

Probabilidad y Estadística

Solución Lab en R

Integrantes:

Emanuel Arturo Rivas Escobar 2127706-3743

Sebastian Quintero Ramirez 2126439-3743

John Alejandro Regino Solis 202128397-3743

Preprocesamiento de datos

c. Reglas aplicadas:

Consistency rules

numerical rules

Tasa.natalidad > 0

Tasa.mortalidad < 50

categorical rules

GRUPOS %in% c("africa", "asia", "iberoamérica", "EuropaOriental", "ORIENTE
MÉDIO", "EO-NA_JAPON_AUSTR_NZ")

Teniendo en cuenta las reglas aplicadas sobre la base de datos de excel, se generaron los siguientes resultados:

```
Console Terminal Background Jobs
R 4.2.2 C:/Users/EMANUEL ARTURO RIVAS/Desktop/UNIVERSIDAD TALLERES Y TRABAJOS/CUARTO SEMESTRE/PROBABILIDAD Y ESTADISTI
> # Verificación de las reglas sobre los datos
> editrules::violatedEdits(Rules, paísesdatos)
edit
record num1 num2 dat1
1 FALSE FALSE FALSE
2 FALSE FALSE FALSE
3 FALSE FALSE FALSE
4 FALSE FALSE FALSE
5 FALSE FALSE FALSE
6 FALSE FALSE FALSE
7 FALSE FALSE FALSE
8 FALSE FALSE FALSE
9 FALSE FALSE FALSE
10 FALSE FALSE FALSE
11 FALSE FALSE FALSE
12 FALSE FALSE FALSE
13 FALSE FALSE FALSE
14 FALSE FALSE FALSE
15 FALSE FALSE FALSE
16 FALSE FALSE FALSE
17 FALSE FALSE FALSE
18 FALSE FALSE FALSE
19 FALSE FALSE FALSE
20 FALSE FALSE FALSE
21 FALSE FALSE FALSE
22 FALSE FALSE FALSE
23 FALSE FALSE FALSE
24 FALSE FALSE FALSE
25 FALSE FALSE FALSE
26 FALSE FALSE FALSE
27 FALSE FALSE FALSE
28 FALSE FALSE FALSE
29 FALSE FALSE FALSE
30 FALSE FALSE FALSE
```

```
96 FALSE FALSE FALSE
97 FALSE FALSE FALSE
> valid_data = editrules::violatedEdits(Rules, paísesdatos)
> summary(valid_data)
No violations detected, 0 checks evaluated to NA
NULL
> |
```

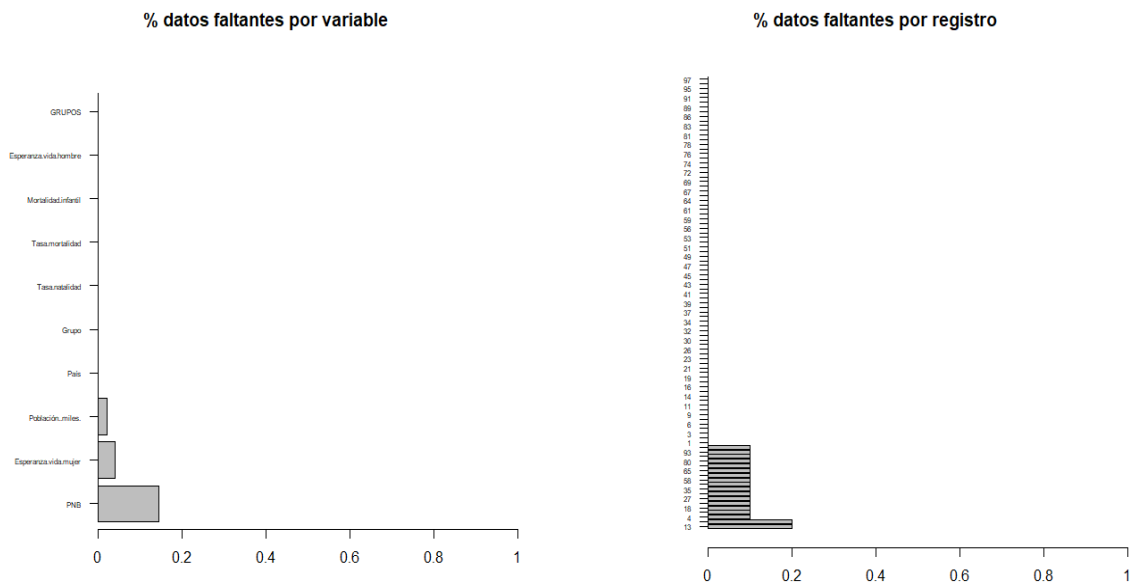
Como se aprecia en los resultados, los registros dentro de la base de datos no violan las reglas propuestas, por lo tanto tenemos datos que son consistentes, sin embargo aún hay datos faltantes, se procederá a mostrar un reporte de los datos faltantes.

d.



Del presente gráfico se puede observar que se encuentran datos faltantes (NA) dentro de las columnas Esperanza.vida.mujer(4.12%), PNB(14,43%) y la Población.miles(2.06%).

