

Representaciones Gráficas

Gráficos para variables cualitativas

Los gráficos más usuales para representar variables de tipo nominal son los siguientes:

Diagramas de barras:

Se representa en el eje de ordenadas las modalidades y en abscisas las frecuencias absolutas o las frecuencias relativas. Si se intentan comparar varias poblaciones entre sí, usando el diagrama, existen otras modalidades, como las mostradas en la figura 2. Cuando los tamaños de las dos poblaciones son diferentes, es conveniente utilizar las frecuencias relativas, ya que en otro caso podrían resultar engañosas.

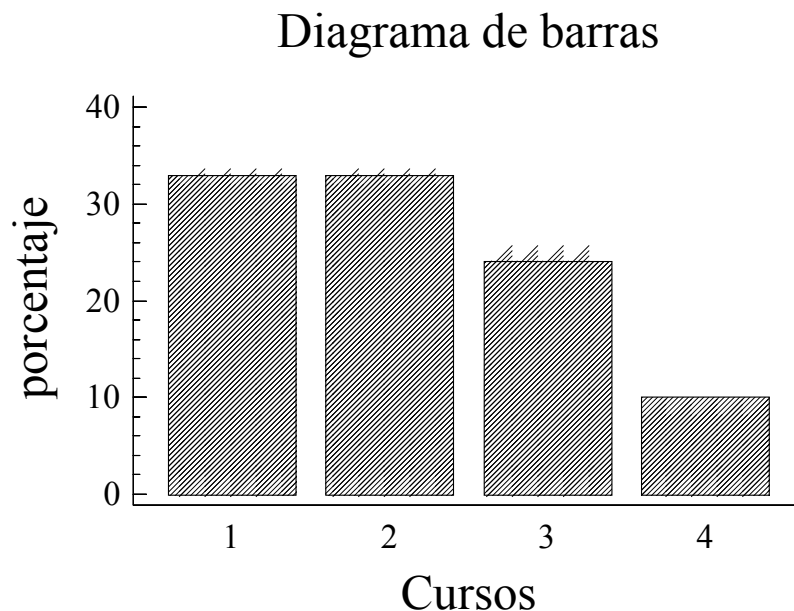


Figura 1. Diagrama de barras para una variable cualitativa

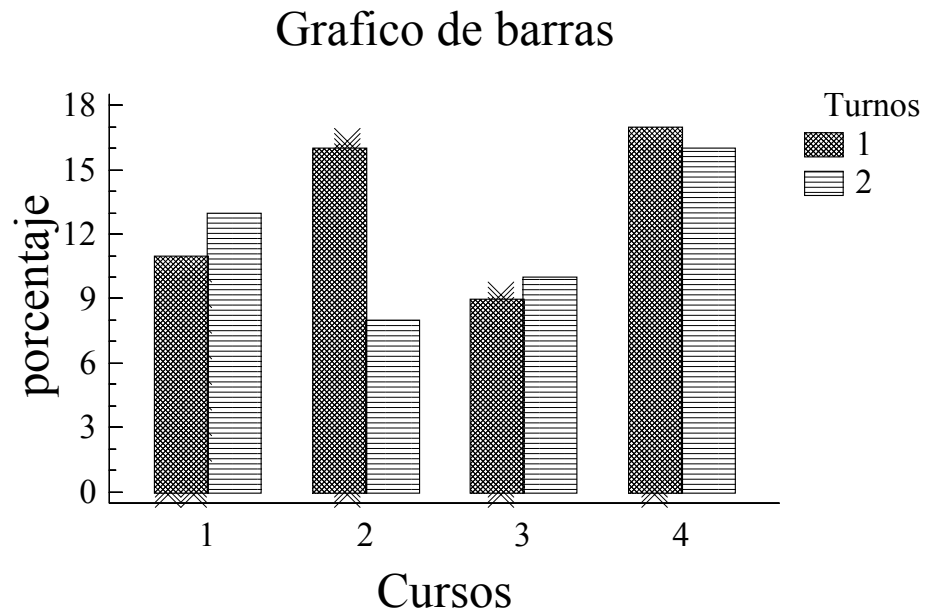


Figura 2. Diagrama de barras para una variable cualitativa

Diagramas de sectores (tartas)

Se divide un círculo en tantas porciones como clases existan, de modo que a cada clase le corresponde un arco de círculo proporcional a su frecuencia absoluta o relativa.

