

Presentación de datos: tablas y gráficos

¿Qué se hace con los datos obtenidos en la muestra o en la población?

En esta lectura nos detendremos en la presentación de los datos obtenidos en el muestreo o en el censo. ¿Cuál será la forma de presentarlos para un mejor análisis y posterior toma de decisiones? Es mucho más amigable presentar un informe con los datos organizados en tablas o gráficos, que como una mera recopilación de datos sin procesar. Seguramente, estas herramientas las ves frecuentemente en diarios, publicaciones y cientos de aplicaciones.

¿Nos acompaña a emprender esta etapa tan útil y tan utilizada en la vida diaria?

- ≡ Presentación de la situación práctica
- ≡ Tabla de frecuencias y gráficos para datos cuantitativos sin agrupar
- ≡ Tabla de frecuencias y gráficos para datos agrupados
- ≡ Tabla de frecuencias y gráficos para datos categóricos
- ≡ Referencias

Presentación de la situación práctica

Desarrollaremos el estudio de la presentación de datos: tablas y gráficos mediante su vinculación con un caso práctico.

Una red de empresas de turismo de nuestro país ha lanzado un programa que consta de varios paquetes turísticos para las vacaciones de invierno. Por este motivo ha publicado en internet, entre otros, dos tipos de paquetes, según se prefiera o no incluir excursiones.

Los datos recogidos en la web han generado una base de datos que se muestra en la planilla de Excel adjunta, llamada ANEXO SOLICITUDES. Al final de ese documento, se explican brevemente el significado y la nomenclatura de las variables intervinientes para que se las pueda identificar en la planilla de Excel.



ANEXO SOLICITUDES.xlsx
37.1 KB



Después de una semana desde que publicó la oferta, la empresa te solicita que realices un estudio sobre los datos recogidos:

1

Realizar un análisis completo para cada una de las variables de la base de datos Solicitudes. Este análisis debe incluir:

- Construir una tabla de distribución de frecuencias para cada variable, según corresponda a cada tipo de variable. En las cuantitativas, decidir si se agrupan o no los datos y porqué.
- Graficar la información de cada variable (elegir los gráficos que corresponden para cada tabla y variable).

2

Responder sobre la base de las tablas construidas:

- ¿Qué porcentaje de solicitudes desea alojarse como máximo 20 noches?
- ¿Qué cantidad de solicitudes registra un total de pasajes por grupo superior a USD5000, aproximadamente?
- ¿Cuál es el porcentaje de solicitudes que viajarían por Aerolíneas Argentinas?

Este caso abarca todo el estudio de lo que llamamos presentación de datos. Para resolverlo necesitamos conocer algunos conceptos previos, de modo de empezar a abordar el problema. ¡Comencemos!

Tabla de frecuencias y gráficos para datos cuantitativos sin agrupar

Arreglo ordenado

Mientras algunos autores también lo llaman serie simple, otros adjudican el nombre de series estadísticas a los datos observados sin procesar. De una u otra forma, a nosotros nos interesa ordenar los datos de menor a mayor, una vez obtenidos. Esto nos facilitará la organización y presentación de datos que nos proponemos estudiar ahora.

“Un arreglo ordenado es una secuencia de datos ordenados del menor al mayor valor” (Berenson, Leviney Krehbiel, 2006, p. 30).

Ejemplo 1:

El Área de Gestión Humana de una empresa desea estudiar las ausencias del personal para realizar un análisis estadístico. Para ello toma, en un año determinado, una muestra aleatoria de 10 días. El registro obtenido es el siguiente:

Tabla 1: Registro de ausencias

La tabla muestra las ausencias del personal de una empresa en 10 días tomados al azar.

N.º de observación	Ausencias
1	2
2	4
3	2
4	1
5	0
6	2

7	1
8	3
9	0
10	2

Fuente: elaboración propia.

El arreglo ordenado de estos datos es:

0 0 1 1 2 2 2 2 3 4

Importante: en el arreglo ordenado, se escriben todas las observaciones, de menor a mayor, aunque se repitan valores.

Terminología: se designa con n al tamaño de la muestra; en este caso, $n=10$.

Pero, como no se trata de trabajar con números sin sentido, es importante que realices un ejercicio simple para ver si estás comprendiendo correctamente el verdadero significado de esos números. De ello dependerá la comprensión de los temas posteriores. Te invitamos a realizar la siguiente actividad.

Observa el arreglo ordenado que se obtiene a partir de la Tabla 1. ¿Qué representa cada dato referido al problema del ejemplo 1?

☐ Ausencias.

☐ Días.

☐ Muestra.

☐ Años.

SUBMIT

Frecuencia absoluta

Cuando se realiza un estudio de cada uno de los elementos que componen la población o muestra bajo análisis, observamos que en general, hay un número de veces en que aparece repetido un mismo valor de la variable, o bien repeticiones de la misma modalidad de un atributo.

Este número de repeticiones de un resultado, recibe el nombre de frecuencia absoluta o simplemente frecuencia.

La frecuencia absoluta es el número de veces que se repite el dato observado.

El procedimiento mediante el cual se realiza el conteo, para así determinar el número de veces que cada dato se repite, recibe el nombre de tabulación. A cada línea de la tabla se la llama clase.

Al agrupar los resultados de las observaciones en término de las veces que éstos se repiten, da lugar a las llamadas series de frecuencias o distribuciones de frecuencias.

La notación de la frecuencia absoluta es f .

El subíndice i se refiere a una posición determinada. En el caso de que dicha frecuencia encabece una columna en una tabla, la i representa la posición de la fila de la tabla a la que se refiere la frecuencia.

Antes comenzar a resolver el caso planteado al principio de la lectura, vamos a continuar con el ejemplo 1 sobre el relevamiento de ausencias de una empresa con el objetivo de comprender bien lo que es una serie de frecuencias o distribución de frecuencias.

Tabla de distribución de frecuencias para datos sin agrupar

Lo estudiado hasta ahora se puede resumir en la siguiente tabla.

Tabla 2: Tabla de distribución de frecuencias para el ejemplo 1

Esta tabla de frecuencias resume y organiza la información. Siempre se indica en la primera columna la variable en estudio ordenada como en el arreglo, pero solo los valores distintos de la variable. En la segunda columna, se indican las veces que se repite esa variable (frecuencia absoluta).

Xi	Fi
0	2
1	2
2	4
3	1

4	1
	10

Xi: cantidad de ausencias.

Fi: cantidad de días que se repite la variable.

Fuente: elaboración propia.



Observa que la sumatoria de las frecuencias es igual al número total de observaciones.



La variable en estudio no está expresada como un intervalo de ausencias. Es por eso que a este tipo de tabla de frecuencias se la denomina tabla de frecuencias para datos sin agrupar.

Clasificación de las frecuencias

FRECUENCIA ABSOLUTA

Indica la cantidad de veces que se repite la variable.

Notación: f_i

En el ejemplo 1: para la variable 2, $f_i=4$; es decir, se registran 4 días con 2 ausencias.

FRECUENCIA RELATIVA

La frecuencia relativa de una variable es el cociente entre su frecuencia absoluta y la suma de todas las frecuencias absolutas. La sumatoria de todas las frecuencias absolutas coincide con el número de observaciones.

Expresa la proporción, el peso que un valor de la variable tiene en el total de observaciones. Cuanto mayor sea la cantidad de repeticiones del valor de la variable en las observaciones, mayor será la frecuencia relativa.

Notación: fr_i

Para una variable: $fr_i = \frac{f_i}{\sum f_i}$

En el ejemplo 1, para la variable 2: $fr_i = \frac{f_i}{\sum f_i} = \frac{4}{10} = 0,4$. Esto significa que, en una proporción de 0,4, se registraron 2 ausencias, o que 2 ausencias representan una proporción de 0,4 de los días.

