

Estadística para la Ciencia de Datos

Exámen final: Estadística y probabilidad aplicada

Manuel Alejandro Serrano Macias

Fecha de modificación: 2025-11-06 18:49:33.042425

Contents

Reporte de Análisis Estadístico Exploratorio de Marginación Municipal 2020	1
Prerequisitos	1
a) Mostrar el rango, rango intercuantílico, media, mediana y desviación estándar	1
b) Gráfico de correlación	13
b) Gráfico de densidad del Índice de marginación normalizado	15
c) Población Total Índice de Marginación	17
d) Población Total Índice de Marginación	18
Distribuciones de Probabilidad	20
1. Ejercicio concurso de TV	20

Reporte de Análisis Estadístico Exploratorio de Marginación Municipal 2020

A continuación se presenta un análisis estadístico exploratorio del Índice de Rezago Municipal 2020. Los datos son extraídos del Consejo Nacional de Población (CONAPO)

Prerequisitos

1. Se instalan y preparan las librerías a utilizar durante el análisis.

```
#install.packages("knitr")
#install.packages("readxl")
#install.packages("corrplot")
library(readxl)
#library(knitr)
library(corrplot)
```

2. Carga y procesamiento de datos para el análisis. Se cargan los datos de Marginación Municipal 2020 (CONAPO)

```
marginacion_base <- read_excel("IMM_2020.xlsx", sheet = "IMM_2020")
```

a) Mostrar el rango, rango intercuantílico, media, mediana y desviación estándar

de las nueve variables porcentuales (ANALF, SBASC, OVSDE, OVSEE, OVSAE, OVPT, VHAC, PL.5000, PO2SM) de cada uno de las 32 entidades federativas.

Se definen las variables que incluiremos en el estudio, iteramos en cada estado para calcular las estadísticas de interés de cada variable mediante las siguientes instrucciones:

```
variables_estudio <-  
  c("ANALF", "SBASC", "OVSDE", "OVSEE", "OVSAE", "OVPT", "VHAC", "PL.5000", "PO2SM")  
  
entidades <- unique(marginacion_base$NOM_ENT)  
  
for (variable in variables_estudio) {  
  marginacion_base[[variable]] <- as.numeric(marginacion_base[[variable]])  
}  
  
resultados <- list()  
  
for (estado in entidades) {  
  datos_estado <- marginacion_base[marginacion_base$NOM_ENT == estado, ]  
  estadisticas_estado <- list()  
  
  for (variable in variables_estudio) {  
    estado_actual <- marginacion_base[[variable]]  
    rango <- range(estado_actual, na.rm = TRUE)  
    rango_intercuantilico <- IQR(estado_actual, na.rm = TRUE)  
    media <- mean(estado_actual, na.rm = TRUE)  
    mediana <- median(estado_actual, na.rm = TRUE)  
    desviacion_estandar <- sd(estado_actual, na.rm = TRUE)  
  
    calculo_estadisticas <- data.frame(  
      Variable = variable,  
      Rango_Min = rango[1],  
      Rango_Max = rango[2],  
      Rango_Intercuantilico = rango_intercuantilico,  
      Media = media,  
      Mediana = mediana,  
      Desviacion_Estandar = desviacion_estandar  
    )  
  
    estadisticas_estado[[variable]] <- calculo_estadisticas  
  }  
  
  df_estado <- do.call(rbind, estadisticas_estado)  
  rownames(df_estado) <- NULL  
  
  resultados[[estado]] <- df_estado  
}
```

Finalmente, se crea una lista principal de data frames para cada estado.