Grafico de sectores

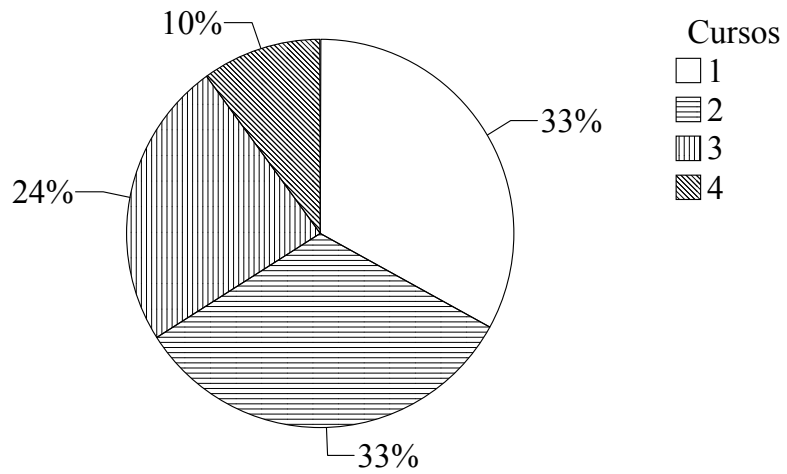


Figura 3. Diagrama de sectores

El arco de cada porción se calcula usando la siguiente *regla de tres*:

$$\begin{aligned} n &\longrightarrow 360^\circ \\ n_i &\longrightarrow x_i = \frac{360 \cdot n_i}{n} \end{aligned}$$

Como en la situación anterior, puede interesar comparar dos poblaciones. En este caso también es aconsejable el uso de las frecuencias relativas (porcentajes) de ambas variable sobre gráficos como los anteriores. Otra posibilidad es comparar las 2 poblaciones usando para cada una de ellas dos círculos concéntricos, al igual que en la figura 4.

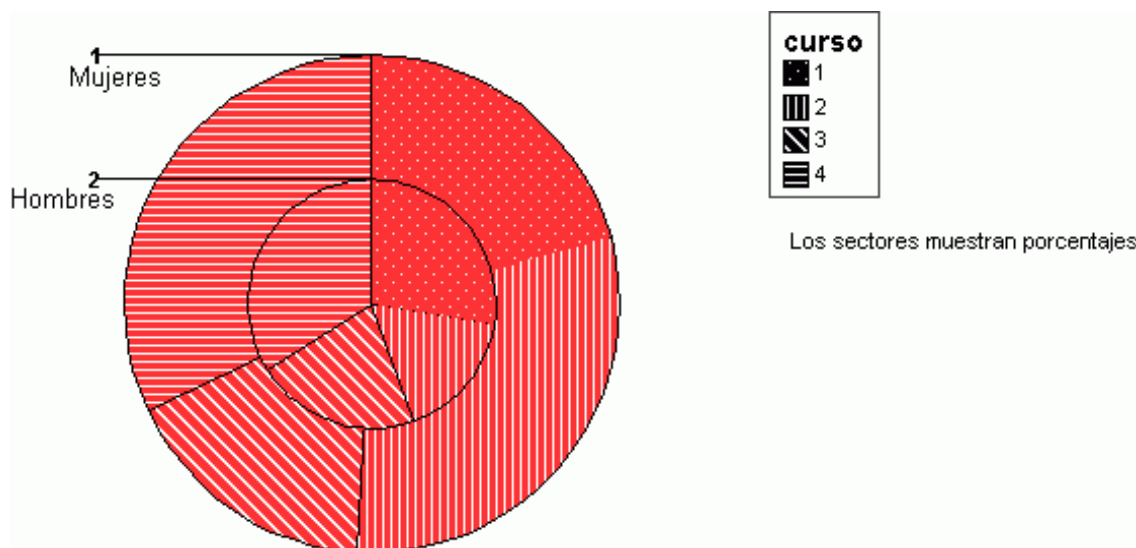


Figura 4.