FRECUENCIA PORCENTUAL

La frecuencia porcentual de una variable es igual a la frecuencia relativa multiplicada por 100. Expresa el porcentaje que el valor de una variable tiene en el total de observaciones.

Notación: $fr_i\%$.

Para una variable: $fr_i\% = fr_i \times 100$.

Para el ejemplo 1, para la variable 2: $fr_i\% = fr_i \times 100 = 0,4 \times 100 = 40\%$. Esto significa que el 40% de los días se registraron 2 ausencias por día.

FRECUENCIA ABSOLUTA ACUMULADA

La frecuencia absoluta acumulada de una variable se obtiene sumando todas las frecuencias absolutas de las variables que le anteceden más la frecuencia absoluta de dicha variable.

Notación: F_i .

Para una variable: $F_i = F_{i-1} + f_i$.

En el ejemplo 1, para la variable 2: $F_i = F_{i-1} + f_i = 2 + 2 + 4 = 8$.

Significa que hubo 8 días en que se registraron 2 faltas o menos.

FRECUENCIA RELATIVA ACUMULADA

La frecuencia relativa acumulada de una variable se obtiene sumando todas las frecuencias relativas de las variables que la anteceden más la frecuencia relativa de dicha variable.

NOTACIÓN: Fr_i .

Para una variable: $Fr_i = Fr_{i-1} + fr_i$.

En el ejemplo 1, para la variable 2: $Fr_i = Fr_{i-1} + fr_i = 0,2 + 0,2 + 0,4 = 0,8$.

Significa que la proporción de días en los que hubo, como máximo, 2 ausencias es 0,8.

FRECUENCIA PORCENTUAL ACUMULADA

La frecuencia porcentual acumulada de una variable se obtiene sumando todas las frecuencias porcentuales de las variables que la anteceden más la frecuencia porcentual de dicha variable.

Notación: $Fr_i\%$.

Para una variable: $Fr_i\% = Fr\%_{i-1} + fr_i\%$.

En el ejemplo 1, para la variable 2: $Fr_i\% = Fr\%_{i-1} + fr_i\% = 20 + 20 + 40 = 80\%$.

Significa que el 80% de los días se registraron, como máximo, 2 ausencias.

FRECUENCIA DESACUMULADA

La frecuencia desacomulada de una variable es igual al número total de observaciones menos su frecuencia acumulada.

NOTACIÓN: Fd_i .

Para una variable: $Fd_i = n - F_i$.

En el ejemplo 1, para la variable 2: $Fd_i = 10 - 8 = 2$.

Significa que hubo 2 días en los que se registraron más de dos faltas.

Para fijar estos conceptos, apliquemos estas frecuencias a nuestro caso sobre la red de empresas de turismo. Iremos calculando e interpretando cada una de las frecuencias definidas anteriormente. Nos limitaremos a buscar las variables que no requieren agrupación en intervalos.

Ahora observa la planilla de Solicitudes; fíjate en las columnas que tengan pocos valores de la variable: se consideran menos de 15 valores distintos (esto puede variar con el investigador, según el objetivo de su estudio). Si tuvieran más datos distintos, entonces se pensaría en agrupar los datos en intervalos.

Analicemos cada variable:

☐

DESTINO: es cualitativa o categórica nominal

☐

CNOCHES: es cuantitativa discreta y tiene menos de 15 valores distintos que se repiten. Esta variable podría ser motivo de nuestro análisis, según lo estudiado hasta aquí.

☐

CPERSONAS: es cuantitativa discreta.

☐

CATEGORÍA: es cualitativa jerarquizada.



TIPOPAQ: es cualitativa nominal.



PASAJE USD: es cuantitativa. Pero ¿será discreta o continua? Los importes están expresados en números enteros, podríamos tratarla como cuantitativa discreta. Pero en general todo lo que es dinero se trata como cuantitativa continua. Además, los valores que adopta la variable son muchos. Es conveniente —como veremos más adelante— resumir esos valores en intervalos y luego contar cuántos datos entran en cada intervalo. A veces el tratamiento de los datos no depende solo del número de observaciones o del tipo de valor numérico, si es entero o real. Por ejemplo, podemos tener una muestra de 100 elementos y tener solo 5 valores que se repiten, ya sean números enteros o fraccionarios. En estos casos conviene hacer una tabla de frecuencias sin agrupar. En conclusión, tomaremos esta variable como cuantitativa continua.



TOTALPAS USD: cuantitativa continua, por las mismas razones que PASAJE USD.



EMPRESA: es cualitativa nominal.

Concluimos, entonces, con que las dos variables que estamos en condiciones de volcar en una tabla de frecuencias son: CNOCHES y CPERSONAS.

Comencemos con CNOCHES: cantidad de noches que los clientes se alojarán en el hotel elegido.

1

El primer paso es realizar el arreglo ordenado, que en este caso obviaremos por tratarse de muchos datos. Pero es importantísimo que recuerdes su concepto: se ordenan todos los datos de menor a mayor (si se repiten, se colocan todas las repeticiones en el arreglo).


2

Construcción de la tabla de frecuencias. Basándonos en la Tabla 2 del ejemplo 1, seguiremos la misma lógica. En la primera columna, van los valores distintos de la variable ordenados de menor a mayor, y en las siguientes columnas, la frecuencia absoluta. Pero ahora le agregaremos más columnas con los distintos tipos de frecuencias estudiadas. Comprueba tú mismo si estos valores son correctos. Puedes utilizar una calculadora o la misma planilla de Excel.

Tabla 3: Tabla de distribución de frecuencias para la variable cantidad de noches, en el caso de las empresas de turismo

Esta tabla de frecuencias resume y organiza la información. En la primera columna, se colocan los valores de la variable ordenados (CNOCHES). En el resto de las columnas, van las distintas frecuencias que corresponden a cada valor de la variable (clase).


<i>Cantidad de Noches</i> <i>xi</i>	<i>fi</i>	<i>fri</i>	<i>fri %</i>	<i>Fi</i>	<i>Fri</i>	<i>Fri%</i>	<i>Fdi</i>
2	8	0,0506	5,06	8	0,0506	5,06	150
5	27	0,1709	17,09	35	0,2215	22,15	123
7	74	0,4684	46,84	109	0,6899	68,99	49
8	4	0,0253	2,53	113	0,7152	71,52	45
10	33	0,2089	20,89	146	0,9241	92,41	12
15	8	0,0506	5,06	154	0,9747	97,47	4
20	1	0,0063	0,63	155	0,9810	98,10	3
25	2	0,0127	1,27	157	0,9937	99,37	1
30	1	0,0063	0,63	158	1,0000	100	0
Total	158	1	100				

<i>Cantidad de Noches xi</i> 	<i>fi</i>	<i>fri</i>	<i>fri %</i>	<i>Fi</i>	<i>Fri</i>	<i>Fri%</i>	<i>Fdi</i>
2	8	0,0506	5,06	8	0,0506	5,06	150
5	27	0,1709	17,09	35	0,2215	22,15	123
7	74	0,4684	46,84	109	0,6899	68,99	49
8	4	0,0253	2,53	113	0,7152	71,52	45
10	33	0,2089	20,89	146	0,9241	92,41	12
15	8	0,0506	5,06	154	0,9747	97,47	4
20	1	0,0063	0,63	155	0,9810	98,10	3
25	2	0,0127	1,27	157	0,9937	99,37	1
30	1	0,0063	0,63	158	1,0000	100	0
Total	158	1	100				


variable en estudio.

<i>Cantidad de Noches xi</i>	<i>fi</i>	<i>fri</i>	<i>fri %</i>	<i>Fi</i>	<i>Fri</i>	<i>Fri%</i>	<i>Fdi</i>
2	8	0,0506	5,06	8	0,0506	5,06	150
5	27	0,1709	17,09	35	0,2215	22,15	123
7	74	0,4684	46,84	109	0,6899	68,99	49
8	4	0,0253	2,53	113	0,7152	71,52	45
10	33	0,2089	20,89	146	0,9241	92,41	12
15	8	0,0506	5,06	154	0,9747	97,47	4
20	1	0,0063	0,63	155	0,9810	98,10	3
25	2	0,0127	1,27	157	0,9937	99,37	1
30	1	0,0063	0,63	158	1,0000	100	0
Total	158	1	100				


frecuencia absoluta.