NOTA: Originalmente hacia el proceso con un ciclo, sin embargo, las tablas no se imprimían en el pdf de manera amigable, investigué y se debe a que kable, que es el formato que especifico en el encabezado del Markdown para impresión de data frames no puede imprimir de manera adecuada dentro de un ciclo, la documentación recomienda tener un chunk de R por cada tabla para así tenerla impresa en el PDF de manera amigable. Seguramente hay una mejor manera de hacerlo, lo investigaré para futuros reportes.

```
# Código original solo para referencia
```

```
# for (estado in sort(names(resultados))) {
```

```
# dataframe_estado <- resultados[[estado]]
#
# cols_numericas <- c("Rango_Min", "Rango_Max", "Rango_Intercuantilico",
#                     "Media", "Mediana", "Desviacion_Estandar")
#
# dataframe_estado[cols_numericas] <-
#   lapply(dataframe_estado[cols_numericas], round, digits = 3)
#
# print(knitr::kable(dataframe_estado, caption = estado))
# }
```

```
imprimir_resultados <- function(data_frame, estado) {
  knitr::kable(data_frame,
    caption = estado,
    digits = 3,
    booktabs = TRUE,
    longtable = FALSE)
}
```

```
imprimir_resultados(resultados[["Aguascalientes"]], "Aguascalientes")
```

Table 1: Aguascalientes

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Baja California"]], "Baja California")
```

Table 2: Baja California

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Baja California Sur"]], "Baja California Sur")
```

Table 3: Baja California Sur

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Campeche"]], "Campeche")
```

Table 4: Campeche

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Coahuila de Zaragoza"]], "Coahuila de Zaragoza")
```

Table 5: Coahuila de Zaragoza

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Colima"]], "Colima")
```

Table 6: Colima

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Chiapas"]], "Chiapas")
```

Table 7: Chiapas

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Chihuahua"]], "Chihuahua")
```

Table 8: Chihuahua

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Ciudad de México"]], "Ciudad de México")
```

Table 9: Ciudad de México

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Durango"]], "Durango")
```

Table 10: Durango

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Guanajuato"]], "Guanajuato")
```

Table 11: Guanajuato

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Guerrero"]], "Guerrero")
```

Table 12: Guerrero

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Hidalgo"]], "Hidalgo")
```

Table 13: Hidalgo

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Jalisco"]], "Jalisco")
```

Table 14: Jalisco

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["México"]], "México")
```

Table 15: México

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Michoacán de Ocampo"]], "Michoacán de Ocampo")
```

Table 16: Michoacán de Ocampo

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Morelos"]], "Morelos")
```

Table 17: Morelos

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Nayarit"]], "Nayarit")
```

Table 18: Nayarit

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Nuevo León"]], "Nuevo León")
```


Table 19: Nuevo León

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Oaxaca"]], "Oaxaca")
```

Table 20: Oaxaca

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Puebla"]], "Puebla")
```

Table 21: Puebla

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Querétaro"]], "Querétaro")
```

Table 22: Querétaro

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Quintana Roo"]], "Quintana Roo")
```

Table 23: Quintana Roo

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["San Luis Potosí"]], "San Luis Potosí")
```

Table 24: San Luis Potosí

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Sinaloa"]], "Sinaloa")
```

Table 25: Sinaloa

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Sonora"]], "Sonora")
```

Table 26: Sonora

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Tabasco"]], "Tabasco")
```

Table 27: Tabasco

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Tamaulipas"]], "Tamaulipas")
```

Table 28: Tamaulipas

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Tlaxcala"]], "Tlaxcala")
```

Table 29: Tlaxcala

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Veracruz de Ignacio de la Llave"]], "Veracruz de Ignacio de la Llave")
```

Table 30: Veracruz de Ignacio de la Llave

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Yucatán"]], "Yucatán")
```

Table 31: Yucatán

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

```
imprimir_resultados(resultados[["Zacatecas"]], "Zacatecas")
```

Table 32: Zacatecas

Variable	Rango_Min	Rango_Max	Rango_Intercuantilico	Media	Mediana	Desviacion_Estandar
ANALF	0.353	53.071	9.360	10.164	8.203	7.634
SBASC	5.535	88.328	20.119	45.853	46.339	13.982
OVSDE	0.000	64.450	2.691	3.160	1.428	5.289
OVSEE	0.000	53.065	1.312	1.501	0.828	2.769
OVSAB	0.000	81.788	6.407	6.118	2.452	9.246
OVPT	0.000	68.150	9.375	7.987	4.714	8.974
VHAC	3.950	69.564	14.096	26.566	25.000	10.587
PL.5000	0.000	100.000	59.870	69.900	100.000	35.268
PO2SM	28.453	100.000	17.005	82.144	84.643	11.830

b) Gráfico de correlación

Variables de estudio (ANALF, SBASC, OVSDE, OVSEE, OVSAE, OVPT, VHAC, PL.5000, PO2SM)

