



# Tarea 1

## Análisis de datos de la población - Sector Educación

**Erick Cervantes Mendieta**

Matrícula: 2032430

*Modelos Probabilistas Aplicados*

Profesora: Dra. Satu Elisa Schaeffer

Doctorado en Ingeniería de Sistemas

Septiembre 2020

## 1. Presentación de los datos

Para esta tarea se tomó una base de datos del INEGI [2], la cual presenta un indicador de eficiencia del sistema educativo nacional: Tasa neta de matriculación por entidad federativa según nivel educativo durante el ciclo escolar 2018-2019. Dicha base de datos se editó en un archivo .dat, para manipular la información en el entorno R (Versión 4.0.2), dichos datos se muestran en la Figura 1, en donde, cada renglón representa un estado de la República Mexicana y cada uno de los datos es la proporción de la matrícula total de un nivel educativo determinado (Preescolar, Primaria, Secundaria, Medio Superior), respecto a la población en edad oficial de cursar el nivel.

Nótese que una mayor tasa bruta de escolarización para un nivel educativo se interpreta como una mayor asistencia a la escuela por parte de la población en las edades reglamentarias.

## Tarea 1

---

	Preescolar	Primaria	Secundaria	MediaSuperior
1	71.8	98.7	84.0	63.6
2	69.3	94.2	80.0	62.2
3	61.4	100.2	85.9	68.9
4	69.1	98.7	85.4	72.7
5	72.2	90.6	77.0	59.5
6	70.2	93.7	80.7	64.9
7	59.7	90.7	78.7	59.1
8	77.9	103.6	75.7	59.1
9	63.6	97.5	79.8	60.8
10	85.5	112.9	107.0	92.2
11	66.3	96.8	79.4	64.8
12	72.0	97.9	82.4	61.2
13	83.5	96.5	78.8	56.6
14	72.6	100.2	90.4	73.0
15	71.3	97.8	80.9	55.0
16	66.7	102.4	88.8	62.2
17	75.6	95.2	76.2	52.1
18	66.8	96.7	85.4	65.1
19	64.6	91.6	79.6	63.7
20	74.1	98.2	83.7	55.9
21	69.1	99.8	83.2	54.8
22	81.6	99.6	84.1	68.6
23	74.9	102.8	85.3	63.8
24	66.1	103.0	84.6	63.6
25	84.2	98.6	84.9	61.5
26	68.4	93.6	80.6	70.3
27	61.7	92.7	81.9	65.2
28	86.6	98.5	91.0	70.6
29	62.1	94.2	79.1	63.0
30	75.6	97.4	87.6	65.9
31	60.6	90.4	77.8	60.4
32	80.9	94.9	80.5	65.4
33	77.9	98.3	83.7	61.1

Figura 1: Base de datos de la tasa neta de matriculación

## 2. Análisis de los datos

Una forma de analizar la base de datos es mediante el uso de herramientas Estadísticas, por lo que calculamos un resumen estadístico (Figura 2). En dicho resumen se puede observar que el nivel Primaria presenta el promedio más alto de matriculación en el país, y que los porcentajes más dispersos entre los estados de la República Mexicana parecieran presentarse en el nivel Medio Superior.

Preescolar	Primaria	Secundaria	MediaSuperior
Min. :59.70	Min. : 90.40	Min. : 75.70	Min. :52.10
1st Qu.:66.30	1st Qu.: 94.20	1st Qu.: 79.60	1st Qu.:60.40
Median :71.30	Median : 97.80	Median : 82.40	Median :63.60
Mean :71.63	Mean : 97.51	Mean : 83.15	Mean :63.84
3rd Qu.:75.60	3rd Qu.: 99.60	3rd Qu.: 85.30	3rd Qu.:65.40
Max. :86.60	Max. :112.90	Max. :107.00	Max. :92.20

Figura 2: Resumen estadístico de la base de datos

Se observa, además, que hay valores mayores al 100 %, esto se debe a que el indicador es sensible a la migración de la población [2], ya que algunos alumnos se inscriben como nuevo ingreso en entidades federativas vecinas.