<i>Cantidad de Noches xi</i>	<i>fi</i>	<i>fri</i> 	<i>fri %</i>	<i>Fi</i>	<i>Fri</i>	<i>Fri%</i>	<i>Fdi</i>
2	8	0,0506	5,06	8	0,0506	5,06	150
5	27	0,1709	17,09	35	0,2215	22,15	123
7	74	0,4684	46,84	109	0,6899	68,99	49
8	4	0,0253	2,53	113	0,7152	71,52	45
10	33	0,2089	20,89	146	0,9241	92,41	12
15	8	0,0506	5,06	154	0,9747	97,47	4
20	1	0,0063	0,63	155	0,9810	98,10	3
25	2	0,0127	1,27	157	0,9937	99,37	1
30	1	0,0063	0,63	158	1,0000	100	0
Total	158	1	100				


frecuencia relativa.

<i>Cantidad de Noches xi</i>	<i>fi</i>	<i>fri</i>	<i>fri %</i> 	<i>Fi</i>	<i>Fri</i>	<i>Fri%</i>	<i>Fdi</i>
2	8	0,0506	5,06	8	0,0506	5,06	150
5	27	0,1709	17,09	35	0,2215	22,15	123
7	74	0,4684	46,84	109	0,6899	68,99	49
8	4	0,0253	2,53	113	0,7152	71,52	45
10	33	0,2089	20,89	146	0,9241	92,41	12
15	8	0,0506	5,06	154	0,9747	97,47	4
20	1	0,0063	0,63	155	0,9810	98,10	3
25	2	0,0127	1,27	157	0,9937	99,37	1
30	1	0,0063	0,63	158	1,0000	100	0
Total	158	1	100				


frecuencia porcentual.

<i>Cantidad de Noches xi</i>	<i>fi</i>	<i>fri</i>	<i>fri %</i>	<i>Fi</i> 	<i>Fri</i>	<i>Fri%</i>	<i>Fdi</i>
2	8	0,0506	5,06	8	0,0506	5,06	150
5	27	0,1709	17,09	35	0,2215	22,15	123
7	74	0,4684	46,84	109	0,6899	68,99	49
8	4	0,0253	2,53	113	0,7152	71,52	45
10	33	0,2089	20,89	146	0,9241	92,41	12
15	8	0,0506	5,06	154	0,9747	97,47	4
20	1	0,0063	0,63	155	0,9810	98,10	3
25	2	0,0127	1,27	157	0,9937	99,37	1
30	1	0,0063	0,63	158	1,0000	100	0
Total	158	1	100				

frecuencia absoluta acumulada.

<i>Cantidad de Noches xi</i>	<i>fi</i>	<i>fri</i>	<i>fri %</i>	<i>Fi</i>	<i>Fri</i> 	<i>Fri%</i>	<i>Fdi</i>
2	8	0,0506	5,06	8	0,0506	5,06	150
5	27	0,1709	17,09	35	0,2215	22,15	123
7	74	0,4684	46,84	109	0,6899	68,99	49
8	4	0,0253	2,53	113	0,7152	71,52	45
10	33	0,2089	20,89	146	0,9241	92,41	12
15	8	0,0506	5,06	154	0,9747	97,47	4
20	1	0,0063	0,63	155	0,9810	98,10	3
25	2	0,0127	1,27	157	0,9937	99,37	1
30	1	0,0063	0,63	158	1,0000	100	0
Total	158	1	100				

frecuencia relativa acumulada.

<i>Cantidad de Noches xi</i>	<i>fi</i>	<i>fri</i>	<i>fri %</i>	<i>Fi</i>	<i>Fri</i>	<i>Fri%</i> 	<i>Fdi</i>
2	8	0,0506	5,06	8	0,0506	5,06	150
5	27	0,1709	17,09	35	0,2215	22,15	123
7	74	0,4684	46,84	109	0,6899	68,99	49
8	4	0,0253	2,53	113	0,7152	71,52	45
10	33	0,2089	20,89	146	0,9241	92,41	12
15	8	0,0506	5,06	154	0,9747	97,47	4
20	1	0,0063	0,63	155	0,9810	98,10	3
25	2	0,0127	1,27	157	0,9937	99,37	1
30	1	0,0063	0,63	158	1,0000	100	0
Total	158	1	100				

frecuencia relativa porcentual acumulada.

<i>Cantidad de Noches xi</i>	<i>fi</i>	<i>fri</i>	<i>fri %</i>	<i>Fi</i>	<i>Fri</i>	<i>Fri%</i>	<i>Fdi</i>
2	8	0,0506	5,06	8	0,0506	5,06	150
5	27	0,1709	17,09	35	0,2215	22,15	123
7	74	0,4684	46,84	109	0,6899	68,99	49
8	4	0,0253	2,53	113	0,7152	71,52	45
10	33	0,2089	20,89	146	0,9241	92,41	12
15	8	0,0506	5,06	154	0,9747	97,47	4
20	1	0,0063	0,63	155	0,9810	98,10	3
25	2	0,0127	1,27	157	0,9937	99,37	1
30	1	0,0063	0,63	158	1,0000	100	0
Total	158	1	100				


frecuencia desacomulada.

<i>Cantidad de Noches xi</i>	<i>fi</i>	<i>fri</i>	<i>fri %</i>	<i>Fi</i>	<i>Fri</i>	<i>Fri%</i>	<i>Fdi</i>
2	8	0,0506	5,06	8	0,0506	5,06	150
5	27	0,1709	17,09	35	0,2215	22,15	123
7	74	0,4684	46,84	109	0,6899	68,99	49
8	4	0,0253	2,53	113	0,7152	71,52	45
10	33	0,2089	20,89	146	0,9241	92,41	12
15	8	0,0506	5,06	154	0,9747	97,47	4
20	1	0,0063	0,63	155	0,9810	98,10	3
25	2	0,0127	1,27	157	0,9937	99,37	1
30	1	0,0063	0,63	158	1,0000	100	0
Total	158	1	100				

0,0506 = 8/158. El cociente conviene expresarlo con cuatro decimales, redondeando el último.

<i>Cantidad de Noches xi</i>	<i>fi</i>	<i>fri</i>	<i>fri %</i>	<i>Fi</i>	<i>Fri</i>	<i>Fri%</i>	<i>Fdi</i>
2	8	0,0506	i 5,06	8	0,0506	5,06	150
5	27	0,1709	17,09	35	0,2215	22,15	123
7	74	0,4684	46,84	109	0,6899	68,99	49
8	4	0,0253	2,53	113	0,7152	71,52	45
10	33	0,2089	20,89	146	0,9241	92,41	12
15	8	0,0506	5,06	154	0,9747	97,47	4
20	1	0,0063	0,63	155	0,9810	98,10	3
25	2	0,0127	1,27	157	0,9937	99,37	1
30	1	0,0063	0,63	158	1,0000	100	0
Total	158	1	100				

5,06 = 0,056 × 100.

<i>Cantidad de Noches xi</i>	<i>fi</i>	<i>fri</i>	<i>fri %</i>	<i>Fi</i>	<i>Fri</i>	<i>Fri%</i>	<i>Fdi</i>
2	8	0,0506	5,06	 8	0,0506	5,06	150
5	27	0,1709	17,09	35	0,2215	22,15	123
7	74	0,4684	46,84	109	0,6899	68,99	49
8	4	0,0253	2,53	113	0,7152	71,52	45
10	33	0,2089	20,89	146	0,9241	92,41	12
15	8	0,0506	5,06	154	0,9747	97,47	4
20	1	0,0063	0,63	155	0,9810	98,10	3
25	2	0,0127	1,27	157	0,9937	99,37	1
30	1	0,0063	0,63	158	1,0000	100	0
Total	158	1	100				

se repite la primera frecuencia absoluta.