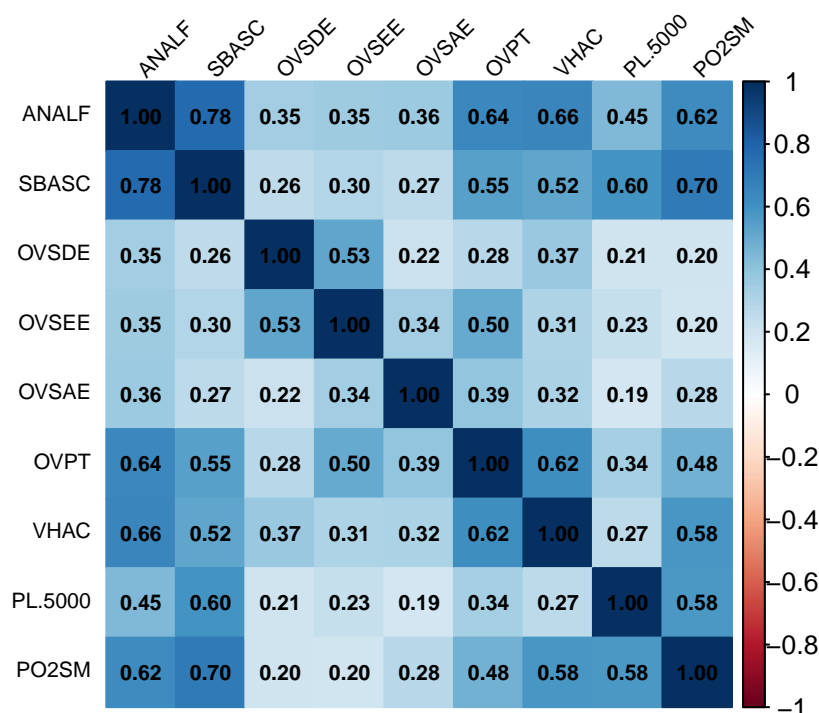
Para generar el gráfico de correlación creamos un nuevo data frame con las variables de estudio, posteriormente la función `cor` calcula la matriz de correlación de Pearson. Se utiliza la librería de `corrplot` para crear el gráfico.

```
datos_correlacion <- marginacion_base[, variables_estudio]
matriz_correlacion <- cor(datos_correlacion, use = "pairwise.complete.obs")
knitr::kable(round(matriz_correlacion, 2))
```

	ANALF	SBASC	OVSDE	OVSEE	OVSAE	OVPT	VHAC	PL.5000	PO2SM
ANALF	1.00	0.78	0.35	0.35	0.36	0.64	0.66	0.45	0.62
SBASC	0.78	1.00	0.26	0.30	0.27	0.55	0.52	0.60	0.70
OVSDE	0.35	0.26	1.00	0.53	0.22	0.28	0.37	0.21	0.20
OVSEE	0.35	0.30	0.53	1.00	0.34	0.50	0.31	0.23	0.20
OVSAE	0.36	0.27	0.22	0.34	1.00	0.39	0.32	0.19	0.28
OVPT	0.64	0.55	0.28	0.50	0.39	1.00	0.62	0.34	0.48
VHAC	0.66	0.52	0.37	0.31	0.32	0.62	1.00	0.27	0.58
PL.5000	0.45	0.60	0.21	0.23	0.19	0.34	0.27	1.00	0.58
PO2SM	0.62	0.70	0.20	0.20	0.28	0.48	0.58	0.58	1.00

```
corrplot(matriz_correlacion,
  method = "color", # Muestra cuadrados de color
  title = "Gráfico de Correlación de Indicadores de Marginación",
  mar = c(0, 0, 4, 0), # Adjust top margin for the title
  cex.main = 1,
  tl.col = "black", # Color de las etiquetas
  tl.srt = 45,      # Rotación de las etiquetas (para que no se encimen)
  addCoef.col = "black", # Agrega los valores numéricos
  number.cex = 0.7, # Tamaño de los números
  tl.cex = 0.7)
```