Pictogramas

Expresan con dibujos alusivos al tema de estudio las frecuencias de las modalidades de la variable. Estos gráficos se hacen representado a diferentes escalas un mismo dibujo, como vemos en la figura siguiente:

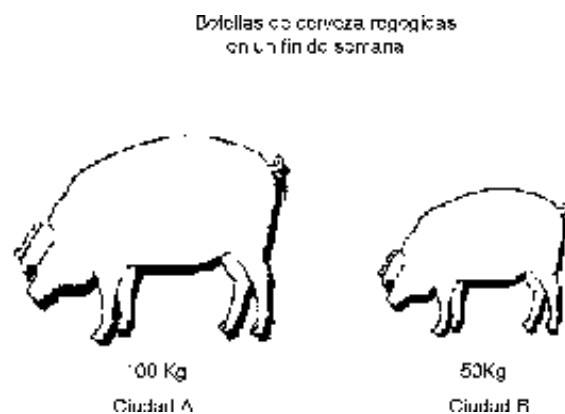


Figura 5.

El escalamiento de los dibujos debe ser tal que el área de cada uno de ellos sea proporcional a la frecuencia de la modalidad que representa. Este tipo de gráficos suele usarse en

los medios de comunicación, para que sean comprendidos por el público no especializado, sin que sea necesaria una explicación compleja.

Gráficos para variables cuantitativas

Para las variables cuantitativas, consideraremos dos tipos de gráficos, en función de que para realizarlos se usen las frecuencias (absolutas o relativas) o las frecuencias acumuladas:

Diagramas diferenciales: Son aquellos en los que se representan frecuencias absolutas o relativas. En ellos se representa el número o porcentaje de elementos que presenta una modalidad dada.

Diagramas integrales:

Son aquellos en los que se representan el número de elementos que presentan una modalidad inferior o igual a una dada. Se realizan a partir de las frecuencias acumuladas, lo que da lugar a gráficos crecientes. Obviamente, este tipo de gráficos no tiene sentido para variables cualitativas.

Según hemos visto existen dos tipos de variables cuantitativas: discretas y continuas. Vemos a continuación las diferentes representaciones gráficas que pueden realizarse para cada una de ellas así como los nombres específicos que reciben.

Gráficos para variables discretas

Cuando representamos una variable discreta, usamos el *diagrama de barras* cuando pretendemos hacer una gráfica diferencial. Las barras deben ser estrechas para representar que los valores que toma la variable son discretos. El diagrama integral o acumulado tiene, por la naturaleza de la variable, forma de escalera. Un ejemplo de diagrama de barras así como su diagrama integral correspondiente están representados en la figura 6.

Ejemplo

Se lanzan tres monedas al aire en 8 ocasiones y se contabiliza el número de caras, X , obteniéndose los siguientes resultados: $X \rightarrow 2, 1, 0, 1, 3, 2, 1, 2$. Representar gráficamente el resultado.

Solución: En primer lugar observamos que la variable X es cuantitativa discreta, presentando las modalidades: $X \in 0, 1, 2, 3$.

Ordenamos a continuación los datos en una tabla estadística, y se representa la misma en la figura 6.

