

CAPÍTULO 3

ESTADÍSTICA DESCRIPTIVA

En este capítulo se verán todas las técnicas que se usan para la organización y presentación de datos en tablas y gráficas, así como el cálculo de medidas estadísticas. Se considerarán solamente datos univariados y bivariados.

Ejemplo 3.1 Los siguientes datos provienen de un cuestionario de 10 preguntas que se hizo a 28 estudiantes de una clase de Estadística Aplicada I en el Recinto Universitario de Mayaguez de la Universidad de Puerto Rico. Un asterisco (*) significa que la pregunta no fue contestada. En lo sucesivo se hará referencia a este conjunto de datos como "<http://math.uprm.edu/~edgar/clase97.dat>"

Row	edad	sexo	escuela	programa	creditos	gpa	familia	hestud	htv
1	21	f	públ	biol	119	3.60	3	35	10
2	18	f	priv	mbio	15	3.60	3	30	10
3	19	f	priv	biot	73	3.61	5	5	7
4	20	f	priv	mbio	*	2.38	3	14	3
5	21	m	públ	pmed	114	3.15	2	25	25
6	20	m	públ	mbio	93	3.17	3	17	6
7	22	m	públ	pmed	120	2.15	5	20	10
8	20	m	priv	pmed	*	3.86	5	15	5
9	20	m	priv	pmed	94	3.19	4	10	2
10	20	f	públ	pmed	130	3.66	6	20	33
11	21	f	priv	mbio	97	3.35	1	15	20
12	20	m	priv	mbio	64	3.17	4	30	2
13	20	f	públ	mbio	*	3.23	2	5	3
14	21	f	públ	mbio	98	3.36	4	15	10
15	21	f	priv	biol	113	2.88	5	15	3
16	21	f	priv	pmed	124	2.80	5	20	10
17	20	f	públ	eagr	*	2.50	4	10	5
18	20	f	priv	mbio	*	3.46	4	18	5
19	22	f	priv	pmed	120	2.74	2	10	15
20	20	f	priv	mbio	95	3.07	3	15	12
21	22	f	priv	biol	125	2.20	3	20	10
22	23	m	públ	eagr	13	2.39	3	10	8
23	21	m	priv	pmed	118	3.05	4	10	10
24	20	f	públ	mbio	118	3.55	5	38	10
25	21	f	públ	mbio	106	3.03	5	36	35
26	20	f	priv	mbio	108	3.61	3	20	10
27	22	f	públ	mbio	130	2.73	5	15	2
28	21	f	priv	pmed	128	3.54	3	18	5

Las variables **edad**, **familia**, **hestud** y **htv** son consideradas como variables cuantitativas discretas. Las variables **créditos** y **gpa** son consideradas como variables cuantitativas continuas. Las variables **sexo**, **escuela** y **programa** son consideradas como variables cualitativas

3.1 Organización de datos Cuantitativos Discretos

3.1.1 Tablas de Frecuencias

Los datos cuantitativos discretos se organizan en tablas, llamadas **Tablas de Distribución de frecuencias**. La primera columna de la tabla contiene los distintos valores que asume la variable ordenados de menor a mayor y las restantes columnas contienen los siguientes tipos de frecuencias.

Frecuencia absoluta: Indica el número de veces que se repite un valor de la variable.

Frecuencia relativa: Indica la proporción con que se repite un valor. Se obtiene dividiendo la frecuencia absoluta entre el tamaño de la muestra. Para una mejor interpretación es más conveniente mutiplicarla por 100 para trabajar con una **Frecuencia relativa porcentual**.

Frecuencia absoluta acumulada: Indica el número de valores que son menores o iguales que el valor dado.

Frecuencia relativa porcentual acumulada: Indica el porcentaje de datos que son menores o iguales que el valor dado.

Para construir una tabla de frecuencias en **R** se sigue la secuencia desde la consola.