Gráfico de Correlación de Indicadores de Marginación



Observaciones

- Correlaciones Positivas Fuertes (Números altos, ej., $r > 0.7$):
 - Educación y Pobreza de Ingresos: Se espera una correlación muy fuerte entre ANALF (analfabetismo) y SBASC (población sin educación básica) con PO2SM (población con ingresos ≤ 2 salarios mínimos). Esto significa que donde hay alta falta de educación básica, hay mayor población con bajos ingresos, reforzando el ciclo de marginación.
 - Vivienda de Calidad: Los indicadores de carencias en la vivienda suelen estar muy correlacionados entre sí. Por ejemplo, OVSDE (sin drenaje/excusado), OVSEE (sin electricidad), OVSAE (sin agua entubada) y OVPT (piso de tierra) mostrarán una correlación positiva alta. Si un municipio carece de un servicio básico de vivienda, es muy probable que carezca de los demás.
 - Hacinamiento (VHAC): También se correlaciona positivamente con las otras carencias de vivienda, pues ambas son manifestaciones de pobreza de infraestructura.
- Correlaciones Negativas o Débiles (Números bajos, ej., $r < 0.3$):
 - Correlación entre Desarrollo de Servicios y Variables Rurales: Es posible encontrar correlaciones más débiles o incluso negativas entre variables como PL.5000 (población en localidades rurales/pequeñas) y algunos indicadores de vivienda, aunque suele ser más compleja. En general, en municipios con alta dispersión rural (PL.5000), las carencias de servicios (OVSDE, OVSAE) pueden ser elevadas, resultando en una correlación positiva.

La falta de correlación indicaría que la carencia de un servicio o condición no está sistemáticamente relacionada con la carencia de otra.

Conclusión: El gráfico de correlación sirve para confirmar la coherencia conceptual del Índice de Marginación. Las variables fueron elegidas porque se espera que se muevan juntas. Las áreas con cuadrados oscuros (cerca de 1) confirman que la marginación es un fenómeno multidimensional donde las carencias se acumulan (mala vivienda \rightarrow baja educación \rightarrow bajos ingresos).

b) Gráfico de densidad del Índice de marginación normalizado

Estados de estudio Nuevo León, Durango, Jalisco, Veracruz y Chiapas.