### 3. Diagrama de Caja y Bigotes

Una ayuda visual para examinar el resumen estadístico presentado anteriormente es mediante el uso de un diagrama de Caja y Bigotes, ya que nos permite visualizar fácilmente el comportamiento de la mediana, la media y los cuartiles cuando se tienen diferentes conjuntos de datos.

En la Figura 3 podemos observar que durante el ciclo escolar 2018-2019 se presentó una mayor asistencia a la escuela en el nivel Primaria, mientras que el porcentaje de menor asistencia se encuentra en el nivel Medio Superior. Se observa una mayor variabilidad en los porcentajes del nivel Preescolar en comparación con los demás niveles. Algo curioso sucede con los datos del nivel Medio Superior, y sería interesante ver qué estados producen los datos atípicos, así, por ejemplo, la Secretaría de Educación Pública podría ver qué medidas tomó el estado con el valor más alto para poder implementar en los demás estados en ese nivel educativo.

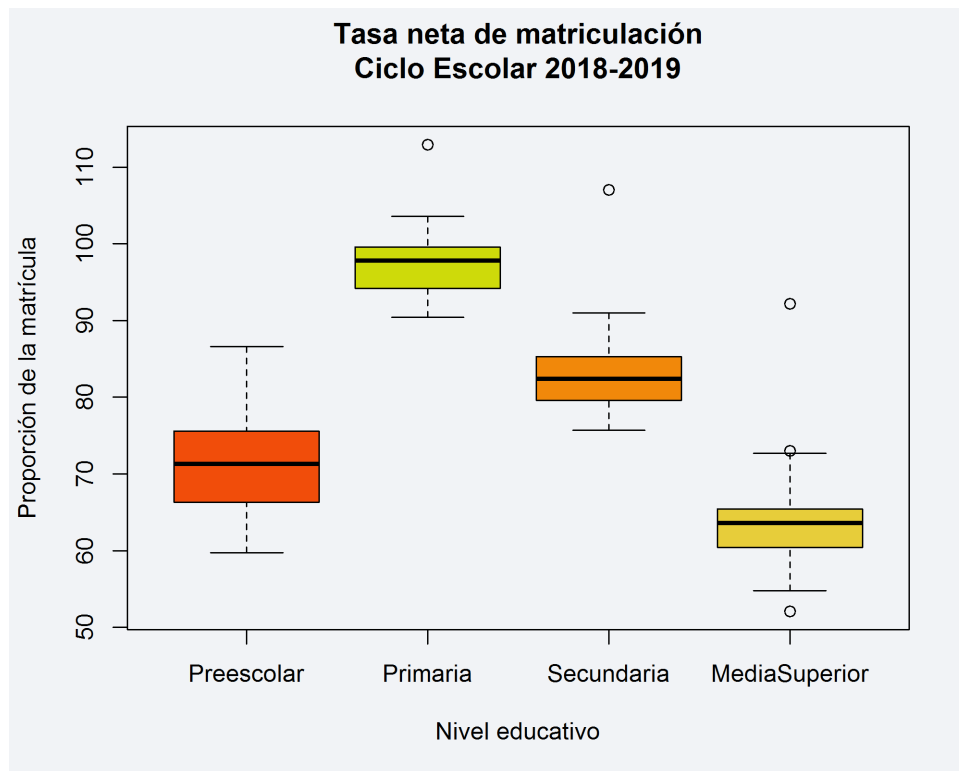


Figura 3: Diagrama de Caja de la base de datos

## Referencias

- [1] R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- [2] Instituto Nacional de Estadística y Geografía. Características educativas de la población. Recuperado el 5 de septiembre de 2020 de [https://www.inegi.org.mx/app/tabulados/interactivos/?px=Educacion\\_09&bd=Educacion](https://www.inegi.org.mx/app/tabulados/interactivos/?px=Educacion_09&bd=Educacion)