<i>Cantidad de Noches xi</i>	<i>fi</i>	<i>fri</i>	<i>fri %</i>	<i>Fi</i>	<i>Fri</i>	<i>Fri%</i>	<i>Fdi</i>
2	8	0,0506	5,06	8	0,0506	5,06	150
5	27	0,1709	17,09	35	0,2215	22,15	123
7	74	0,4684	46,84	109	0,6899	68,99	49
8	4	0,0253	2,53	113	0,7152	71,52	45
10	33	0,2089	20,89	146	0,9241	92,41	12
15	8	0,0506	5,06	154	0,9747	97,47	4
20	1	0,0063	0,63	155	0,9810	98,10	3
25	2	0,0127	1,27	157	0,9937	99,37	1
30	1	0,0063	0,63	158	1,0000	100	0
Total	158	1	100				


35 = 8 + 27. Se le suma a la primera frecuencia absoluta anterior (8) la frecuencia absoluta de la clase (27).

<i>Cantidad de Noches xi</i>	<i>fi</i>	<i>fri</i>	<i>fri %</i>	<i>Fi</i>	<i>Fri</i>	<i>Fri%</i>	<i>Fdi</i>
2	8	0,0506	5,06	8	0,0506	5,06	150
5	27	0,1709	17,09	35	0,2215	22,15	123
7	74	0,4684	46,84	109	0,6899	68,99	49
8	4	0,0253	2,53	113	0,7152	71,52	45
10	33	0,2089	20,89	146	0,9241	92,41	12
15	8	0,0506	5,06	154	0,9747	97,47	4
20	1	0,0063	0,63	155	0,9810	98,10	3
25	2	0,0127	1,27	157	0,9937	99,37	1
30	1	0,0063	0,63	158	1,0000	100	0
Total	158	1	100				

0,2215 = 0,0506 + 0,1709.

<i>Cantidad de Noches xi</i>	<i>fi</i>	<i>fri</i>	<i>fri %</i>	<i>Fi</i>	<i>Fri</i>	<i>Fri%</i>	<i>Fdi</i>
2	8	0,0506	5,06	8	0,0506	5,06	150
5	27	0,1709	17,09	35	0,2215	22,15	123
7	74	0,4684	46,84	109	0,6899	68,99	49
8	4	0,0253	2,53	113	0,7152	71,52	45
10	33	0,2089	20,89	146	0,9241	92,41	12
15	8	0,0506	5,06	154	0,9747	97,47	4
20	1	0,0063	0,63	155	0,9810	98,10	3
25	2	0,0127	1,27	157	0,9937	99,37	1
30	1	0,0063	0,63	158	1,0000	100	0
Total	158	1	100				

22,15 = $0,2215 \times 100$.

<i>Cantidad de Noches xi</i>	<i>fi</i>	<i>fri</i>	<i>fri %</i>	<i>Fi</i>	<i>Fri</i>	<i>Fri%</i>	<i>Fdi</i>
2	8	0,0506	5,06	8	0,0506	5,06	 150
5	27	0,1709	17,09	35	0,2215	22,15	123
7	74	0,4684	46,84	109	0,6899	68,99	49
8	4	0,0253	2,53	113	0,7152	71,52	45
10	33	0,2089	20,89	146	0,9241	92,41	12
15	8	0,0506	5,06	154	0,9747	97,47	4
20	1	0,0063	0,63	155	0,9810	98,10	3
25	2	0,0127	1,27	157	0,9937	99,37	1
30	1	0,0063	0,63	158	1,0000	100	0
Total	158	1	100				

150 = 158 – 8. Al total de observaciones se le resta la frecuencia absoluta de esa clase.

<i>Cantidad de Noches xi</i>	<i>fi</i>	<i>fri</i>	<i>fri %</i>	<i>Fi</i>	<i>Fri</i>	<i>Fri%</i>	<i>Fdi</i>
2	8	0,0506	5,06	8	0,0506	5,06	150
5	27	0,1709	17,09	35	0,2215	22,15	123
7	74	0,4684	46,84	109	0,6899	68,99	49
8	4	0,0253	2,53	113	0,7152	71,52	45
10	33	0,2089	20,89	146	0,9241	92,41	12
15	8	0,0506	5,06	154	0,9747	97,47	4
20	1	0,0063	0,63	155	0,9810	98,10	3
25	2	0,0127	1,27	157	0,9937	99,37	1
30	1	0,0063	0,63	158	1,0000	100	0
Total	158	1	100				

$123 = 150 - 27$. A la frecuencia des acumulada anterior se le resta la frecuencia absoluta de esa clase.

Aclaraciones:



Esta tabla se puede hacer directamente en una planilla de Excel, utilizando funciones para contar los datos que se repiten, como la función Contar Si. Además, hay una gran variedad de software que lo hace y que están disponibles en el mercado.



Por otro lado, hay que decir que muchas veces no es conveniente presentar tablas con todas las frecuencias, sino solo con aquellas que más interesan para la investigación. En este caso lo hemos hecho así con fines pedagógicos.

Para interpretar la tabla correctamente, te invitamos a mirar la Tabla 3 de frecuencias. Luego, podrás responder a la siguiente actividad:

El porcentaje de solicitudes que desean quedarse 8 noches es del ____ %.

Escriba su respuesta aquí

SUBMIT

Gráficos

A fin de poder realizar una lectura rápida de la distribución de datos y sacar conclusiones inmediatas de ella, es muy conveniente presentar la información recolectada en gráficos. Para cada tipo de datos, existen gráficos diferentes. A continuación, citaremos los más comunes y estándares:



Para datos cuantitativos sin agrupar en clases: diagrama de bastones, gráficos circulares.



Para datos cuantitativos agrupados en clases: histogramas y polígonos.



Para datos categóricos: diagramas circulares y de barras, entre una gran variedad.



Los gráficos circulares o de torta pueden utilizarse para cualquier tipo de variable si se hacen las referencias correspondientes.

Gráficos para datos sin agrupar

Los gráficos que más se adaptan a estos tipos de datos (discretos y con pocos valores de la variable) son los diagramas de bastones.

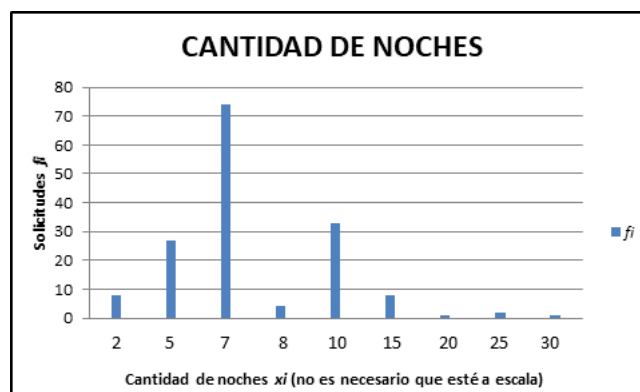
Diagrama de bastones

Los gráficos para variables cuantitativas —en general, como veremos más adelante— tienen un eje horizontal en el que se representan los valores de la variable en estudio y un eje vertical en el que se marcan a escala las frecuencias que se quieren representar.

Para el caso que venimos estudiando de la red de empresas de turismo, el gráfico que muestra la cantidad de noches solicitadas por los clientes se muestra a continuación.

Figura 1: Diagrama de bastones

El gráfico muestra la frecuencia absoluta correspondiente a cada cantidad de noches solicitadas.



Fuente: elaboración propia.

Obsérvese que cada bastón parte de un valor; no son barras porque cada valor de la variable es un punto en el eje horizontal. Si pudiera dibujarse un segmento, mejor. La altura de cada bastón está dada por la frecuencia del valor de la variable. Podría hacerse un gráfico para cada tipo de frecuencia que se quiera visualizar.