Para generar el gráfico de densidad se filtran los datos de las entidades federativas de interes, se calculan las densidades de cada estado, posteriormente se determinan los rangos para que las curvas quepan y creamos el gráfico.

```
entidades_interes <- c("Nuevo León", "Durango", "Jalisco", "Veracruz de Ignacio de la Llave", "Chiapas")
variable_imn <- "IMN_2020"
datos_entidades_estudio <- marginacion_base[marginacion_base$NOM_ENT %in% entidades_interes, ]
datos_entidades_estudio[[variable_imn]] <- as.numeric(datos_entidades_estudio[[variable_imn]])

# --- PASO 1: Filtrado de Datos por Entidad ---

nuevo_leon_imn <- na.omit(subset(marginacion_base, NOM_ENT == "Nuevo León")[[variable_imn]])
durango_imn <- na.omit(subset(marginacion_base, NOM_ENT == "Durango")[[variable_imn]])
jalisco_imn <- na.omit(subset(marginacion_base, NOM_ENT == "Jalisco")[[variable_imn]])
veracruz_imn <- na.omit(subset(marginacion_base, NOM_ENT == "Veracruz de Ignacio de la Llave")[[variable_imn]])
chiapas_imn <- na.omit(subset(marginacion_base, NOM_ENT == "Chiapas")[[variable_imn]])

densidad_nuevo_leon <- density(nuevo_leon_imn)
densidad_durango <- density(durango_imn)
densidad_jalisco <- density(jalisco_imn)
densidad_veracruz <- density(veracruz_imn)
densidad_chiapas <- density(chiapas_imn)

colores <- c("cadetblue", "coral", "antiquewhite4", "brown", "goldenrod")

min_x_vals <- c(min(densidad_nuevo_leon$x),
                min(densidad_durango$x),
                min(densidad_jalisco$x),
                min(densidad_veracruz$x),
                min(densidad_chiapas$x))

max_x_vals <- c(max(densidad_nuevo_leon$x),
                max(densidad_durango$x),
                max(densidad_jalisco$x),
                max(densidad_veracruz$x),
                max(densidad_chiapas$x))

rango_x <- c(min(min_x_vals), max(max_x_vals))

max_y_vals <- c(max(densidad_nuevo_leon$y),
                max(densidad_durango$y),
                max(densidad_jalisco$y),
                max(densidad_veracruz$y),
                max(densidad_chiapas$y))

rango_y <- c(0, max(max_y_vals))

plot(densidad_nuevo_leon,
      xlim = rango_x,
      ylim = rango_y,
```

```

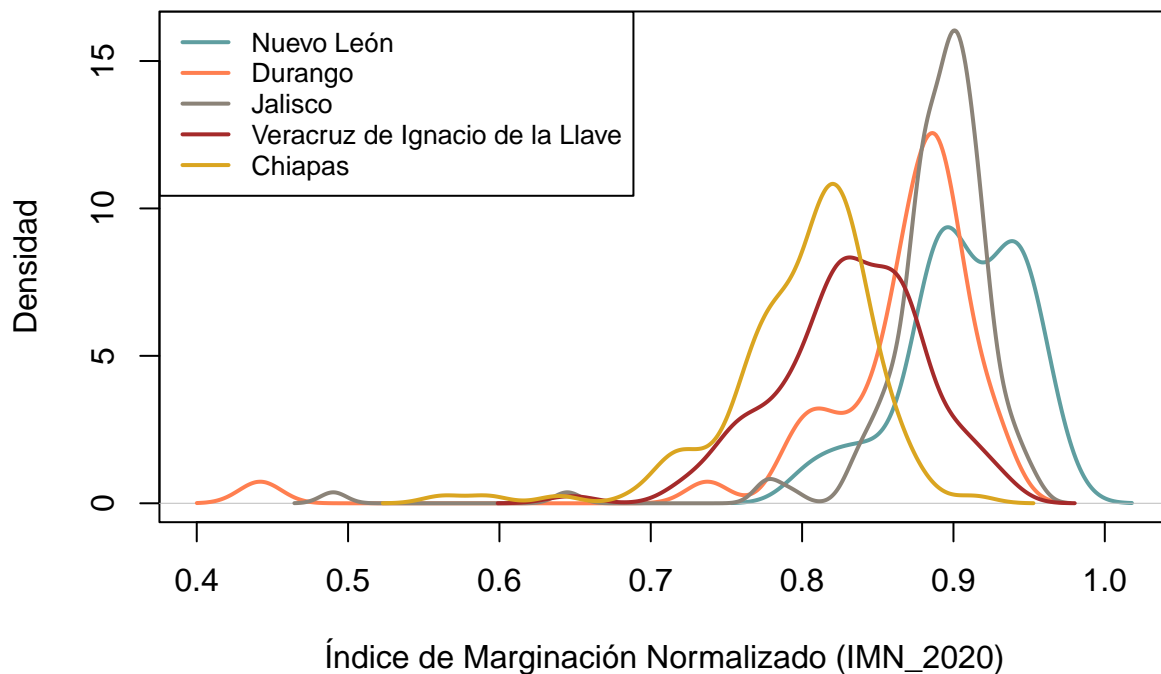
main = "Densidad del Índice de Marginación Normalizado (IMN_2020) por Entidad",
xlab = "Índice de Marginación Normalizado (IMN_2020)",
ylab = "Densidad",
col = colores[1],
lwd = 2)

lines(densidad_durango, col = colores[2], lwd = 2)
lines(densidad_jalisco, col = colores[3], lwd = 2)
lines(densidad_veracruz, col = colores[4], lwd = 2)
lines(densidad_chiapas, col = colores[5], lwd = 2)

legend("topleft",
      legend = entidades_interes,
      col = colores,
      lwd = 2,
      cex = 0.8)

