

Creación, selección y acoplamiento de diversas variables

Joel Alejandro Zavala Prieto

Contents

Informacion de contacto	2
Descripción del análisis	3
Factores	4
Selección de variables	4
Agregando nuevas variables	6
Distribuciones de los datos	7

Informacion de contacto

mail: alejandro.zavala1001@gmail.com

Facebook: <https://www.facebook.com/AlejandroZavala1001>

```
## Loading required package: zoo
```

```
##
```

```
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      as.Date, as.Date.numeric
```

Descripción del análisis

El siguiente análisis consiste en interactuar con el entorno y lógica de diversas funciones en R. Se simulara un dataframe con diversas variables para el análisis y descripción del mismo

Viendo las primeras 20 observaciones del dataframe

	edad	sexo	ingresos	raza	auto	casa
1	29	Femenino	2365	blanco	Si	No
2	32	Masculino	1630	negro	No	No
3	11	Femenino	1037	negro	No	Si
4	24	Femenino	9757	blanco	No	No
5	21	Masculino	6961	moreno	No	No
6	32	Femenino	6043	negro	No	Si
7	19	Femenino	5508	negro	No	Si
8	16	Femenino	4058	moreno	No	No
9	10	Femenino	2673	moreno	No	No
10	27	Masculino	1199	moreno	No	No
11	12	Masculino	7133	blanco	No	No
12	12	Masculino	2127	blanco	Si	Si
13	15	Femenino	9549	moreno	Si	No
14	31	Masculino	2769	negro	Si	Si
15	29	Femenino	6477	blanco	No	Si
16	20	Femenino	7648	moreno	No	No
17	17	Masculino	1922	moreno	No	Si
18	13	Masculino	4652	negro	No	Si
19	23	Masculino	3266	blanco	No	Si
20	18	Femenino	5518	moreno	Si	No

Viendo las ultimas 20 observaciones del dataframe

	edad	sexo	ingresos	raza	auto	casa
9981	14	Femenino	5494	blanco	No	Si
9982	15	Masculino	1843	blanco	Si	Si
9983	31	Femenino	6441	negro	No	Si
9984	23	Masculino	3002	negro	No	No
9985	25	Femenino	9549	moreno	Si	Si
9986	19	Femenino	5440	blanco	No	No
9987	20	Masculino	876	negro	Si	No
9988	10	Masculino	6222	moreno	No	No
9989	13	Femenino	6801	blanco	No	No
9990	11	Masculino	1034	moreno	No	Si
9991	18	Masculino	6351	negro	Si	Si
9992	29	Masculino	3524	negro	No	Si
9993	15	Femenino	734	blanco	Si	No
9994	32	Femenino	4830	moreno	No	No
9995	22	Femenino	8518	negro	No	Si
9996	10	Masculino	471	blanco	No	No
9997	14	Masculino	8483	moreno	Si	Si
9998	32	Masculino	5209	blanco	Si	No
9999	35	Femenino	4399	moreno	No	Si
10000	30	Masculino	9303	blanco	No	No

Viendo el nombre de columna del dataframe

```
## [1] "edad"      "sexo"      "ingresos" "raza"      "auto"      "casa"
```

Donde:

1. Edad: representa la edad del individuo
2. Sexo: representa el sexo del individuo
3. Ingresos: Ingresos mensuales del individuo mensual
4. Raza: raza del individuo
5. Auto: si el individuo tiene auto propio
6. casa: si el individuo tiene casa propio

Factores

Para conocer la estructura del dataframe, usamos la función **str**

```
str(datos_ingresos)
```

```
## 'data.frame':    10000 obs. of  6 variables:
## $ edad      : int  29 32 11 24 21 32 19 16 10 27 ...
## $ sexo      : chr  "Femenino" "Masculino" "Femenino" "Femenino" ...
## $ ingresos: int  2365 1630 1037 9757 6961 6043 5508 4058 2673 1199 ...
## $ raza      : chr  "blanco" "negro" "negro" "blanco" ...
## $ auto      : chr  "Si" "No" "No" "No" ...
## $ casa      : chr  "No" "No" "Si" "No" ...
```