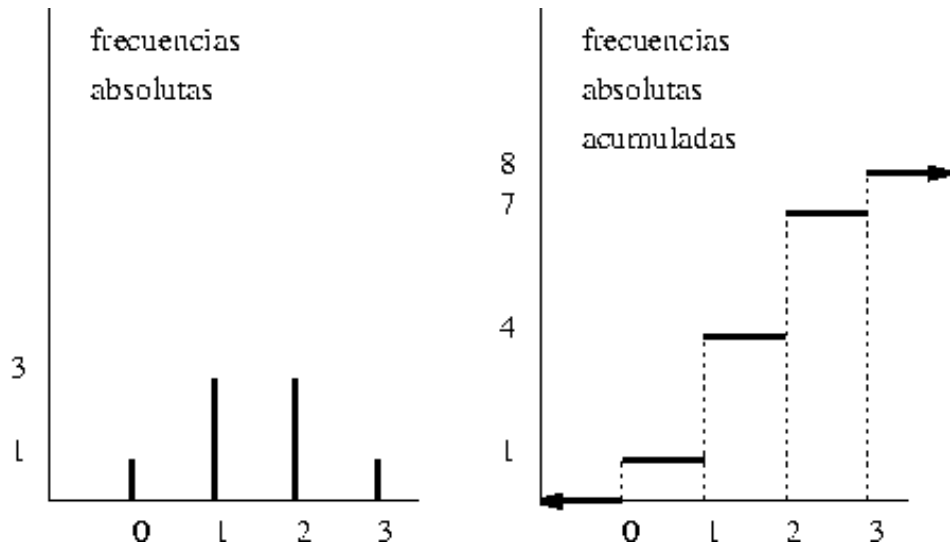


Figura 6. Diagrama diferencial (barras) e integral para una variable discreta. Obsérvese que el diagrama integral (creciente) contabiliza el número de observaciones de la variable inferiores o iguales a cada punto del eje de abscisas.

x_i	n_i	f_i	N_i	F_i
0	1	1/8	1	1/8
1	3	3/8	4	4/8
2	3	3/8	7	7/8
3	1	1/8	8	8/8
	$n = 8$	1		

Ejemplo

Clasificadas 12 familias por su número de hijos se obtuvo:

Número de hijos (x_i)	1	2	3	4
Frecuencias (n_i)	1	3	5	3

Comparar los diagramas de barras para frecuencias absolutas y relativas. Realizar el diagrama acumulativo creciente.

Solución: En primer lugar, escribimos la tabla de frecuencias en el modo habitual:

Variables	F. Absol.	F. Rel.	F. Acum.
x_i	n_i	f_i	N_i
1	1	0,083	1
2	3	0,250	4
3	5	0,416	9
4	3	0,250	12
	12	1	

Con las columnas relativas a x_i y n_i realizamos el diagrama de barras para frecuencias absolutas, lo que se muestra en la figura 7. Como puede verse es idéntico (salvo un cambio de escala en el eje de ordenadas) al diagrama de barras para frecuencias relativas y que ha sido calculado usando las columnas de x_i y f_i . El diagrama escalonado (acumulado) se ha construido con la información procedente de las columnas x_i y N_i .