Otro reporte teniendo en cuenta las variables y registros:



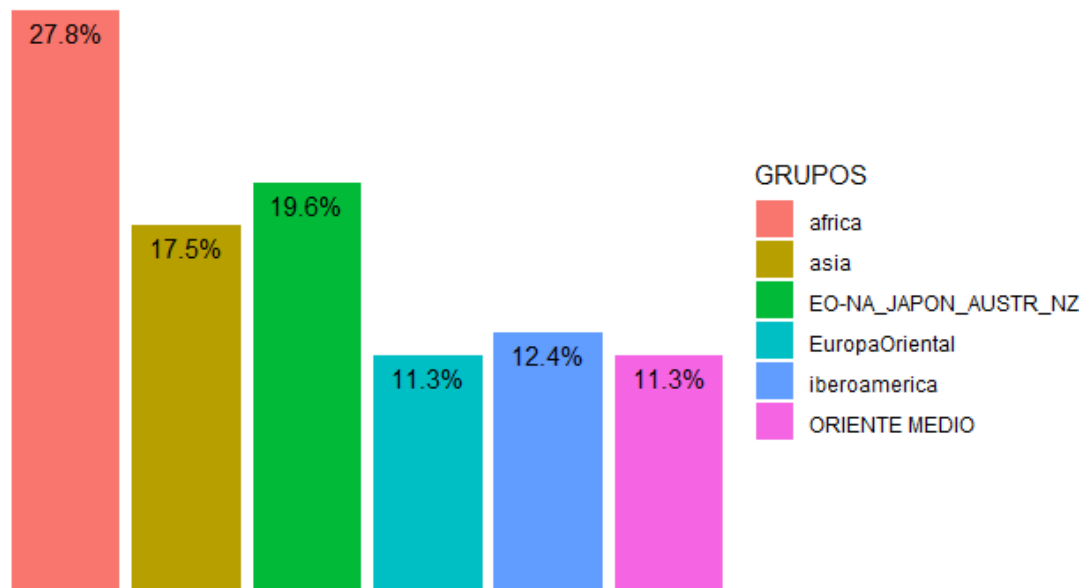
Para corregir los datos faltantes, se utilizará uno de los métodos de imputación propuestos en clase, en éste caso se utilizará el de “Imputación por la Media”. Los siguientes datos serán los que se van a imputar en los datos faltantes NA en su correspondiente columna.

```
> mean(Esperanza.vida.mujer,na.rm =T)
[1] 65.90538
> mean(PNB,na.rm = T)
[1] 5838
> mean(Población..miles.,na.rm = T)
[1] 51123.11
summary(mis_datos)
```