Selección de variables

Si deseamos mostrar las primeras 10 filas de nuestro dataframe

edad	sexo	ingresos	raza	auto	casa
29	Femenino	2365	blanco	Si	No
32	Masculino	1630	negro	No	No
11	Femenino	1037	negro	No	Si
24	Femenino	9757	blanco	No	No
21	Masculino	6961	moreno	No	No
32	Femenino	6043	negro	No	Si
19	Femenino	5508	negro	No	Si
16	Femenino	4058	moreno	No	No
10	Femenino	2673	moreno	No	No
27	Masculino	1199	moreno	No	No

Si deseamos ver las primeras 10 observaciones (renglones) de las primeras 2 columnas

	edad	sexo
1	29	Femenino
2	32	Masculino
3	11	Femenino
4	24	Femenino
5	21	Masculino
6	32	Femenino
7	19	Femenino
8	16	Femenino
9	10	Femenino
10	27	Masculino

Si deseamos ver los elementos únicos de una columna, tomando la de sexo

sexo
Femenino
Masculino

Si deseamos filtrar nuestro dataframe por individuos que sean del sexo “Femenino”. Mostraremos las primeras 15 observaciones

	edad	sexo	ingresos	raza	auto	casa
1	29	Femenino	2365	blanco	Si	No
3	11	Femenino	1037	negro	No	Si
4	24	Femenino	9757	blanco	No	No
6	32	Femenino	6043	negro	No	Si
7	19	Femenino	5508	negro	No	Si
8	16	Femenino	4058	moreno	No	No
9	10	Femenino	2673	moreno	No	No
13	15	Femenino	9549	moreno	Si	No
15	29	Femenino	6477	blanco	No	Si
16	20	Femenino	7648	moreno	No	No
20	18	Femenino	5518	moreno	Si	No
22	21	Femenino	384	blanco	No	Si
23	16	Femenino	147	moreno	Si	No
24	17	Femenino	7550	moreno	No	Si
25	28	Femenino	5461	negro	Si	Si

Si queremos extraer a todos los individuos de sexo “Masculino” y que sean de raza “Moreno”. Mostrando las primeras 10 observaciones

	edad	sexo	ingresos	raza	auto	casa
5	21	Masculino	6961	moreno	No	No
10	27	Masculino	1199	moreno	No	No
17	17	Masculino	1922	moreno	No	Si
27	16	Masculino	9803	moreno	Si	No
28	11	Masculino	8250	moreno	Si	Si
30	35	Masculino	8884	moreno	Si	No
41	30	Masculino	4643	moreno	No	No
44	14	Masculino	2497	moreno	Si	No
45	15	Masculino	9300	moreno	Si	Si
62	27	Masculino	3162	moreno	No	No

Si deseamos filtrar por individuos de sexo femenino y con ingresos mayores iguales a 1000 dolares. Mostrando las primeras 10 observaciones

	edad	sexo	ingresos	raza	auto	casa
1	29	Femenino	2365	blanco	Si	No
3	11	Femenino	1037	negro	No	Si
4	24	Femenino	9757	blanco	No	No
6	32	Femenino	6043	negro	No	Si
7	19	Femenino	5508	negro	No	Si
8	16	Femenino	4058	moreno	No	No
9	10	Femenino	2673	moreno	No	No
13	15	Femenino	9549	moreno	Si	No
15	29	Femenino	6477	blanco	No	Si
16	20	Femenino	7648	moreno	No	No

Agregando nuevas variables

Si al dataframe inicial queremos convertir la cantidad a pesos mexicanos. Listando las primeras 10 observaciones

	edad	sexo	ingresos	raza	auto	casa	ingresos_MXN
1	29	Femenino	2365	blanco	Si	No	47300
2	32	Masculino	1630	negro	No	No	32600
3	11	Femenino	1037	negro	No	Si	20740
4	24	Femenino	9757	blanco	No	No	195140
5	21	Masculino	6961	moreno	No	No	139220
6	32	Femenino	6043	negro	No	Si	120860
7	19	Femenino	5508	negro	No	Si	110160
8	16	Femenino	4058	moreno	No	No	81160
9	10	Femenino	2673	moreno	No	No	53460
10	27	Masculino	1199	moreno	No	No	23980

Distribuciones de los datos

Veamos la distribucion de los ingresos de todo nuestro dataframe