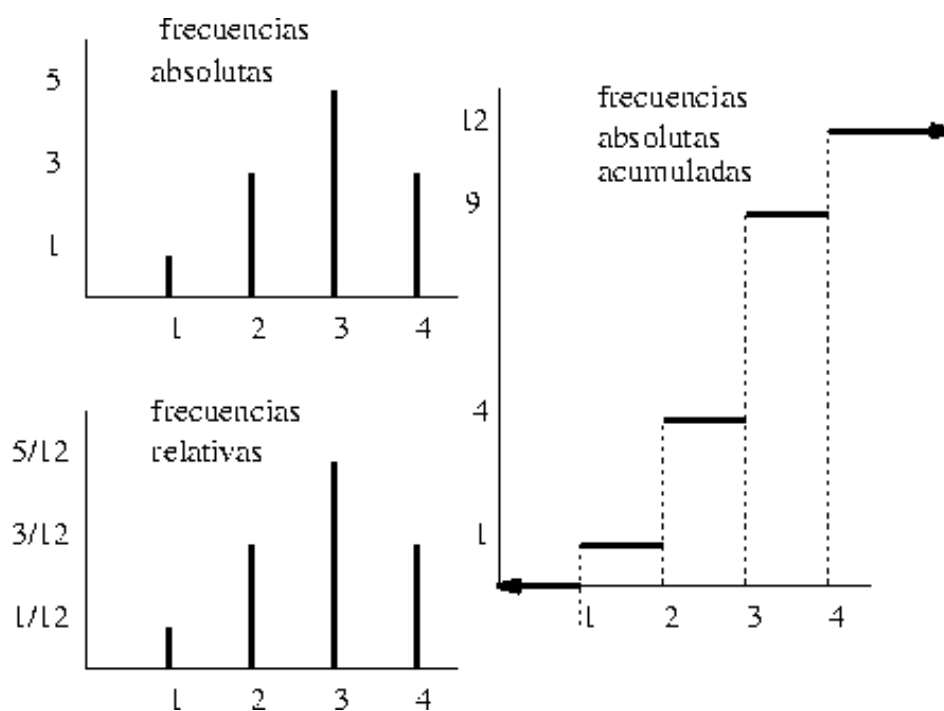


Figura 7. Diagramas de frecuencias para una variable discreta

Gráficos para variables continuas

Cuando las variables son continuas, utilizamos como diagramas diferenciales los histogramas y los polígonos de frecuencias.

Un histograma se construye a partir de la tabla estadística, representando sobre cada intervalo, un rectángulo que tiene a este segmento como base. El criterio para calcular la altura de cada rectángulo es el de mantener la proporcionalidad entre las frecuencias absolutas (o relativas) de cada intervalo y el área de los mismos.

El polígono de frecuencias se construye fácilmente si tenemos representado previamente el histograma, ya que consiste en unir mediante líneas rectas los puntos del histograma que corresponden a las marcas de clase. Para representar el polígono de frecuencias en el primer y último intervalo, suponemos que adyacentes a ellos existen otros intervalos de la misma amplitud y frecuencia nula, y se unen por una línea recta los puntos del histograma que corresponden a sus marcas de clase. Obsérvese que de este modo, el polígono de frecuencias tiene en común con el histograma el que las áreas de la gráficas sobre un intervalo son idénticas. Véanse ambas gráficas diferenciales representadas en la parte superior de la figura 8.

El diagrama integral para una variable continua se denomina también polígono de frecuencias acumulado, y se obtiene como la poligonal definida en abscisas a partir de los extremos de los intervalos en los que hemos organizado la tabla de la variable, y en ordenadas por alturas que son proporcionales a las frecuencias acumuladas. Dicho de otro modo, el polígono de frecuencias absolutas es una primitiva del histograma. Véase la parte inferior de la figura 8, en la que se representa a modo de ilustración los diagramas correspondientes a la variable cuantitativa continua expresada en la tabla siguiente:

Intervalos	c_i	n_i	N_i
0 - 2	1	2	2
2 - 4	3	1	3
4 - 6	5	4	7
6 - 8	7	3	10
8 - 10	9	2	12
		12	

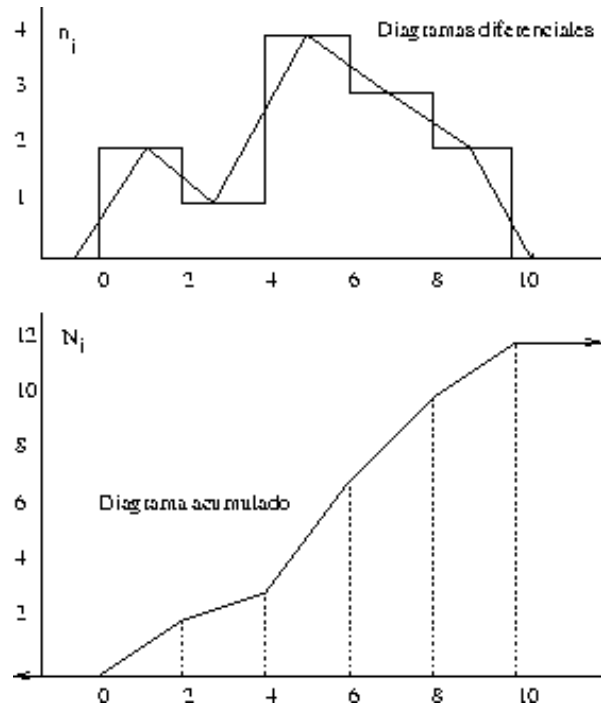


Figura 8. Diagramas diferenciales e integrales para una variable continua.

