

Estadística descriptiva con datos ordinales

Ramon Ceballos

29/1/2021

1. Descripción de datos ordinales

Los **datos ordinales** son parecidos a los cualitativos, en el sentido de que son cualidades de los individuos u objetos.

La diferencia existente entre los datos cualitativos y los ordinales reside en las características que expresan. En el caso de los ordinales, éstas tienen un orden natural que permite “acumular” observaciones.

Es decir, podemos contar, para cada nivel, cuantas veces hemos observado un dato menor o igual a este.

Un ejemplo frecuente de datos ordinales es el empleo de las **Escalas Likert**.

2. Frecuencias para datos ordinales

Al trabajar con datos ordinales, el orden de los niveles de los datos nos permite calcular no solo *frecuencias absolutas y relativas*, sino también **frecuencias acumuladas**.

Es decir, podemos contar cuantas veces hemos observado un dato menor o igual a este.

Ejemplo 1

Suponed que tenemos una muestra de 15 estudiantes de los cuales sabemos su nota en el examen de Estadística. Clasificamos todos estos resultados en Suspenso (*S*), Aprobado (*A*), Notable (*N*) y Excelente (*Ex*) y consideramos su orden natural $S < A < N < Ex$.

Las notas obtenidas han sido las siguientes

S, A, N, Ex, S, S, Ex, Ex, N, A, A, A, A, N, S

Como recordaréis, para saber cuantas hay de cada una (su frecuencia absoluta), utilizamos la función **table()**.

Para crear el array de datos ordinales se emplea la función **ordered()** en lugar de **factor()**.

```
#Creo el array de notas
#En vez de emplear la función factor(),
#se emplea la función ordered() para generar el array de datos ordinales
notas = ordered(c("S","A", "N", "Ex",
                  "S", "S", "Ex", "Ex",
                  "N", "A", "A",
                  "A","A", "N", "S"),
               levels = c("S", "A", "N", "Ex"))
```

```
#Genero la table contingencia para esta variable ordinal
table(notas)
```

```
## notas
##  S  A  N Ex
##  4  5  3  3
```

Como podréis observar, hay 4 *S*, 5 *A*, 3 *N* y 3 *Ex*.

En lo referente a **frecuencias absolutas acumuladas**, hay

- 4 estudiantes con *S* o menos. Ello implica que la frecuencia acumulada de *S* es 4
- 9 estudiantes que han obtenido *A* o menos. Entonces, la frecuencia acumulada de *A* es 9
- 12 estudiantes los cuales han obtenido *N* o menos. Así, la frecuencia acumulada de *N* es 12
- 15 estudiantes (todos) que han obtenido *Ex* o menos. De este modo, la frecuencia acumulada de *Ex* es 15, o sea, el total.

Frecuencia relativa acumulada. Es la fracción del total de las observaciones en tanto por 1 que representa su frecuencia absoluta acumulada

Así, las frecuencias relativas acumuladas respectivas son

- $S : \frac{4}{15} \approx 0.27$
- $A : \frac{9}{15} \approx 0.6$
- $N : \frac{12}{15} \approx 0.8$
- $Ex : \frac{15}{15} = 1$

En general, supongamos que realizamos n observaciones:

$$x_1, \dots, x_n$$

Estas observaciones son de un cierto tipo de datos ordinales, cuyos posibles niveles ordenados son:

$$l_1 < l_2 < \dots < l_k$$

Por tanto, cada una de las observaciones x_j es igual a algún l_i . Diremos que todas estas observaciones forman una variable ordinal. En nuestro ejemplo anterior, los 4 niveles eran:

$$S < A < N < Ex$$

Además, nuestro $n = 15$ y nuestros x_1, \dots, x_{15} son las calificaciones obtenidas por los alumnos.