```

Densidad del Índice de Marginación Normalizado (IMN_2020) por Entidad



El análisis comparativo de las curvas de densidad del Índice de Marginación Normalizado (IMN_2020) revela contrastes fundamentales en la distribución de la marginación a nivel municipal entre las entidades seleccionadas. Entidad Posición de la Curva (Tendencia Central) Forma de la Curva (Dispersión/Heterogeneidad)

Conclusión Clave

- Nuevo León Muy hacia la izquierda (valores bajos, ≈ 0 o negativos).
- Alta y Estrecha (Baja dispersión). La mayoría de sus municipios presentan una marginación muy baja y homogénea, con el núcleo de municipios altamente industrializados dominando la distribución.

- Jalisco Hacia la izquierda (valores bajos). Ancha/Asimétrica (Dispersión Media-Alta). La distribución muestra heterogeneidad. Coexisten municipios de muy baja marginación (Zona Metropolitana de Guadalajara) y un grupo considerable de municipios rurales con marginación media o alta.
- Durango Cerca del centro (valores cercanos al promedio nacional). Ancha (Dispersión Media). Muestra una distribución más equilibrada alrededor del promedio nacional. Su curva es más plana, indicando una mezcla continua de municipios de baja y alta marginación.
- Veracruz Hacia la derecha (valores altos). Ancha y Plana (Alta dispersión). Presenta una alta dispersión de la marginación, reflejando el fuerte contraste entre su franja costera y sus extensas regiones rurales. La mayoría de sus municipios se encuentran en la zona de marginación media-alta.
- Chiapas Significativamente hacia la derecha (valores muy altos). Alta y Estrecha (sesgada a la derecha). Sus municipios tienen la marginación más alta y concentrada. El pico de la curva está lejos a la derecha, indicando que la mayor parte de la población municipal se ubica en los niveles más severos de marginación.

c) Población Total Índice de Marginación

```
variable_pob <- "POB_TOT"
marginacion_base[[variable_pob]] <- as.numeric(marginacion_base[[variable_pob]])

# Creamos un data frame vacío para almacenar los resultados
resumen_poblacion <- data.frame(
  Entidad = character(),
  Poblacion_Total_Municipal = numeric(),
  stringsAsFactors = FALSE
)

for (estado in entidades) {
  # Seleccionamos la POB_TOT para el estado actual
  poblacion_estado <- subset(marginacion_base, NOM_ENT == estado)[[variable_pob]]

  # Calculamos la suma
  suma_poblacion <- sum(poblacion_estado)

  # Creamos una nueva fila
  nueva_fila <- data.frame(
    Entidad = estado,
    Poblacion_Total_Municipal = suma_poblacion,
    stringsAsFactors = FALSE
  )

  # Añadimos la nueva fila al data frame de resultados usando rbind (base R)
  resumen_poblacion <- rbind(resumen_poblacion, nueva_fila)
}

resumen_poblacion <- resumen_poblacion[order(resumen_poblacion$Entidad), ]

#resumen_poblacion$Poblacion_Total_Municipal_Formato <-
# format(resumen_poblacion$Poblacion_Total_Municipal, big.mark = ",", scientific = FALSE)

resumen_poblacion$Poblacion_Total <- as.numeric(resumen_poblacion$Poblacion_Total_Municipal)

knitr::kable(resumen_poblacion[, c("Entidad", "Poblacion_Total")],
```

```

digits = 3,
booktabs = TRUE,
longtable = FALSE)