Ejemplo

La siguiente distribución se refiere a la duración en horas (completas) de un lote de 500 tubos:

Duración en horas	Número de tubos
300 – 500	50
500 – 700	150
700 – 1.100	275
más de 1.100	25
	Total 500

Representar el histograma de frecuencias relativas y el polígono de frecuencias. Trazar la curva de frecuencias relativas acumuladas. Determinar el número mínimo de tubos que tienen una duración inferior a 900 horas.

Solución: En primer lugar observamos que la variable en estudio es discreta (*horas completas*), pero al tener un rango tan amplio de valores resulta más conveniente agruparla

en intervalos, como si de una variable continua se tratase. La consecuencia es una ligera pérdida de precisión.

El último intervalo está abierto por el límite superior. Dado que en él hay 25 observaciones puede ser conveniente cerrarlo con una amplitud *razonable*. Todos los intervalos excepto el tercero tienen una amplitud de 200 horas, luego podríamos cerrar el último intervalo en 1300 horas.

Antes de realizar el histograma conviene observar que el histograma representa las frecuencias de los intervalos mediante áreas y no mediante alturas. Sin embargo nos es mucho más fácil hacer representaciones gráficas teniendo en cuenta estas últimas. Si todos los intervalos tienen la misma amplitud no es necesario diferenciar entre los conceptos de área y altura, pero en este caso el tercer intervalo tiene una amplitud doble a los demás, y por tanto hay que repartir su área en un rectángulo de base doble (lo que reduce su altura a la mitad).

Así será conveniente añadir a la habitual tabla de frecuencias una columna que represente a las amplitudes a_i de cada intervalo, y otra de frecuencias relativas rectificadas, f_i' , para representar la altura del histograma. Los gráficos requeridos se representan en las figuras 9 y 10.

Intervalos	a_i	n_i	f_i	f_i'	F_i
300 – 500	200	50	0,10	0,10	0,10
500 – 700	200	150	0,30	0,30	0,40
700 – 1.100	400	275	0,55	0,275	0,95
1.100 – 1.300	200	25	0,05	0,05	1,00
		$n = 500$			

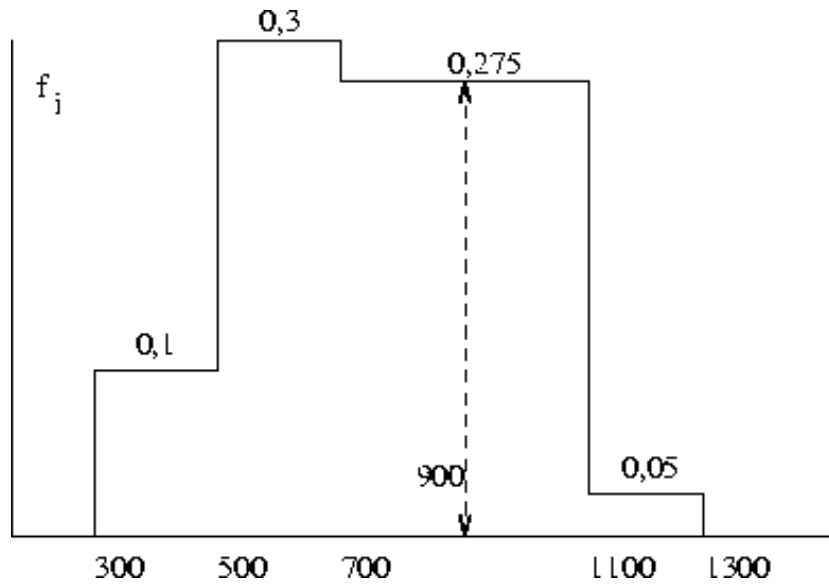


Figura 9. Histograma. Obsérvese que la altura del histograma en cada intervalo es f'_i que coincide en todos con f_i salvo en el intervalo 700 – 1.100 en el que $f'_i = 1/2f_i$ ya que la amplitud de ese intervalo es doble a la de los demás.