Siempre en el eje vertical va la frecuencia elegida, a escala.

Además, el eje horizontal no es necesario que esté a escala porque lo que importa es comparar las frecuencias de cada variable. No obstante, hay autores que sostienen hacerlo a escala y dejar con frecuencia cero los valores que no están en la recolección de datos.

Lo importante es que, en el eje vertical —el de las frecuencias—, siempre se tome una escala.

Diagramas circulares o de sectores

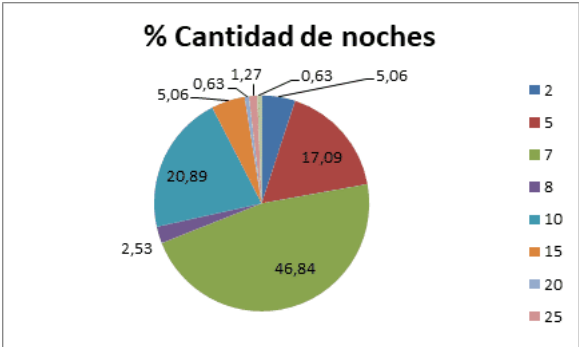
En este tipo de gráfico, consideraremos que los 360° del ángulo central del círculo representan el 100% de los datos relevados. Además, cada una de las clases estará dada por un sector cuyo ángulo central será proporcional a su frecuencia porcentual. Por eso, con una regla de tres simple directa se procede a graficarlas.

En algunas ocasiones es interesante mostrar la relación porcentual entre las distintas frecuencias de la variable. Los gráficos de sectores o circulares se adaptan muy bien a estas situaciones.

Observemos cómo quedaría, en nuestro caso de estudio, la representación porcentual de la cantidad de noches solicitadas por los clientes.

Figura 2: Diagrama de sectores

El gráfico muestra la frecuencia absoluta porcentual correspondiente a cada cantidad de noches solicitadas.



Fuente: elaboración propia.

Conviene elegir qué tipo de gráfico presentar en un informe. Al igual que la tabla de distribución de frecuencias, dijimos que no es aconsejable presentar una tabla con todos los tipos de frecuencias. Aquí también conviene seleccionar qué tipo de gráfico será más amigable para presentar la información.

Dejamos, a modo de práctica personal, realizar la tabla de frecuencias y los gráficos que elijas hacer para la otra variable discreta CPERSONAS (cantidad de personas) referida al caso de las empresas de turismo. Puedes hacer gráficos de frecuencia absoluta, relativa, porcentual o acumulada; es decir, cada columna de la tabla es un gráfico distinto.

Es conveniente que, antes de seguir, repases la clasificación que hicimos de cada variable en el apartado de "Clasificación de las frecuencias" para poder seguir resolviendo la situación planteada al comienzo de la lectura.

Tabla de frecuencias y gráficos para datos agrupados

Intervalos de clase

Cuando los valores que adopta la variable son demasiados (15 o más, según algunos autores), no resulta útil un ordenamiento de datos en tablas para datos sin agrupar. Conviene, entonces, resumirlos en intervalos de clase. De otra manera, sería engorroso y poco amigable leerlos. Este caso suele suceder con mayor frecuencia cuando las variables son cuantitativas continuas. Para estas situaciones, necesitamos resumir la información sin que pierda su calidad al momento de utilizarla en su interpretación y en la toma de decisiones.

Partiremos del caso presentado al comienzo de la lectura. Según el análisis de las variables que intervienen (Excel), ¿qué variable es adecuada para ser presentada en una tabla de datos agrupados?


En principio, tenemos dos:

- ☐ PASAJE USD.
- ☐ TOTALPAS USD.

Vamos a resolver una de las dos variables, dejando a modo de estudio personal la construcción de la tabla de frecuencias y los gráficos de la otra.


Elijamos TOTALPAS USD: cantidad de dólares totales que arroja en pasajes cada solicitud. Para ello tendrás que buscar la planilla “ANEXO SOLICITUDES” y calcular el total de dólares, como se indica en el anexo ($CPERSONAS \times PASAJE\ USD$), para cada observación. Luego deberás hacer el arreglo ordenado para poder visualizar cuál es el menor y el mayor importe.

Para facilitar el análisis, abre el “ANEXO SOLICITUDES CON TOTALPAS”, y verás que ya están hechas las operaciones.



ANEXO SOLICITUDES CON TOTALPAS.xlsx

20.9 KB



Comencemos con la construcción de la tabla que, como veremos, tiene ciertas reglas para respetar con el fin de evitar un sesgo en la información.

Condiciones para construir una tabla con intervalos

1

¿Qué cantidad de intervalos conviene hacer? La determinación del número de intervalos (k) está relacionada con la cantidad de datos que tenemos (n). Algunos técnicos definen:

$$k = \sqrt{n}$$

Otros se inclinan por la expresión del método Sturges:

$$k = 1 + \log_2 n$$

$$k = 1 + 3,332 \log n$$

Pero, en un plano aún más general, se puede definir el valor de k teniendo en cuenta:

$$5 \leq k \leq 15$$

Es decir, el número de intervalos debe estar comprendido entre 5 y 15. Menos de 5 intervalos da lugar a una distribución con información insuficiente y frecuencias muy altas, y más de 15 intervalos, genera una distribución que resulta complicado operar.

Elijamos una, la más utilizada es la raíz cuadrada del número de observaciones. En el caso de nuestra red de empresas de turismo, $n = 158$, por lo que nos conviene tomar un intervalo de

$$k = \sqrt{158} \cong 12,7$$

Como la cantidad de intervalos debe ser un número natural, debemos elegir entre 12 y 13. Por lo general, se acerca al valor más próximo, aunque todo depende de la investigación que se quiera realizar. En este caso, si elegimos 13, corremos el peligro de que quede más de una clase en cero, pues hay un salto en los precios superiores que queremos que se refleje de alguna manera en la tabla.

2

¿Cómo calculamos la amplitud de cada intervalo? A la amplitud de intervalo la calculamos así:

$$\Delta x = \frac{\text{Dato mayor} - \text{Dato menor}}{k}$$

Porque todos los datos deben tener la misma longitud para que se aprecie bien la distribución de frecuencias, sino no se podrían comparar las clases.

En el caso que venimos analizando (ver anexo "SOLICITUDES CON TOTALPAS"):

$$\Delta x = \frac{7990 - 338}{12} = 637,7$$

Es decir, cada intervalo es de aproximadamente USD637,7. Es cierto que, con los redondeos, se pierde algo de información, pero es más práctico tomar números enteros o con un decimal a lo sumo. Eso depende del tipo de investigación que se realice: no es lo mismo trabajar con mg que con dólares. Probablemente, a un bioquímico sí le interese redondear a la segunda o tercera cifra decimal, **pero la longitud del intervalo debe redondearse siempre hacia arriba.**

En este caso tomaremos USD638 porque, si truncamos los decimales, podrían quedar importes fuera del rango. Con esto nos aseguramos de que todos los valores queden dentro de todos los intervalos.

3

Debe tratarse de que no existan intervalos con frecuencia cero; si esa situación se presentara, sería necesario modificar la cantidad de intervalos o la amplitud de cada intervalo. Esta condición no es tan estricta, pues depende del tipo de investigación que se quiera realizar. Si la cantidad de intervalos es la correcta y los límites de cada intervalo están bien calculados, no es imprescindible que todas las clases tengan frecuencias distintas de cero. Si esta situación se da, es una información extra para el investigador, quien deberá decidir si cambia la cantidad de intervalos o la amplitud.

4

Al fijar los extremos de los intervalos, debe atenderse a la posibilidad de que uno de ellos pueda ser cerrado y el otro abierto. En general, como son intervalos reales, se toman cerrados por izquierda y abiertos por derecha: $[L_i, L_s)$. El límite inferior del intervalo se considera dentro, y el superior corresponde al intervalo siguiente.