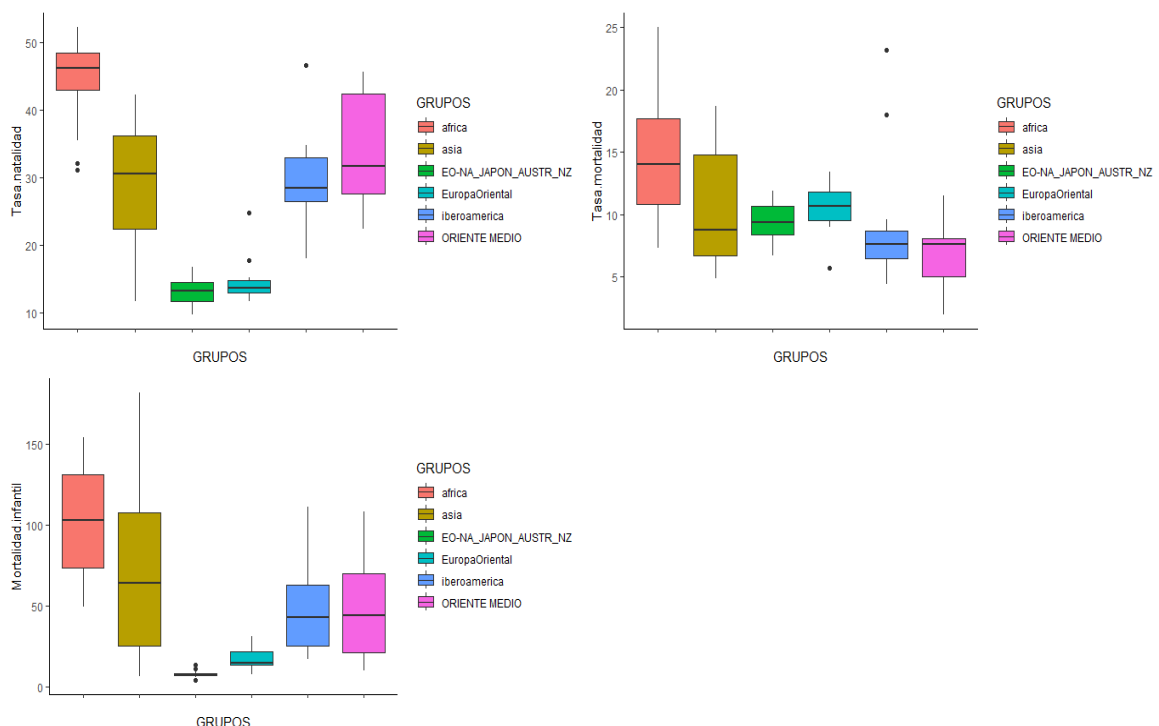
Visualización de Datos - Actividad

1. Al tratarse de una gráfica con varios datos de observación lo primero al analizar es el cómo deberíamos simplificar esos datos para así construir la tabla más eficiente posible. (En este caso hablamos de grupos conformados por países de los distintos continentes, como áfrica, asia, etc). Aquí lo más adecuado sería el manejar gráficas de tipo diagrama de barra o tablas de frecuencia dado que estos nos dan las cifras lo suficientemente exactas.



Teniendo en cuenta el anterior diagrama de barras, se puede observar que se concentra una mayor cantidad de países pertenecientes a África.

2

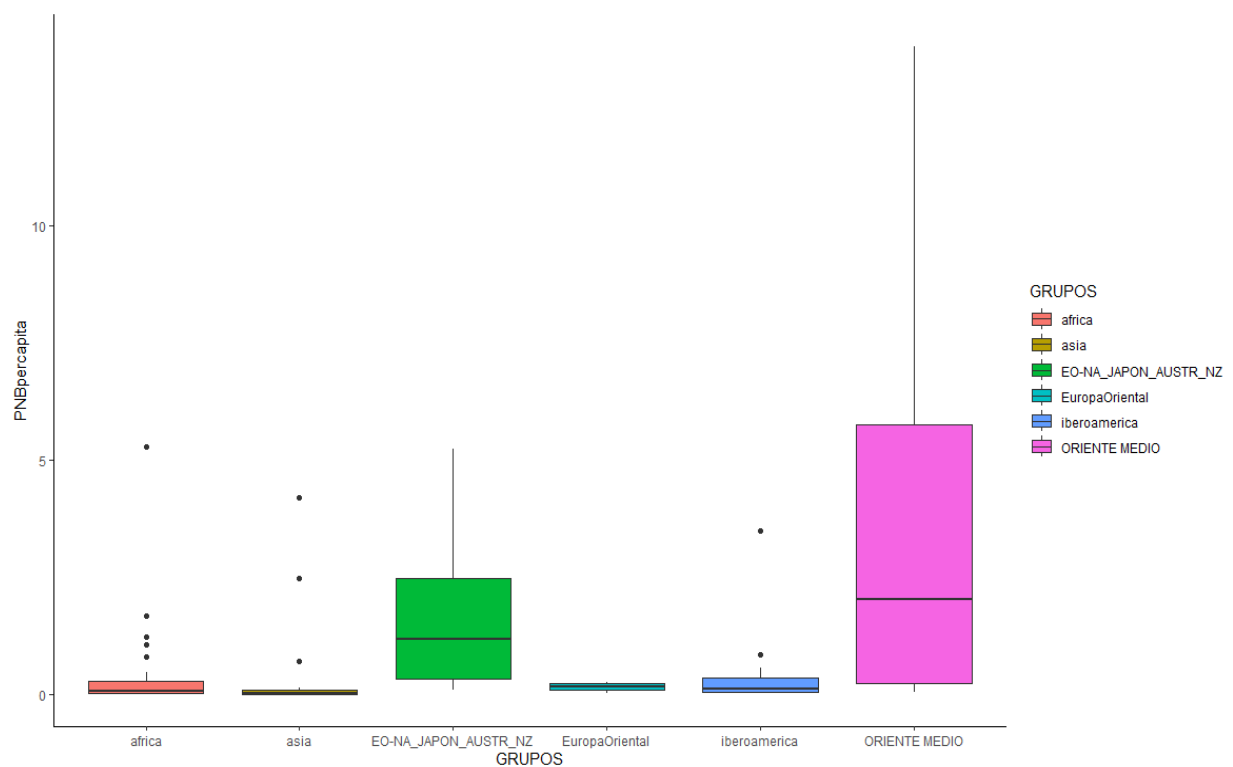


A partir del anterior gráfico, las principales diferencias que se concluyen es que África predomina 3 categorías, una gran tasa de natalidad pero a su vez una gran tasa de mortalidad, esto determina que en dicha zona no hay una calidad de vida estable.

Otra principal diferencia es que se presenta una mayor cantidad de tasa de natalidad en África y en el Oriente Medio.

Comparando la tasa de natalidad y mortalidad de Oriente Medio, se concluye que puede que tengan sobrepoblación, debido a que nacen muchos niños pero no se presentan muchos fallecidos durante un tiempo determinado.

3.



A partir del anterior gráfico, se concluye que la zona geográfica con más PNB per cápita entre los grupos, es el de Oriente Medio: esto quiere decir que en promedio el Oriente Medio tiene un nivel de riqueza entre cada uno de sus habitantes durante un determinado año, además de su productividad económica, mayor que los otros grupos.