```

	Entidad	Poblacion_Total
1	Aguascalientes	1425607
2	Baja California	3769020
3	Baja California Sur	798447
4	Campeche	928363
7	Chiapas	5543828
8	Chihuahua	3741869
9	Ciudad de México	9209944
5	Coahuila de Zaragoza	3146771
6	Colima	731391
10	Durango	1832650
11	Guanajuato	6166934
12	Guerrero	3540685
13	Hidalgo	3082841
14	Jalisco	8348151
15	México	16992418
16	Michoacán de Ocampo	4748846
17	Morelos	1971520
18	Nayarit	1235456
19	Nuevo León	5784442
20	Oaxaca	4132148
21	Puebla	6583278
22	Querétaro	2368467
23	Quintana Roo	1857985
24	San Luis Potosí	2822255
25	Sinaloa	3026943
26	Sonora	2944840
27	Tabasco	2402598
28	Tamaulipas	3527735
29	Tlaxcala	1342977
30	Veracruz de Ignacio de la Llave	8062579
31	Yucatán	2320898
32	Zacatecas	1622138

```

#poblacion_nacional_calculada <- sum(resumen_poblacion$Poblacion_Total_Municipal)
#print(paste0("Total Nacional Calculado (Suma de las 32 Entidades): ",
#             format(poblacion_nacional_calculada, big.mark = ",", scientific = FALSE)))

```

d) Población Total Índice de Marginación

```

entidades_boxplot <- c("Chiapas", "Guerrero", "Nuevo León", "Tamaulipas", "Durango")

# --- PASO 1: Preparación de Datos ---

# 2. Filtrar los datos para incluir solo las 5 entidades
datos_boxplot <- subset(marginacion_base, NOM_ENT %in% entidades_boxplot)

```

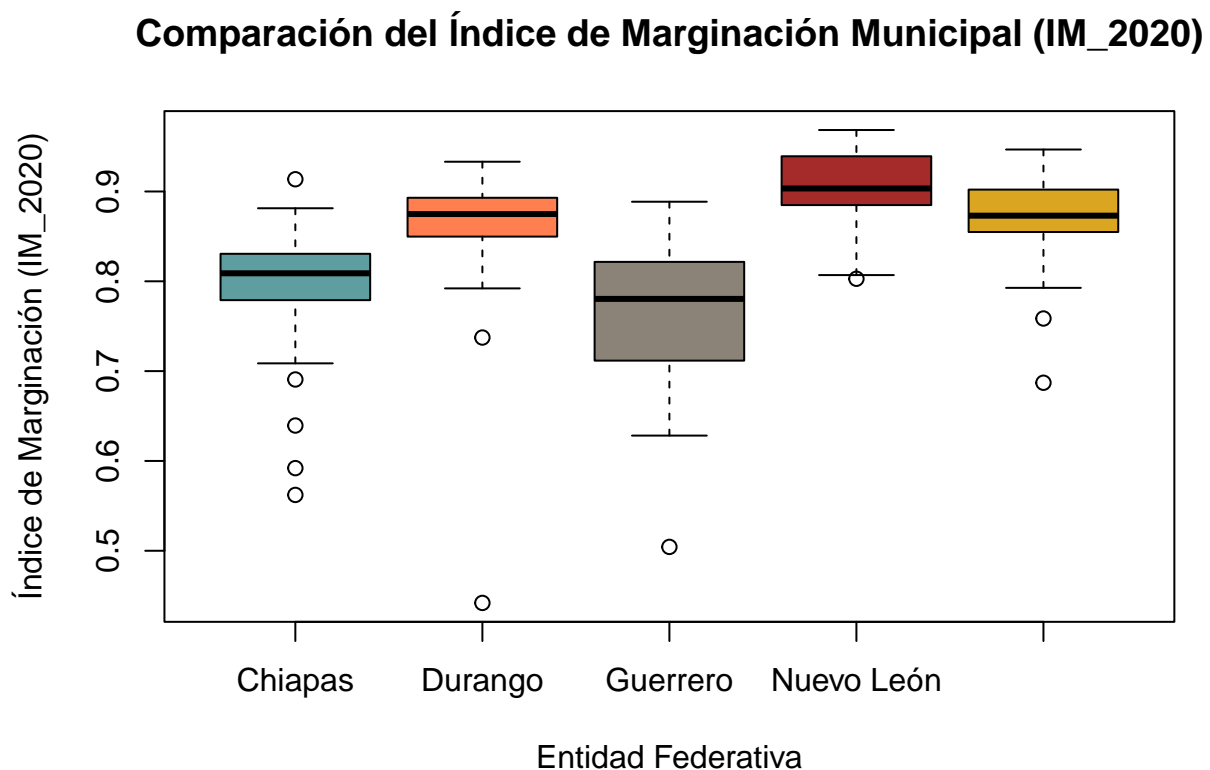
```
# 3. Eliminar filas con NA en el índice de marginación
datos_boxplot <- na.omit(datos_boxplot[, c("NOM_ENT", variable_imn)])

# --- PASO 2: Generación del Boxplot ---

print("Generando Gráfico de Boxplots para el Índice de Marginación (IM_2020)...")

## [1] "Generando Gráfico de Boxplots para el Índice de Marginación (IM_2020)..."

# 1. Generar el boxplot
# Fórmula: Variable ~ Grupo (IM_2020 por NOM_ENT)
boxplot(IMN_2020 ~ NOM_ENT,
        data = datos_boxplot,
        main = "Comparación del Índice de Marginación Municipal (IM_2020)",
        xlab = "Entidad Federativa",
        ylab = "Índice de Marginación (IM_2020)",
        col = colores, # Asignación de colores
        las = 2, # Rotar etiquetas del eje X verticalmente para mejor lectura
        outline = TRUE) # Mostrar outliers (puntos fuera de los bigotes)
```



```
print("Gráfico de Boxplots generado.")
```

```
## [1] "Gráfico de Boxplots generado."
```