5

Todos los intervalos deben tener el mismo tamaño y, cuando esto no sea posible, los intervalos de diferente amplitud deben ubicarse en los extremos.

6

Hay que tratar de evitar que las observaciones coincidan con los extremos de cada intervalo, así desaparece la duda en cuanto a la pertenencia de una observación a un intervalo o a su adyacente.

Teniendo en cuenta todas estas condiciones, nos proponemos ahora comenzar a hacer la distribución de intervalos correspondiente a nuestro caso de la red de empresas de turismo.

Tabla de frecuencias con intervalos de clase

Lo primero es establecer los límites de cada intervalo. Estos van en la primera columna de la tabla—lugar que ocupa la variable en estudio—, que, en el caso que estamos analizando, es el total que cada grupo debe pagar (solicitud) en concepto de pasajes, llamado TOTALPAS.

Para ser más gráficos, diagramamos la situación y calculamos los extremos de cada intervalo. Para construir 12 intervalos del mismo tamaño, es evidente que tenemos que dividir el rango (Dato mayor – Dato menor) en K =12.

Figura 3: Esquema auxiliar para determinar los límites de cada intervalo en el caso de estudio

El esquema es una visualización para comprender cómo calcular los límites de los intervalos de la variable TOTALPAS, en dólares, en el caso de la red de empresas de turismo.



Fuente: elaboración propia.

Nos proponemos entonces delimitar los límites de cada clase, que son puntos sobre el eje horizontal. Son intervalos de la variable en estudio.

Verifica los siguientes límites sumando al límite inferior de cada intervalo el tamaño del intervalo, que es 638

338	976	1614	2252	2890	3528	4166	4804	5442	6080	6718	7356	7994
-----	-----	------	------	------	------	------	------	------	------	------	------	------

Como vemos, el límite superior del último intervalo excede en 4 dólares al dato mayor. En muchos casos, como este, para no deformar la distribución y no darle mayor oportunidad al último intervalo que incluya alguna frecuencia que no le corresponda, se distribuyen los excesos por redondeo. Este exceso (por considerar 638 en lugar de 637,7) puede salvarse comenzando antes del valor mínimo, que es de 338. Igualmente, no estamos aumentando el tamaño ni del primer ni del último intervalo porque se corren todos los límites.

Lo más conveniente es comenzar en 336 y terminar en 7992, y la distribución minimizaría los errores que pudieran sobrevenir si cargamos todos los excesos al último intervalo. Vemos que se verifica que el dato menor y el dato mayor quedan dentro del rango.

Los límites de los intervalos quedarían de la siguiente forma:

		161	225	288	352	416	480	544	607	671	735	799	
336	974	2	0	8	6	4	2	0	8	6	4	2	

Tabla 4: Tabla de distribución de frecuencias para datos agrupados en clases para la variable TOTALPAS del caso de las empresas de turismo

La tabla muestra las doce clases de la variable cantidad de dólares por solicitud en concepto de pasajes, con sus correspondientes frecuencias.

N.º de clase	Total, Pasajes (U\$S) [Li-Ls)	Solicitudes de pasajes fi	fri	fri%	Fi	Fdi
1	336-974	24	0,1519	15,19	24	134
2	974-1612	13	0,0823	8,23	37	121
3	1612-2250	4	0,0253	2,53	41	117
4	2250-2888	53	0,3354	33,54	94	64
5	2888-3526	15	0,0949	9,49	109	49
6	3526-4164	25	0,1582	15,82	134	24
7	4164-4802	5	0,0316	3,16	139	19
8	4802-5440	11	0,0696	6,96	150	8
9	5440-6078	3	0,0190	1,90	153	5
10	6078-6716	3	0,0190	1,90	156	2
11	6716-7354	0	0,0000	0,00	156	2
12	7354-7992	2	0,0127	1,27	158	0
	158	1	100			

Fuente: elaboración propia.

Solo hemos indicado algunas frecuencias, pero observa que hay una clase con frecuencia cero. Tenemos un problema con respecto a las condiciones de la construcción de intervalos, pero puede dejarse así. En caso contrario, habría que construir una tabla de menos cantidad de intervalos y se perdería información. Seguramente, tal situación se presenta por el salto que da el importe en dólares de la clase 10 a la 12.

Marca de clase

Este concepto es muy importante en la construcción de una tabla para datos agrupados, pues, como veremos más adelante, tiene muchas aplicaciones. Por ejemplo, a los efectos de futuros cálculos, como el promedio, necesitaremos un valor que represente a la clase.

La marca de clase es el punto medio de cada intervalo. Es el promedio entre los límites de la clase. En definitiva, representa a esa clase.

Se la designa: xm_i .

Entonces

$$xm_i = \frac{Li+Ls}{2}$$

Li: límite inferior del intervalo.

Ls: límite superior del intervalo.

Por ejemplo, en la variable TOTALPAS del caso que venimos analizando, las marcas de clase quedarían de la siguiente manera:

Tabla 5: Tabla de intervalos de clase con la marca de clase

La tabla solo muestra las columnas pertenecientes a la variable en estudio TOTALPAS. Recuerda que es solo una parte de la tabla la que se muestra con fines pedagógicos, pues faltan las columnas de las frecuencias que se visualizan en la Tabla 4.

N.º clase	[Li-Ls)	xmi
1	336-974	655
2	974-1612	1293
3	1612-2250	1931
4	2250-2888	2569
5	2888-3526	3207
6	3526-4164	3845
7	4164-4802	4483

8	4802-5440	5121
9	5440-6078	5759
10	6078-6716	6397
11	6716-7354	7035
12	7354-7992	7673

Fuente: elaboración propia.

Gráficos para datos agrupados

Los gráficos más utilizados para este tipo de datos, como ya dijimos, son los histogramas y los polígonos. En algunos casos también se utilizan los circulares o de sectores ya estudiados.

HISTOGRAMAS

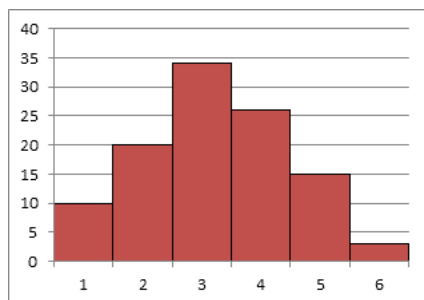
POLÍGONOS

Los histogramas son rectángulos pegados uno a continuación del otro. ¿Por qué? Pues, si se trata de variables continuas, el extremo inferior de un intervalo coincide con el extremo superior del intervalo siguiente.

Según el tipo de frecuencia que se quiera representar, los histogramas pueden ser:

- de frecuencias absolutas;
- de frecuencias relativas;
- de frecuencias porcentuales;
- de frecuencias absolutas acumuladas;
- de frecuencias relativas acumuladas;
- de frecuencias desacomuladas.

Figura 4: Histograma de frecuencias absolutas



HISTOGRAMAS

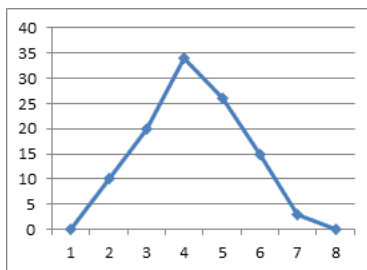
POLÍGONOS

Los polígonos están formados por segmentos consecutivos. Cada punto que une dos segmentos tiene como coordenadas la marca de clase de cada intervalo y la frecuencia de dicho intervalo.

Según el tipo de frecuencia que se quiera representar, los polígonos pueden ser:

- de frecuencias absolutas;
- de frecuencias relativas;
- de frecuencias porcentuales;
- de frecuencias absolutas acumuladas, lo que se conoce como ojiva;
- de frecuencias relativas acumuladas;
- de frecuencias desacomuladas.

