad (-0.373), y el intercepto (0.24) el cual integra de manera implica el valor base de los prestadores de la Región Pacífico. Estos análisis indican que el número de prestadores por habitante aumentan a mayor población, generando un desbalance sobre comunidades con pocos habitantes y que a mayor ruralidad se tienen menores prestadores. Se observa también que las regiones que contienen mayor concentración de prestadores por habitante serían el Llano, Eje cafetero y Centro sur.

Continuando el análisis comparando por Índice de ruralidad, se observa que el número de prestadores por 10.000 habitantes tiene un comportamiento decreciente según el nivel de ruralidad (Figura 2). Se observa que el mayor número de prestadores por 10.000 habitantes se encuentra en el municipio de Puerto Colombia (Región Caribe) con un valor de 52.42, presentando un índice de ruralidad de 27. Se observa que solo existe un municipio con más de 10 prestadores por 10.000 habitantes y un índice de ruralidad superior a 70. Se observa que el Llano tiene la menor cantidad de prestadores por 10.000 habitantes en sus municipios, mientras que el Eje cafetero y la Región Centro Oriente tiene la mayor cantidad de municipios con un alto número de prestadores según el número de habitantes en ruralidades bajas.

|  |
| --- |
| Figura 4: Prestadores promedio por nivel según el índice de ruralidad del municipio. |

Se observa que la gran mayoría de prestadores no tienen un nivel reportado. Considerando exclusivamente aquellos prestadores con un nivel 1,2 o 3 reportado (Figura 4), se observa que el promedio más alto de prestadores por municipio se encuentra en centros urbanos con un nivel de ruralidad 6.

Existe un prestador de nivel 3 en un municipio con nivel de ruralidad de 47 y fuera de este, solo se ven presentes en municipios con una ruralidad de nivel 34. No se observan prestadores con ningún nivel para valores superiores a los 82 puntos de ruralidad.