De este modo, con estas notaciones

- Las definiciones de **frecuencias absolutas** n_j y las **relativas** f_j , para cada nivel l_j son las mismas que en una variable cualitativa.
- Las **frecuencia absoluta acumulada** del nivel l_j en esta variable ordinal es el número N_j de observaciones x_i tales que $x_i \leq l_j$. Es decir,

$$N_j = \sum_{i=1}^j n_i$$

- La **frecuencia relativa acumulada** del nivel l_j en esta variable ordinal es la fracción en tanto por 1 F_j de observaciones x_i tales que $x_i \leq l_j$. Es decir,

$$F_j = \frac{N_j}{n} = \sum_{i=1}^j f_i$$

Ejemplo 2

En un estudio, a un grupo de clientes de un restaurante, se les hizo la siguiente pregunta:

“¿Estás contento con el trato ofrecido por los trabajadores del establecimiento?”

Las posibles respuestas forman una escala ordinal con $1 < 2 < 3 < 4 < 5$. El valor 1 se corresponde con un trato malo por parte de los trabajadores mientras que el valor 5 refiere a un trato excelente.

Supongamos que se recogieron las siguientes respuestas de 50 técnicos:

```
#set.seed() permite que cada vez que se ejecuten las líneas salgan los mismo resultados  
#Se pone un nº facil de recordar  
set.seed(2018)
```

```
#sample() genera un array aleatorio  
clientes = sample(1:5, 50, replace = TRUE)  
clientes
```

```
## [1] 3 4 5 2 5 1 3 4 2 4 3 3 1 1 5 3 1 3 3 5 1 4 2 5 3 4 5 1 2 2 1 5 5 2 1 2 5 5  
## [39] 2 1 2 1 3 2 1 2 3 3 1 2
```

```
#Permite que la semilla aleatoria vuelva a partir de cero  
set.seed(NULL)
```

En este caso tenemos 5 niveles ($k = 5$) y 50 observaciones ($n = 50$) que forman una variable ordinal a la que hemos llamado `clientes`.

Hemos calculado todas sus frecuencias (absoluta, relativa, acumulada y relativa acumulada) y las hemos representado en la siguiente tabla.

```
#FREC. ABSOLUTA  
absolut = table(clientes)  
  
#FREC. RELATIVA ABSOLUTA  
relative = prop.table(absolut)  
  
#FREC. ABSOLUTA ACUMULADA  
acumul = cumsum(absolut)  
  
#FREC. RELATIVA ACUMULADA  
rel.acumul = cumsum(relative)  
  
#Guardo todas las frecuencias en una misma tabla  
  
absolut = (as.matrix(absolut))  
relative = (as.matrix(relative))  
acumul = (as.matrix(acumul))  
rel.acumul = (as.matrix(rel.acumul))  
  
clientela = data.frame(absolut,relative,acumul,rel.acumul)  
  
colnames(clientela) = c("Absoluta", "Relativa", "Acumulada", "Rel. Acumulada")  
  
clientela
```

| ## | Absoluta | Relativa | Acumulada | Rel. Acumulada |
|------|----------|----------|-----------|----------------|
| ## 1 | 12 | 0.24 | 12 | 0.24 |
| ## 2 | 12 | 0.24 | 24 | 0.48 |
| ## 3 | 11 | 0.22 | 35 | 0.70 |
| ## 4 | 5 | 0.10 | 40 | 0.80 |
| ## 5 | 10 | 0.20 | 50 | 1.00 |

Los gráficos para frecuencias absolutas y relativas absolutas de variables ordinales son exactamente los mismos que para las variables cualitativas.

También podemos utilizar diagramas de barras para describir frecuencias acumuladas: en este caso, la altura de cada barra debe ser igual a la frecuencia acumulada del nivel respectivo. Además, estos niveles deben de aparecer ordenados de manera ascendente, de forma que las alturas de las barras también tengan un orden ascendente.

No obstante, se recomienda no hacer uso de diagramas circulares a la hora de representar frecuencias acumuladas, debido a que éstos no representan la información sobre la acumulación de datos de forma fácil de entender a simple vista.