Figura 5: Polígono de frecuencias absolutas



Es importante destacar lo siguiente:



En los histogramas y polígonos, lo importante es hacer un corte en el eje horizontal, mostrando que la escala no siempre comienza en cero. Lo importante es, a partir de un punto, tomar segmentos iguales sobre dicho eje que representen los intervalos con igual longitud.



En los histogramas: el área de cada rectángulo está dada por el producto entre la base Δx y la altura, que está determinada por la frecuencia. Por lo tanto, si sumamos las áreas de todos los rectángulos de un histograma, nos da como resultado el número de todas las observaciones: n .



Los histogramas de frecuencias relativas y porcentuales tienen la misma forma que los de frecuencias absolutas, solo cambian las escalas en el eje vertical. Esto era de esperar, ya que se guardan las proporciones entre estas tres frecuencias.



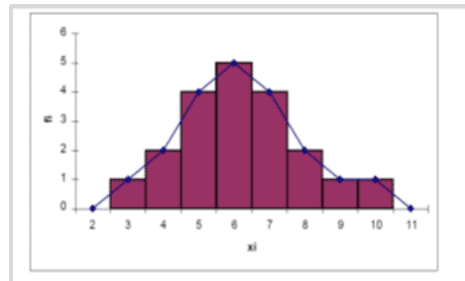
Los polígonos de frecuencias absolutas, relativas y porcentuales deben cerrarse sobre el eje horizontal, suponiendo una clase anterior y una posterior con frecuencia cero.



Si, en un mismo gráfico, superponemos el histograma y el polígono de frecuencias, observamos que el área encerrada por el polígono de frecuencias es igual al área encerrada por el histograma de frecuencias, ya que, en cada rectángulo, se compensan entre sí áreas en exceso con áreas en defecto.

Figura 6: Histograma y polígono de frecuencias absolutas

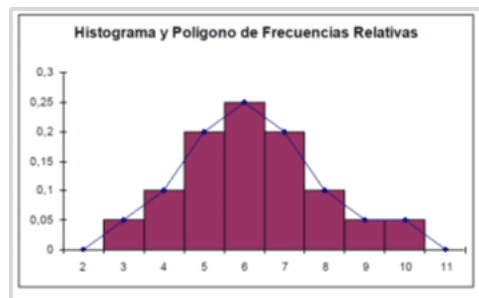
En el gráfico se muestran histograma y polígono superpuestos de una misma distribución.



Fuente: elaboración propia.

Figura 7: Histograma y polígono de frecuencias relativas

En el gráfico se muestran histograma y polígono superpuestos de una misma distribución. Observa que la diferencia con el histograma y el polígono de frecuencias absolutas está en las escalas del eje vertical.



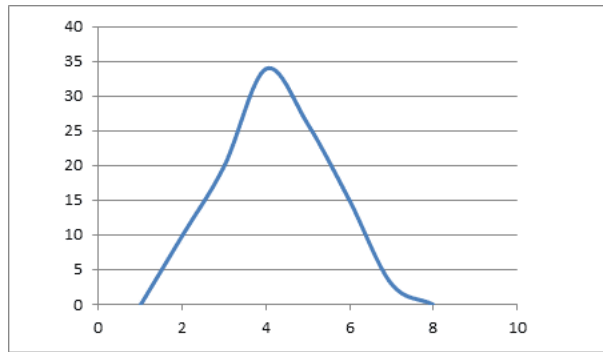
Fuente: elaboración propia.



Cuando el número de intervalos tiende a infinito y la longitud de cada clase tiende a cero, se suaviza la forma del polígono y nos queda una curva de frecuencias.

Figura 8: Curva de frecuencias

En el gráfico se muestra cómo queda un polígono si se aumenta indefinidamente la cantidad de intervalos y se disminuye la amplitud de estos a un punto. Esto muestra un caso extremo en que a cada punto de la variable en estudio le corresponde una frecuencia determinada. Es una forma de visualización porque sería imposible graficarlo punto por punto.



Fuente: elaboración propia.

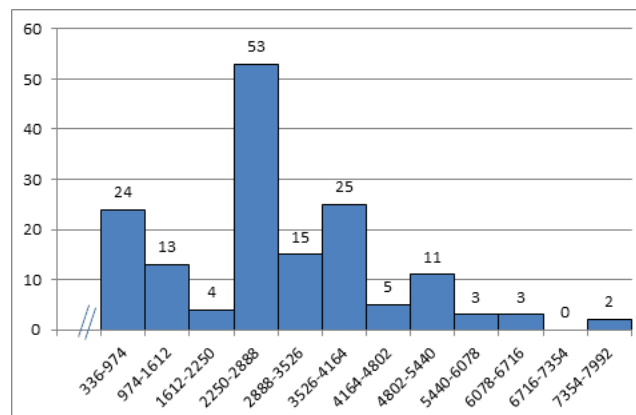
Volvamos al caso de la red de empresas de turismo. En el informe se pide el análisis de las variables, las tablas y los gráficos.

Hasta ahora hemos analizado todas las variables, construido la tabla de frecuencias para una variable discreta CNOCHES y para una variable continua TOTALPAS. Además, hemos construido el diagrama de bastones para la variable CNOCHES. Podemos elegir ahora que gráfico hacer para TOTALPAS, que es una variable continua.

Con lo estudiado hasta aquí, podríamos elegir el histograma de frecuencias absolutas, que nos quedaría de la siguiente manera:

Figura 9: Histograma de frecuencias absolutas para la variable TOTALPAS

El gráfico muestra un histograma de frecuencias absolutas para el caso de la red de empresas de turismo. En el eje horizontal, se colocan los intervalos en dólares que representan los importes totales en concepto de pasajes. El eje vertical indica la cantidad de solicitudes comprendidas en cada intervalo.



Fuente: elaboración propia.

También podrías elegir otro gráfico, por ejemplo, el polígono, un histograma de frecuencias porcentuales o también el de frecuencias acumuladas.

Recordemos lo que se pedía en la primera parte del informe que solicita la empresa de turismo:



Realizar un análisis completo para cada una de las variables de la base de datos "SOLICITUDES". Este análisis debe incluir lo siguiente:

- Construir una tabla de distribución de frecuencias para cada variable, según corresponda a cada tipo de variable. En las cuantitativas, decide si se agrupan o no los datos y porqué.
- Graficar la información de cada variable (elige los gráficos que corresponden para cada tabla y variable).

Ya hemos visto los gráficos más utilizados según la variable que se está estudiando. Y además hemos graficado dos de las variables del caso en estudio. Solo nos faltarían las tablas y gráficos para los datos categóricos.

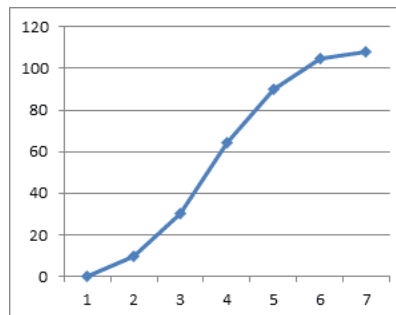
Antes de comenzar con los datos cualitativos o categóricos, afiancemos conceptos sobre las frecuencias acumuladas.

Gráficos de frecuencias acumuladas

Dedicamos un apartado para las frecuencias acumuladas porque estas, por definición, no son proporcionales a las frecuencias absolutas. Son el resultado de una acumulación de frecuencias de clases anteriores y, por lo tanto, sus gráficas tienen características particulares. Observa cómo quedaría una ojiva: polígono de frecuencias acumuladas.

Figura 10: Polígono de frecuencias absolutas acumuladas

El gráfico muestra un ejemplo para visualizar cómo queda el gráfico de frecuencias acumuladas. Es una poligonal cuyos puntos están formados por la marca de clase y la frecuencia absoluta correspondiente. Queda abierto por derecha. Por izquierda se supone una clase anterior imaginaria con frecuencia cero para poder cerrarlo.



