

Estadística discreta con pandas

Una pequeña empresa de seguros está realizando un estudio de los accidentes de tráfico que han tenido sus asegurados en el último mes, con intención de revisar el precio de las primas. Ha elegido aleatoriamente 40 asegurados y obtenido los siguientes datos

2, 1, 3, 1, 1, 0, 0, 3, 0, 1, 4, 2, 1, 0, 0, 0, 1, 1, 2, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 1, 3, 2

1. Completa la tabla de frecuencias
2. Representa la distribución utilizando un diagrama de barras
3. Representa la distribución utilizando un diagrama de sectores
4. Calcula las medidas de centralización: moda, mediana y media.
5. Calcula las medidas de dispersión: rango, desviación típica y coeficiente de variación.

Solución

1. x_i valores que toma la variable

f_i **frecuencia absoluta**: número de veces que aparece un determinado resultado de la variable estadística entre todos los datos que se han recogido

$F_i = \sum_{j \leq i} f_j$ **frecuencia absoluta acumulada**: suma de todas las frecuencias absolutas de ese resultado y todos lo que están antes que él.

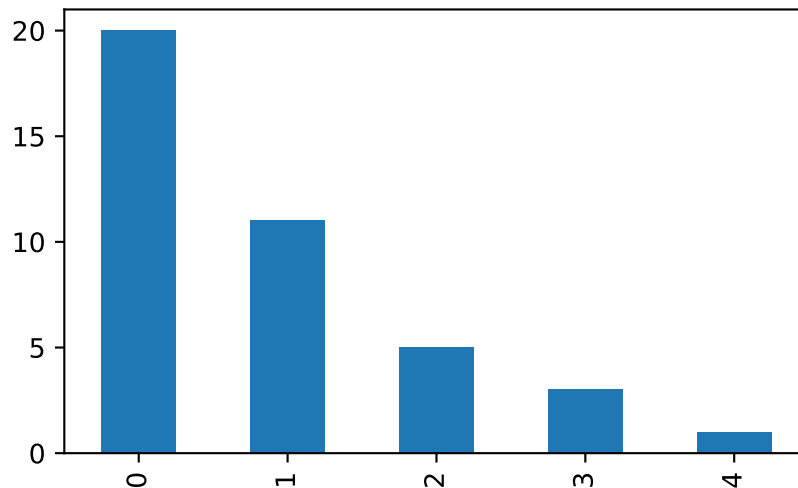
$h_i = \frac{f_i}{N}$ **frecuencia relativa**: cociente entre su frecuencia absoluta y el número total de datos.

$H_i = \frac{F_i}{N} = \sum_{j \leq i} h_j$ **frecuencia relativa acumulada**: suma de las frecuencia relativa del resultado y de todos los anteriores.

$\% = h_i * 100$ **porcentaje**

x_i	f_i	F_i	h_i	H_i	$\%$
0.0	20.0	20.0	0.500	0.500	50.0
1.0	11.0	31.0	0.275	0.775	27.5
2.0	5.0	36.0	0.125	0.900	12.5
3.0	3.0	39.0	0.075	0.975	7.5
4.0	1.0	40.0	0.025	1.000	2.5

2. Representa la distribución utilizando un diagrama de barras

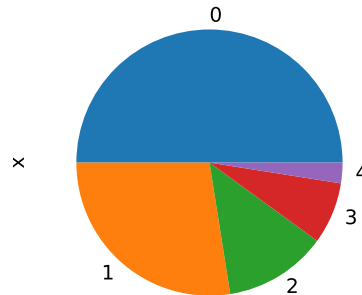


3. Representa la distribución utilizando un diagrama de sectores

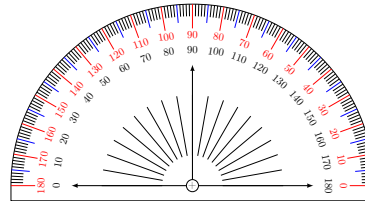
En primer lugar tendremos que obtener al ángulo de cada sector, teniendo en cuenta que $\text{ángulo} = h_i \cdot 360$. Podemos usar la tabla de frecuencias y añadir una última columna para la amplitud del ángulo de cada sector circular:

x_i	f_i	F_i	h_i	H_i	%	Ángulo
0.0	20.0	20.0	0.500	0.500	50.0	180.0
1.0	11.0	31.0	0.275	0.775	27.5	99.0
2.0	5.0	36.0	0.125	0.900	12.5	45.0
3.0	3.0	39.0	0.075	0.975	7.5	27.0
4.0	1.0	40.0	0.025	1.000	2.5	9.0

DIAGRAMA DE SECTORES



TRANSPORTADOR DE ÁNGULOS



4. Calcula las medidas de centralización: moda, mediana y media.

M_o **moda**: valor que más se repite entre los datos de que disponemos

M_e **mediana**: si suponemos que los datos están ordenados numéricamente de menor a mayor, es el valor que está en el centro, es decir, el valor que tiene por delante la mitad de los valores y por detrás la otra mitad.

media (aritmética): suma de todos los valores obtenidos y divididos entre el número total de datos.

$$\bar{x} = \frac{\sum f_i \cdot x_i}{N}$$

Para las dos primera nos basaremos en la tabla de frecuencias.

$M_o = 0$, $M_e = 0,5$ ya que $\frac{40}{2} = 20,0$ y tenemos que coger la media aritmética de los valores que están en la posición nº 20,0 y 21,0

x_i	f_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$
0	20	0	0
1	11	11	11
2	5	10	20
3	3	9	27
4	1	4	16
\sum	40	34	74

A partir de la tabla tenemos que: $\bar{x} = \frac{\sum f_i \cdot x_i}{N} = \frac{34}{40} = 0,85$

5. Calcula las medidas de dispersión: rango, desviación típica y coeficiente de variación.

rango: diferencia entre el mayor y el menor valor de la variable

varianza: media aritmética de los cuadrados de las diferencias entre los valores de la variable y la media.

$$s^2 = \sigma^2 = Var(X) = \frac{\sum f_i \cdot (x_i - \bar{x})^2}{N} = \frac{\sum f_i \cdot x_i^2}{N} - \bar{x}^2$$

desviación típica: raíz cuadrada de la varianza.

$$s = \sigma = \sqrt{Var(X)}$$

coeficiente de variación: es un número neutro, es decir, no está referido a ninguna medida. Por ello nos permite comparar datos correspondientes a elementos distintos.

$$cv = \frac{\sigma}{\bar{x}}$$

■ rango: $4 - 0 = 4$

■ $\sigma^2 = \frac{\sum f_i \cdot x_i^2}{N} - \bar{x}^2 = \frac{74}{40} - 0,85^2 = 1,128 \Rightarrow \sigma = \sqrt{1,128} = 1,062$

■ $cv(X) = \frac{\sigma}{\bar{x}} = \frac{1,062}{0,85} = 1,249$