2.1. Frecuencias Acumuladas (absolutas y relativas)

Función `cumsum()`

¿Recordáis la función `cumsum()`? Pues esta puede ser utilizada a la hora de calcular frecuencias acumuladas. Para obtenerlas se hace:

- Frecuencias absolutas acumuladas. Se aplica `cumsum()` a la tabla de frecuencias absolutas.
- Frecuencias relativas globales acumuladas. Se aplica `cumsum()` a la tabla de frecuencias relativas globales

Retomemos el ejemplo anterior de las notas de los estudiantes y calculemos y representemos en un diagrama de barras las frecuencias acumuladas de la muestra de notas.

```
#Vector de estudio
notas
```

```
## [1] S A N Ex S S Ex Ex N A A A A N S
## Levels: S < A < N < Ex
```

```
#Calculo de Frec. abs.
fAbs = table(notas)
```

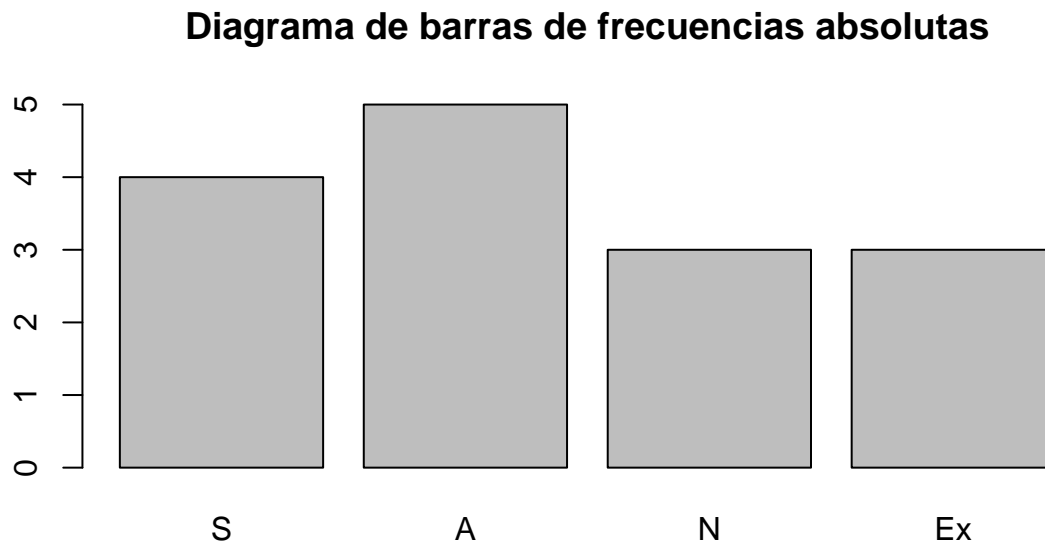
```
#Calculo de Frec. abs. acumuladas
cumsum(fAbs)
```

```
## S A N Ex
## 4 9 12 15
```

```
#calculo de Frec. relativas acumuladas
cumsum(prop.table(fAbs))
```

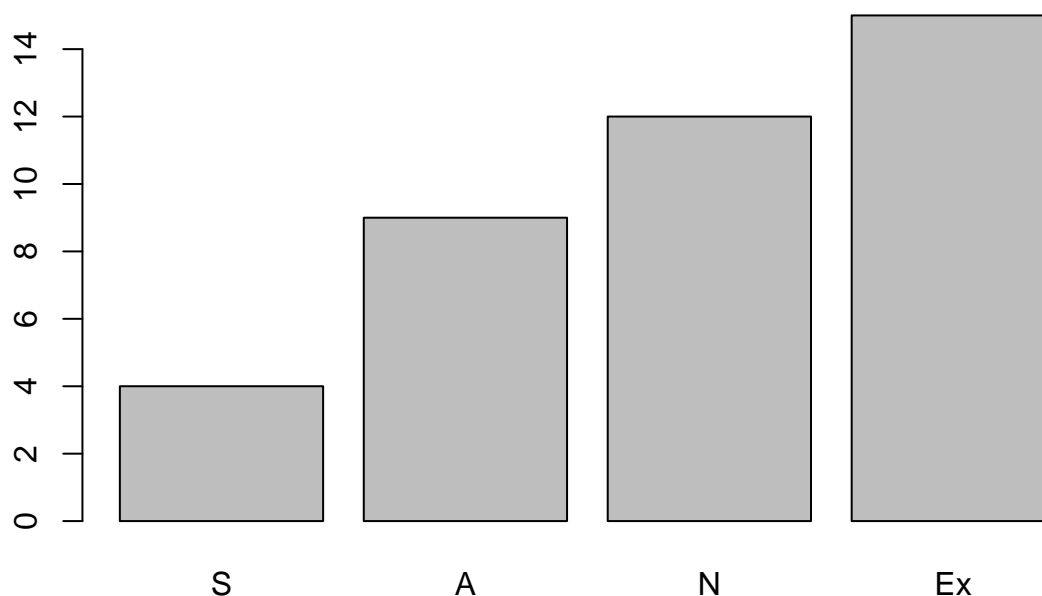
```
## S A N Ex
## 0.2666667 0.6000000 0.8000000 1.0000000
```

```
#Diagrama de barras para las frec. absolutas  
barplot(fAbs, main = "Diagrama de barras de frecuencias absolutas")
```



```
#Diagrama de barras para las frec. absolutas acumuladas  
barplot(cumsum(fAbs), main = "Diagrama de barras de frecuencias absolutas acumuladas")
```