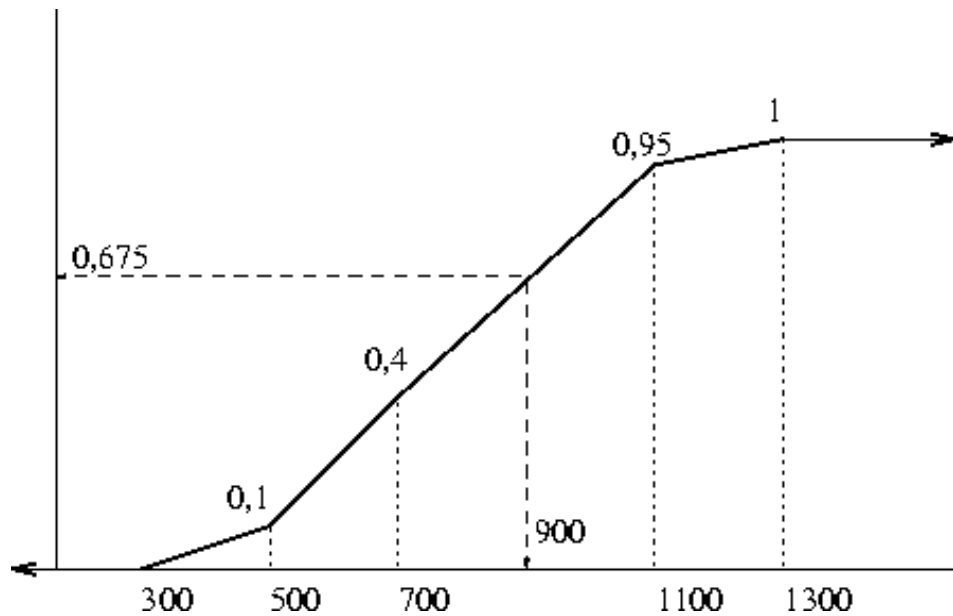


Figura 10. Diagrama acumulativo de frecuencias relativas

Por otro lado, mirando la figura 9 se ve que sumando frecuencias relativas, hasta las 900 horas de duración hay $0,10 + 0,30 + 0,275 = 0,675 = 67,5\%$ de los tubos.

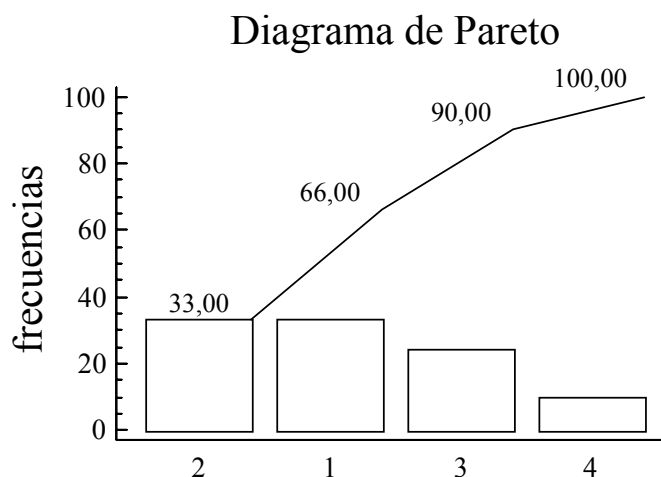
Esta cantidad se obtiene de modo más directo viendo a qué altura corresponde al valor 900 en el diagrama de frecuencias acumuladas (figura 10). Como en total son 500 tubos, el número de tubos con una duración igual o menor que 900 horas es $0,675 \times 500 = 337,5$, esto es, redondeando, 338 tubos.

Gráficos de Pareto

W. Pareto fue un economista italiano de siglo XIX que enunció el principio de la distribución de la riqueza, diciendo que el 80% de la misma está en manos del 20% de la población.

El diagrama de Pareto consiste en clasificar los factores que intervienen en un proceso por su orden de importancia, para poder tratar cada uno de ellos de forma adecuada a su peso específico. En realidad, no deja de ser un histograma que ordena las clases de mayor a menor frecuencia, junto a un polígono de frecuencias acumulado.