Entidad Posición de la Caja (Mediana) Tamaño de la Caja y Bigotes (Dispersión) Valores Atípicos (Outliers)
 Conclusión Principal Chiapas y Guerrero Más alta. El 50% central está en los valores más altos del índice.

Compacta/Alta: Esto indica que sus municipios son homogéneamente marginados. La gran mayoría se agrupa en índices muy altos. Pocos o nulos hacia la baja. Alta Marginación Concentrada. Son los estados con la peor situación generalizada a nivel municipal. Nuevo León Más baja. La mediana se encuentra en los valores más bajos. Muy compacta y baja. La dispersión es mínima y se concentra en valores muy positivos (baja marginación). Posibles atípicos hacia la alta marginación (municipios rurales aislados). Baja Marginación Concentrada. Sus municipios son, en su mayoría, los menos marginados.

Tamaulipas y Durango Posición intermedia (más cerca de la zona media del gráfico). Más amplias. La caja y los bigotes se extienden significativamente. Probables atípicos en ambas direcciones (baja y alta marginación). Heterogeneidad. Estos estados muestran la mayor desigualdad interna. Coexisten municipios de baja marginación (ciudades principales) y municipios con problemas severos (zonas rurales o fronterizas).

Distribuciones de Probabilidad

1. Ejercicio concurso de TV

El 30% de un determinado pueblo ve un concurso que hay en televisión. Desde el concurso se llama por teléfono a 10 personas del pueblo elegidas al azar. Determinar la probabilidad que, de las 10 personas elegidas:

Datos:

Número de eventos $n = 10$

Probabilidad de éxito $p = 0.30$

- a) Más de ocho personas estén viendo el programa

$$P(X = 9) + P(X = 10)$$

```
prob_mas_de_ocho <- dbinom(x = 9, size = 10, prob = 0.30) +  
                    dbinom(x = 10, size = 10, prob = 0.30)  
prob_mas_de_ocho
```

```
## [1] 0.0001436859
```

$$P(X > 8) = 0.0001436859$$

- b) Al menos una persona de las diez esté viendo el programa

$$P(X \geq 1) = 1 - P(X < 1)$$

```
prob_al_menos_1 <- 1 - pbinom(q = 0, size = 10, prob = 0.30)  
prob_al_menos_1
```

```
## [1] 0.9717525
```

$$P(X \geq 1) = 0.9717525$$