Diagrama de barras de frecuencias absolutas acumuladas



Podríamos haber calculado las frecuencias relativas acumuladas de la siguiente forma.

```
#dividir las frecuencias absolutas acumuladas por la longitud del vector  
cumsum(table(notas))/length(notas)
```

```
##           S           A           N           Ex  
## 0.2666667 0.6000000 0.8000000 1.0000000
```

```
#hacer cumsum() de las frecuencias relativas globales  
cumsum(table(notas)/length(notas))
```

```
##           S           A           N           Ex  
## 0.2666667 0.6000000 0.8000000 1.0000000
```

Pero no podemos hacer `prop.table(cumsum(table(notas)))`. Da un valor incorrecto ya que toma la tabla de frecuencias absolutas acumuladas como una tabla de contingencia.

Ejemplo 3

Se ha evaluado el tamaño de los cuellos de 100 jirafas. Los niveles que se han utilizado se los considera ordenados de la siguiente manera:

Muy.corto < Corto < Normal < Largo < Muy.largo

Los valores obtenidos en dicho estudio han sido los siguientes:

```
set.seed(2018)

longitud = sample(1:5,100, replace = TRUE)

longitud = ordered(longitud)

levels(longitud) = c("Muy.corto","Corto","Normal","Largo","Muy.largo")
```

```
#vector a estudiar del cuello de las jirafas
longitud
```

```
##      [1] Normal      Largo      Muy.largo Corto      Muy.largo Muy.corto Normal
##      [8] Largo      Corto      Largo      Normal      Normal      Muy.corto Muy.corto
##     [15] Muy.largo Normal      Muy.corto Normal      Normal      Muy.largo Muy.corto
##     [22] Largo      Corto      Muy.largo Normal      Largo      Muy.largo Muy.corto
##     [29] Corto      Corto      Muy.corto Muy.largo Muy.largo Corto      Muy.corto
##     [36] Corto      Muy.largo Muy.largo Corto      Muy.corto Corto      Muy.corto
##     [43] Normal      Corto      Muy.corto Corto      Normal      Normal      Muy.corto
##     [50] Corto      Normal      Muy.corto Largo      Largo      Corto      Muy.corto
##     [57] Corto      Normal      Normal      Normal      Normal      Muy.corto Normal
##     [64] Muy.corto Corto      Largo      Muy.corto Corto      Muy.corto Muy.largo
##     [71] Muy.corto Corto      Muy.largo Largo      Muy.largo Normal      Corto
##     [78] Corto      Normal      Largo      Largo      Corto      Corto      Muy.largo
##     [85] Largo      Largo      Normal      Normal      Muy.corto Normal      Corto
##     [92] Normal      Muy.corto Corto      Muy.corto Normal      Corto      Corto
##     [99] Muy.corto Corto
## Levels: Muy.corto < Corto < Normal < Largo < Muy.largo
```

```
set.seed(NULL)
```