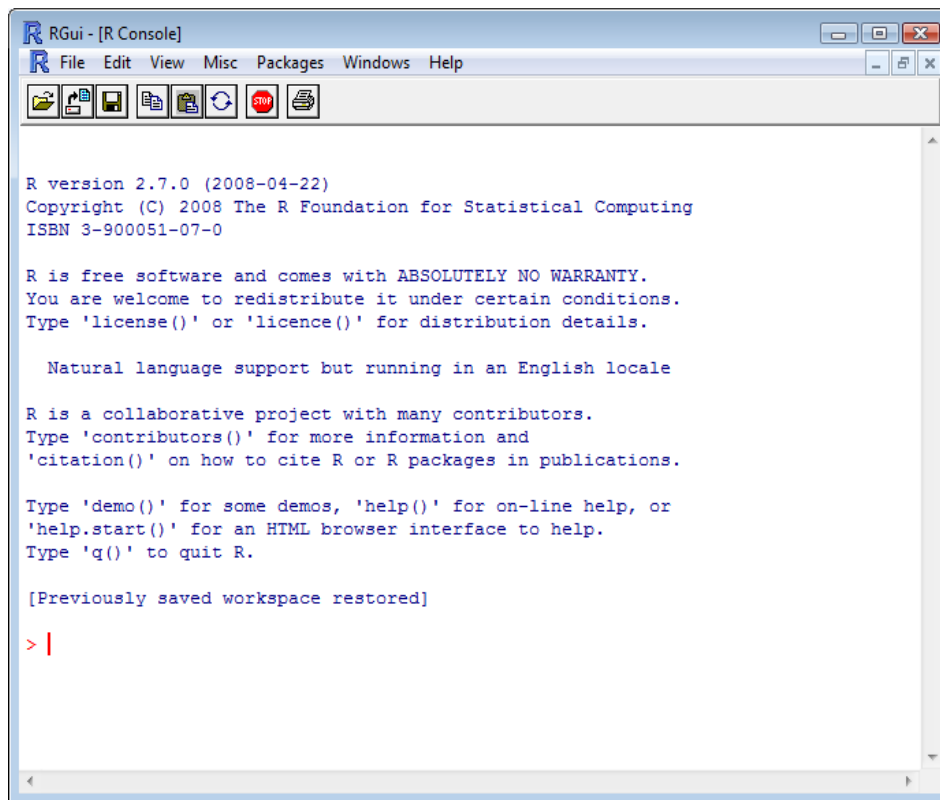


Figura 3.1. Consola de R Statistical Computing

```
> datos<-read.table("http://math.uprm.edu/~edgar/clase97.dat",header=T)
> attach(datos)
> mytable<-table(familia)
> mytable
familia
1 2 3 4 5 6
1 3 9 6 8 1
> (mytable)/margin.table(mytable)
familia
```

```

      1      2      3      4      5      6
0.03571429 0.10714286 0.32142857 0.21428571 0.28571429 0.03571429

> cumsum(mytable)
  1  2  3  4  5  6
  1  4 13 19 27 28

> cumsum(mytable)/margin.table(mytable)
      1      2      3      4      5      6
0.03571429 0.14285714 0.46428571 0.67857143 0.96428571 1.00000000

```

Explicación de los comandos.

```
> datos<-read.table("http://math.uprm.edu/~edgar/clase97.dat",header=T)
```

Lee el archivo en formato de tabla y genera un conjunto de datos de él. El subcomando header=T nos provee el nombre de cada columna.

```
> attach(datos)
```

La base de datos es adherida a la dirección donde R busca cuando va a evaluar una variable.

```
> mytable<-table(familia)
```

Crea una tabla de la variable familia. Nos provee la frecuencia absoluta. Por ejemplo el tamaño familiar que más predomina es 3.

```
> (mytable)/margin.table(mytable)
```

Representa la frecuencia relativa porcentual. Por ejemplo, sólo 3.57 por ciento de las familias de los estudiantes entrevistados son de tamaño 6.

```
> cumsum(mytable)
```

Representa la frecuencia absoluta acumulada. Por ejemplo 27 de los 28 entrevistados tienen una familia de tamaño menor o igual que 5.

```
> cumsum(mytable)/margin.table(mytable)
```

Representa la frecuencia relativa porcentual acumulada. Por ejemplo, el 94.93% de las familias son de tamaño menor o igual que 5.