

UNIDAD N° 1

ESTADÍSTICA DESCRIPTIVA

Organización de los datos. Métodos tabulares: Tabla de frecuencias para variables discretas. Tabla de frecuencias para variables continuas.

UNIDAD N° 1: ESTADÍSTICA DESCRIPTIVA

ORGANIZACIÓN DE LOS DATOS – MÉTODOS TABULARES

¿Por qué y cómo construir una tabla de frecuencias?

En estadística, uno de los primeros desafíos al trabajar con datos es encontrar una forma clara y ordenada de presentarlos. Cuando la cantidad de observaciones es grande o los valores son diversos, se vuelve difícil detectar patrones, comparar frecuencias o sacar conclusiones simplemente observando la lista de datos.

En este apunte nos centraremos en una herramienta fundamental para resolver ese problema, la tabla de frecuencias. Este recurso permite organizar los datos para identificar cuántas veces aparece cada valor o grupo de valores, facilitando su lectura, análisis y comunicación.

Según el tipo de variable, las tablas pueden construirse listando cada valor individual (cuando se trata de datos discretos o categorías) o agrupando los datos en intervalos (cuando se trata de variables continuas o discretas con muchos valores posibles). En ambos casos, las tablas de frecuencia nos brindan una primera síntesis del conjunto de datos y nos permiten reconocer rápidamente regularidades, distribuciones y concentraciones.

En esta unidad trabajaremos con distintos tipos de tablas y aprenderemos a interpretar cada una de sus columnas: frecuencia absoluta, frecuencia relativa, frecuencias acumuladas y marcas de clase, entre otras. Comprender el significado y el uso de estos elementos es esencial para el análisis estadístico y para tomar decisiones fundamentadas a partir de la información recolectada.

Más adelante, aprenderemos a construir estas tablas utilizando el lenguaje R. Pero antes de usar esta herramienta, es importante comprender el proceso que subyace a su construcción. Esto te permitirá interpretar correctamente los resultados, detectar posibles errores y aplicar los conocimientos estadísticos con mayor criterio en contextos reales, especialmente en el campo de la programación y el desarrollo de software.

Elementos que componen una tabla de frecuencias

Una tabla de frecuencias organiza los valores observados de una variable y permite

responder preguntas como: ¿cuáles son los valores más comunes?, ¿cómo se acumulan?, ¿qué proporción representan?

A continuación, se describen los principales elementos que la componen:

- Valor observado de la variable (x_i)

Es cada uno de los valores distintos que toma la variable. En variables cualitativas, pueden ser categorías; en variables cuantitativas, son números.

Ejemplo: Si la variable es “cantidad de tickets resueltos por semana”, los valores posibles podrían ser: 0, 1, 2, 3, etc.

- Frecuencia absoluta (f_i)

Es el número de veces que se repite cada valor en el conjunto de datos.

Ejemplo: Si hubo 8 semanas con 2 tickets resueltos, entonces $f_2 = 8$

La suma de todas las frecuencias absolutas nos da el total de observaciones, que se denota con n

- Frecuencia absoluta acumulada (F_i)

Es la suma de todas las frecuencias absolutas desde el primer valor hasta el valor x_i . Nos indica cuántas observaciones tienen un valor menor o igual a ese.

Ejemplo: Si $F_3 = 22$, significa que 22 observaciones tienen 3 tickets o menos.

- Frecuencia relativa (f_r)

Es la proporción de observaciones que toma cada valor. Se calcula dividiendo la frecuencia absoluta entre el total de datos.

$$f_r = \frac{f_i}{n}$$

Ejemplo: Si $f_2 = 8$ y el total de datos es 40, entonces $f_r = 8/40 = 0,20$.

La suma de todas las frecuencias relativas siempre es igual a 1.

- Frecuencia relativa acumulada (F_r)

Indica la proporción acumulada de datos hasta cierto valor, y se puede obtener sumando las frecuencias relativas o dividiendo la frecuencia absoluta acumulada entre el total de observaciones.

$$F_r = \frac{F_i}{n} = \sum f_r$$

TABLAS DE FRECUENCIAS PARA DATOS NO AGRUPADOS

Cuando trabajamos con variables cualitativas o con variables cuantitativas discretas que no presentan muchos valores distintos, podemos organizar los datos en una tabla de frecuencias no agrupadas. Esto significa que cada fila de la tabla corresponde a un valor específico observado en el conjunto de datos.

Este tipo de tablas es ideal para visualizar rápidamente:

- qué valores se repiten más o menos,
- cómo se acumulan los datos,
- qué proporción representa cada valor dentro del total.

Ejemplo: Supongamos que en una empresa de desarrollo de software se lleva registro de la cantidad de tickets de soporte resueltos cada mes. Los datos correspondientes a los últimos 36 meses fueron los siguientes:

0 ; 1 ; 2 ; 1 ; 2 ; 0 ; 3 ; 2 ; 4 ; 0 ; 4 ; 2 ; 1 ; 0 ; 3 ; 0 ; 0 ; 3 ; 4 ; 2 ; 0 ; 1 ; 1 ; 3 ; 0 ; 1 ; 2 ; 1 ; 2 ; 3 ; 1 ; 0 ; 3 ; 4 ; 2 ; 1

Construir la tabla de frecuencias

Solución:

x_i	f_i	F_i	f_r	F_r
0	9	9	0,250	0,250
1	9	18	0,250	0,500
2	8	26	0,222	0,722
3	6	32	0,167	0,889
4	4	36	0,111	1,000
TOTAL	36		1,000	

¿Qué nos dice esta tabla?