En STATGRAPHICS se obtiene en el menú Avanzado ->Control de Calidad ->Análisis de Pareto.



Pareto Chart with Cumulative Frequencies

Class			Weighted		Cum.		Cum.
Label	Rank	Count	Weight	Score	Score	Percent	Percent
2	1	33	1	33	33	33,00	33,00
1	2	33	1	33	66	33,00	66,00
3	3	24	1	24	90	24,00	90,00
4	4	10	1	10	100	10,00	100,00

Total		100		100			
-------	--	-----	--	-----	--	--	--

The StatAdvisor

This table shows the frequency of occurrence of the 4 unique values of x. The classes are sorted according to the counts, with the most frequently occurring class first. The highest class is 2 with a count of 33, which represents 33,0% of the total.

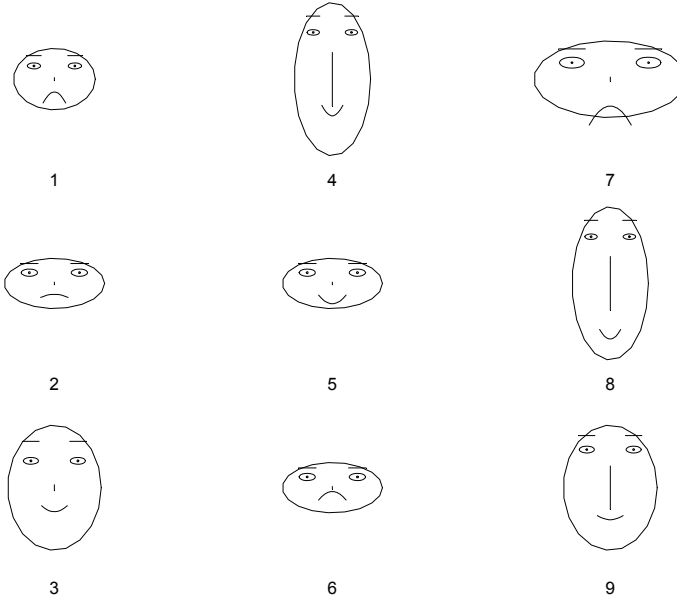
Caras de Chernoff y gráficos de estrellas (sun ray plots)

Caras de Chernoff: Este gráfico permite ilustrar tendencias en datos donde se miden varias variables al mismo tiempo. Se relacionan los datos con rasgos de una cara, de modo que se puede diferenciar entre observaciones.

Table 1: Description of facial features of Chernoff face

Dimension	Facial Feature
1	Face width
2	Ear level
3	Half face height
4	Eccentricity of upper ellipse of face
5	Eccentricity of lower ellipse of face
6	Length of nose
7	Position of centre of mouth
8	Curvature of mouth
9	Length of mouth
10	Height of centre of eyes
11	Separation of eyes
12	Slant of eyes
13	Eccentricity of eyes
14	Half length of eye
15	Position of pupil
16	Height of eyebrow
17	Angle of brow
18	Length of brow
19	Radius of ear
20	Nose width

9 observaciones de 5 variables



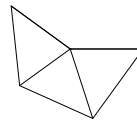
Las caras de Chernoff permiten identificar fácilmente los datos, y los valores que se representan mediante los rasgos de las caras, deben ser identificados perfectamente junto a la gráfica. El conocimiento de las tendencias de los datos permite clasificarlos de manera visual. No obstante, no contienen ningún tipo de información sobre los verdaderos valores que se van a dibujar.

Gráficos de estrellas: Cada estrella representa una observación, y las variables empiezan a representarse desde la derecha y en dirección a las agujas del reloj. El tamaño de cada línea, respecto al centro de la estrella, está relacionado con los valores reescalados de cada variable.

9 observaciones de 5 variables



1



4



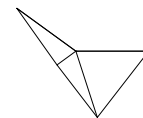
7



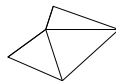
2



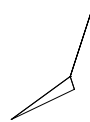
5



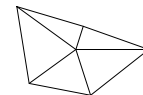
8



3



6



9