Fuente: elaboración propia.

Los polígonos de frecuencias acumuladas, llamados ojivas, solo se cierran sobre el eje horizontal por el lado izquierdo. Como estudiamos anteriormente, también pueden construirse histogramas de frecuencias acumuladas.

Interpretación de la frecuencia acumulada

La frecuencia acumulada (F_i) de una clase indica la cantidad de valores menores o iguales a la variable de esa clase.

Observa la Tabla 4, correspondiente a las frecuencias para cada clase de la variable TOTALPAS. ¿Qué significa la frecuencia acumulada de la clase 4? Nos fijamos en la tabla y dicha frecuencia es 94. Podemos concluir con que hay 94 solicitudes que arrojan un total de pasajes entre 336 y 2888 dólares.

Tabla de frecuencias y gráficos para datos categóricos

Tabla de frecuencias para datos categóricos

Los datos categóricos tienen un tratamiento mucho más sencillo. ¿Para qué nos interesaría una tabla de frecuencias para datos categóricos? Por ejemplo, vayamos al "ANEXO SOLICITUDES" que venimos trabajando y seleccionemos la variable EMPRESA. Identifiquemos las distintas modalidades de la variable EMPRESA: TAM, ALITALIA, AA, IBERIA, AIREUROPA, DELTA y TAP AIR PORTUGAL. Hay 7 empresas distintas. Podemos resumir esta información en la siguiente tabla:

Tabla 6: Tabla de frecuencias absolutas y porcentuales para la variable EMPRESA, en el caso de la red de empresas de turismo

En la tabla se muestran las empresas aéreas que están involucradas en las solicitudes. Para cada una se muestra la cantidad de solicitudes y qué porcentaje del total representan.

Empresas (xi)	fi	fri%
AA	72	45,57
AIR EUROPA	16	10,13
ALITALIA	7	4,43
DELTA	7	4,43
IBERIA	3	1,90
TAM	52	32,91
TAP AIR PORTUGAL	1	0,63
	158	100

Fuente: elaboración propia.

¡Observa que no hacen falta más frecuencias! ¿Para qué sería útil una frecuencia acumulada, si el orden de las empresas no tiene importancia? En este caso, optamos por escribirlas en orden alfabético, pero podrían haber estado en otro orden. Además, lo único significativo es saber qué empresa es la que más se utiliza

o la que menos es requerida conforme al trayecto que elijan los usuarios.

La frecuencia porcentual puede ser útil porque nos habla de una relación con respecto al total de solicitudes de las empresas involucradas. Pero aquí terminaría el análisis. Como veremos más adelante, no podemos calcular un promedio, sino solo la moda (que está dada por la variable que más veces se repite), que en este caso es Aerolíneas Argentinas (AA).

Gráficos para datos categóricos

Los gráficos para este tipo de variables son muy variados. El más común es el diagrama de barras y el circular.

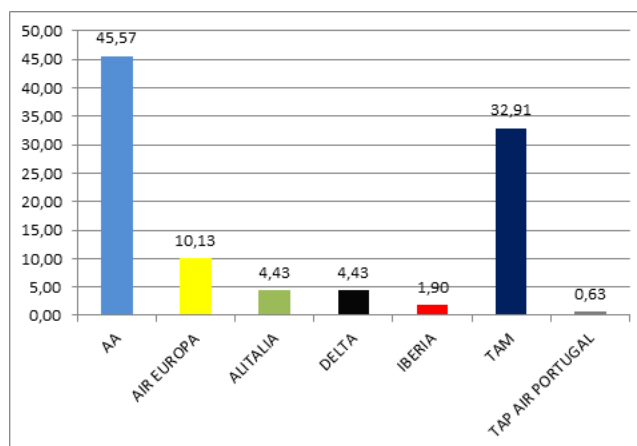
Diagrama de barras

Estos diagramas proporcionan información en un gráfico de dos dimensiones. En el eje horizontal, se muestran los valores de la variable (las denominaciones o modalidades que estamos midiendo), como las calificaciones en una evaluación o los diferentes deportes practicados en un club social. En el vertical, señalamos las frecuencias de las clases mostradas en el eje horizontal. De esta manera, la altura de las barras mide el número de observaciones que hay en cada clase señalada en el eje horizontal.

A continuación, elegimos un gráfico para representar el caso de la variable EMPRESAS.

Figura 11: Gráfico de frecuencias porcentuales de barras verticales

En este gráfico vemos cómo la altura de cada barra representa los porcentajes de solicitudes de pasajes aéreos en la empresa que figura en el eje horizontal.



Fuente: elaboración propia.

También podríamos hacer gráficos de barras horizontales o circulares. En la planilla de Excel, existen muy variados gráficos, pero ¡cuidado!, hay que elegir gráficos que no confundan la lectura de lo que se quiere informar, por más vistosos que sean.

Responderemos a la última parte del informe sobre la red de empresas de turismo. En esta parte se pide lo siguiente:



Responder sobre la base de las tablas construidas:

- ¿Qué porcentaje de solicitudes desea alojarse, como máximo, 20 noches? Verifica que es el 98,1 % (ver columna de frecuencias porcentuales acumuladas en la Tabla 2 de esta lectura, correspondiente al valor 20).
- ¿Qué cantidad de solicitudes registran un total de pasajes por grupo superior a 5000 dólares, aproximadamente? Verifica que son aproximadamente 13 solicitudes.
- ¿Cuál es el porcentaje de solicitudes que viajarían por Aerolíneas Argentinas? El porcentaje es 45,57 %.

Te animamos a releer el caso de lectura obligatoria de la Lectura 1, sobre la violencia y la criminalidad en Uruguay, y avanzar en ese documento con los temas estudiados en esta lectura. En esta investigación verás cómo en estadística descriptiva algunos criterios se adaptan, mientras que otros no cambian. Hay terminologías distintas, pero el objetivo es que puedas comprender un estudio estadístico, aunque cambien algunos enfoques. Lo importante es enriquecer lo ya estudiado.

Lectura obligatoria: Curso de estadística básica aplicada a la gestión

En este trabajo están organizados de manera sistemática todos los conceptos estudiados en esta lectura. Se trata de una investigación sobre la violencia y la criminalidad en Uruguay. Se abarcan conceptos teórico-prácticos ya estudiados con el fin de reforzarlos mediante un caso real. Se recomienda la lectura desde la página 08 hasta la página 18 inclusive.



Curso de Estadística Básica aplicada a la Gestión.pdf
209.4 KB



Fuente: Repetto Pereira, D. M. (2009). Curso de Estadística Básica aplicada a la Gestión. Publicado en Escuela Nacional de Administración Pública (ENAP). Oficina Nacional del Servicio Civil. Recuperado de link https://www.onsc.gub.uy/onsc1/images/stories/Enap/Material/PDM2/PDM2_MII06.pdf

Diagrama de Pareto

No podemos dejar de mencionar esta eficiente técnica que te ayudará en tus decisiones y está basada en todo lo que estudiamos en esta lectura. ¡Te sorprenderá! Es una aplicación directa de las tablas y gráficos, y se utiliza para datos categóricos en todas las disciplinas. También se llama regla 20/80 y su lectura es ágil.

Lectura obligatoria: Diagrama de Pareto

En la lectura se explica qué es un diagrama de Pareto, para qué sirve y cómo se construye.



Diagrama de Pareto - Blog de La Calidad.pdf

438.9 KB



Fuente: Arenhart De Bastiani, J. y Martins, R. (2018). Diagrama de Pareto. Recuperado de <https://blogdelacalidad.com/diagrama-de-pareto/>

Referencias

Berenzon, M., Levine, D. y Krehbiel, T. (2006). *Presentación de datos en tablas y gráficas*. En Autores, Estadística descriptiva (pp. 21-70). Naucalpan de Juárez, MX. Pearson.