Estudiemos las frecuencias absolutas y relativas del vector del ejemplo.

```
#frec. abs.
Fr.Abs = table(longitud)
Fr.Abs
```

```
## longitud
## Muy.corto      Corto      Normal      Largo Muy.largo
##           23         26         24         13         14
```

```
#frec. rel. global
Fr.Rel = prop.table(Fr.Abs)
Fr.Rel
```

```
## longitud
## Muy.corto      Corto      Normal      Largo Muy.largo
##          0.23         0.26         0.24         0.13         0.14
```

```
#frec. abs. acumuladas
Fr.Acum = cumsum(Fr.Abs)
Fr.Acum
```

```
## Muy.corto    Corto    Normal    Largo Muy.largo
##           23      49      73      86      100
```

```
#frec. rel. acumuladas
Fr.RAcum = cumsum(Fr.Rel)
Fr.RAcum
```

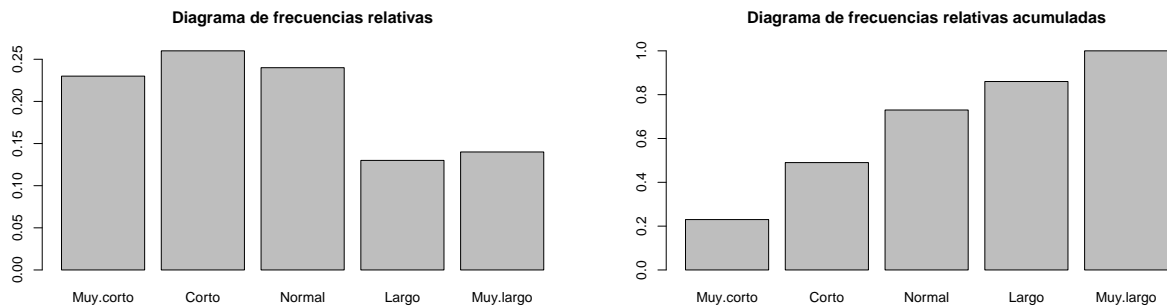
```
## Muy.corto    Corto    Normal    Largo Muy.largo
##           0.23    0.49    0.73    0.86    1.00
```

La instrucción `barplot` produce el siguiente diagrama de barras de frecuencias relativas acumuladas.

```
par(mfrow=c(1,2))

barplot(Fr.Rel, main = "Diagrama de frecuencias relativas")

barplot(Fr.RAcum, main = "Diagrama de frecuencias relativas acumuladas")
```



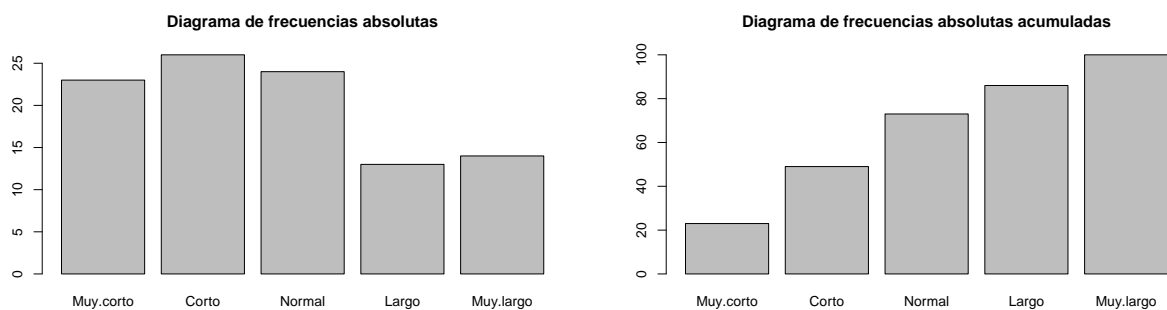
```
par(mfrow=c(1,1))
```

La instrucción `barplot` produce el siguiente diagrama de barras de frecuencias absolutas acumuladas.

```
par(mfrow=c(1,2))

barplot(Fr.Abs, main = "Diagrama de frecuencias absolutas")

barplot(Fr.Acum, main = "Diagrama de frecuencias absolutas acumuladas")
```




```
par(mfrow=c(1,1))
```