- Los valores más frecuentes son 0 y 1 tickets resueltos por mes, con 9 apariciones cada uno.
- En el 50 % de los meses se resolvieron como máximo 1 ticket.
- Solo en el 11,1 % de los meses se resolvieron 4 tickets.
- El 72,2 % de los datos corresponden a valores iguales o inferiores a 2.

Claves para interpretar una tabla de frecuencias no agrupadas

- *Frecuencia absoluta* te dice cuántas veces ocurrió cada valor.
- *Frecuencia acumulada* indica cuántas observaciones son menores o iguales a un determinado valor.
- *Frecuencia relativa* muestra la importancia de cada valor respecto del total (es decir, su “peso” proporcional).
- *Frecuencia relativa acumulada* es la proporción acumulada hasta un cierto valor. Al multiplicar la frecuencia relativa por 100 obtenemos el correspondiente porcentaje.

TABLAS DE FRECUENCIAS PARA DATOS AGRUPADOS

Cuando los datos corresponden a variables cuantitativas continuas, o a variables discretas con muchos valores distintos, no resulta práctico listar cada uno por separado. En esos casos, organizamos los valores en intervalos, agrupando los datos según el rango en el que se encuentran. A esta organización la llamamos tabla de frecuencias para datos agrupados.

¿Por qué agrupar los datos?

Agrupar los datos facilita:

- Sintetizar conjuntos de observaciones muy extensos o diversos,
- Analizar distribuciones continuas,
- Generar representaciones gráficas claras (como histogramas o polígonos de frecuencias),

- Calcular medidas estadísticas en variables continuas que aprenderemos más adelante en la materia.

Elementos específicos de una tabla de frecuencias agrupadas.

Además de los elementos ya conocidos, una tabla agrupada incorpora:

Intervalos de clase: representan los rangos en los que se agrupan los datos (por ejemplo, de 0 a 4 kg, de 4 a 8 kg, etc.)

Marca de clase (x_i): es el punto medio de cada intervalo. Este valor representa al intervalo y se usa para representar todo el intervalo en los cálculos numéricos.

Amplitud de la clase: es la diferencia entre el límite superior y el límite inferior de cada intervalo.

Ejemplo: A continuación, se presentan los pesos en Kg. que han logrado descender un grupo de pacientes en tratamiento con una dieta determinada. El conjunto está compuesto por 20 observaciones ($n = 20$). Construir la tabla de frecuencias.

0,2	8,4	14,3	6,5	3,4	4,6	9,1	4,3	3,5	1,5
6,4	15,2	16,1	19,8	5,4	12,1	9,6	8,7	12,1	3,2

Solución:

Paso 1: Calcular la cantidad de clases

Para calcular la cantidad de clases se utiliza la “Regla de Sturges” que establece:

$$C = 1 + \log_2(n) = 1 + \frac{\log(n)}{\log(2)} \approx 1 + 3,322 \log n$$

$$C = 1 + 3,322 \log n = 1 + 3,322 \log 20 = 5,32.$$

Redondeamos y utilizaremos $C = 5$.

Paso 2: Calcular la amplitud

Calculamos la amplitud de cada clase:

$$A = \frac{x_{\max} - x_{\min}}{C} = \frac{19,8 - 0,2}{5} = 3,92 \Rightarrow \text{Se redondea } A = 4$$

Paso 3: Construir la tabla de frecuencias agrupadas

<i>Intervalo</i>	x_i	f_i	F_i	f_r	F_r
[0 , 4)	2	5	5	0,250	0,250
[4 , 8)	6	5	10	0,250	0,500
[8 , 12)	10	4	14	0,200	0,700
[12 , 16)	14	4	18	0,200	0,900
[16 , 20)	18	2	20	0,100	1,000
		20		1,000	

Convención sobre los intervalos

En esta tabla, todos los intervalos están cerrados en su límite inferior y abiertos en su límite superior, es decir, se expresan como:

$$[a , b) \text{ lo cual significa que } a \leq x < b$$

Esta convención permite que cada valor del conjunto de datos se ubique en un único intervalo, sin superposición.

Si el valor máximo del conjunto de datos coincidiera exactamente con el límite superior del último intervalo, en ese caso se suele cerrar el último intervalo a derecha, es decir, usar la notación $[a , b]$, para asegurarse de que dicho valor quede incluido en la tabla.

En este ejemplo, como el valor máximo es 19,8, y es estrictamente menor que 20, no es necesario modificar la convención general: el valor ya queda incluido en el último intervalo $[16 , 20)$.

A MODO DE CIERRE DE ESTE APUNTE

Organizar los datos mediante tablas de frecuencias es uno de los primeros pasos fundamentales del análisis estadístico. Estas tablas nos permiten estructurar la información, observar patrones, reconocer concentraciones y preparar el terreno para un análisis más profundo.

Comprender qué significan las frecuencias absolutas, relativas y acumuladas, y saber cómo construir tablas tanto para datos no agrupados como para datos agrupados, brinda una base sólida para interpretar la realidad a partir de datos.

En la materia avanzaremos hacia nuevas herramientas de análisis descriptivo con medidas numéricas que permiten resumir la información de forma aún más precisa y mediante representaciones gráficas.

Pero ninguna de esas herramientas tendría sentido sin una adecuada organización previa de los datos. Por eso, dominar estas técnicas no es solo un paso inicial, es una instancia clave para un proceso de análisis estadístico riguroso posterior.