

Razonamiento Aproximado. Ingeniería Informática. 2do Año.

Tema II: Estadística Descriptiva.

L/T. Walpole (pág. 1-34); L/T. Miller (5-44)

Estadística Descriptiva. Tablas univariantes.

Cuando coloquialmente se habla de estadística, se suele pensar en una relación de datos numéricos presentada de forma ordenada y sistemática. Esta idea es la consecuencia del concepto popular que existe sobre el término y que cada vez está más extendido debido a la influencia de nuestro entorno, ya que hoy día es casi imposible que cualquier medio de difusión, periódico, radio, televisión, etc, no nos aborde diariamente con cualquier tipo de información estadística sobre accidentes de tráfico, índices de crecimiento de población, turismo, tendencias políticas, resultados deportivos, etc.

Sólo cuando nos adentramos en un mundo más específico como es el campo de la investigación empezamos a percibir que la Estadística no sólo es algo más, sino que se convierte en la única herramienta que, hoy por hoy, permite dar luz y obtener resultados, y por tanto beneficios, en cualquier tipo de estudio, cuyos movimientos y relaciones, por su variabilidad intrínseca, no puedan ser abordadas desde la perspectiva de las leyes deterministas.

Podríamos, desde un punto de vista más amplio, definir la estadística como la ciencia que estudia cómo debe emplearse la información y cómo dar una guía de acción en situaciones prácticas que entrañan incertidumbre.

La **Estadística** se ocupa de los métodos y procedimientos para recoger, clasificar, resumir, hallar regularidades y analizar los *datos*, siempre y cuando la variabilidad e *incertidumbre* sea una causa intrínseca de los mismos; así como de realizar *inferencias* a partir de ellos, con la finalidad de ayudar a la toma de *decisiones* y en su caso formular *predicciones*.

Es posible clasificar la Estadística en *descriptiva*, cuando los resultados del análisis no pretenden ir más allá del conjunto de datos, e *inferencial* cuando el objetivo del estudio es derivar las conclusiones obtenidas a un conjunto de datos más amplio.

Estadística descriptiva: Describe, analiza y representa un grupo de datos utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos.

Estadística inferencial: Apoyándose en el cálculo de probabilidades y a partir de datos muestrales, efectúa estimaciones, decisiones, predicciones u otras generalizaciones sobre un conjunto mayor de datos.

Ante todo debemos tener claro algunos conceptos:

- **Población:** Colección de individuos o elementos que presentan el objeto de interés (seres vivos o inanimados).
- **Censo:** Observación y estudio de todos los elementos que componen la población.
- **Muestra:** Cualquier subconjunto de la población.
- **Muestreo:** Procedimiento mediante el cuál se extrae una muestra.
- **Característica:** Es el signo o detalle que interesa observar en la población.

Las datos que se obtienen en una investigación pueden ser, generalmente, de dos tipos:

- a) **Cuantitativos.** Son los caracteres que toman valores numéricos, se pueden medir.
Por ejemplo: Altura, peso, edad, tiempo, precipitación en una ciudad, petróleo extraído en un yacimiento
- b) **Cualitativos.** Describen una condición o cualidad, se utilizan palabras para referirse a ellos. Se denomina modalidad a cada una de las diferencias que pueden establecerse dentro de un carácter cualitativo.
Por ejemplo: el sexo, color de los ojos, estado civil.

Def: Se denomina variable estadística al conjunto de los valores que puede tomar un carácter estadístico. La variable se suele representar por letras latinas mayúsculas Ej.X y los valores que toma se representan por $x_1, x_2, x_3, \dots, x_n$.

A su vez las variables estadísticas cuantitativas se dividen en:

a) **Variables estadísticas continuas** son aquellas que pueden tomar, al menos teóricamente, cualquier valor de un intervalo fijado previamente.

Por ejemplo: La altura de un individuo. Tiempo en una carrera.

Son en general todas las magnitudes relacionadas con el espacio, tiempo, masa, etc.

b) **Variables estadísticas discretas** cuando pueden tomar valores aislados y concretos, normalmente números enteros.

Por ejemplo: Número de medallas. Número de hits, etc.

Nota Para trabajar con variables estadísticas cualitativas solemos convertirlas en variables discretas asignando a las distintas modalidades de cada carácter valores numéricos. Por ejemplo: Ante una pregunta con respuesta sí, no, indiferente se le puede asignar respectivamente valores -1, 1, 0.

Organización, representación y reducción de la información estadística.

Los datos estadísticos obtenidos de encuestas, experimentos o cualquier serie de mediciones con frecuencia son tan numerosos que prácticamente resultan inútiles a menos que se condensen o se reduzcan en forma más adecuada.

Tablas de Distribución de Frecuencias.

- Elementos que la forman:

Clases (x_i): Valores diferentes de la variable .

Frecuencia absoluta (n_i): Número de veces que se repite el i-ésimo elemento de la variable (Clases).

Frecuencia relativa (f_i): Frecuencia absoluta de la clase dividida por el total de elementos.

Frecuencia absoluta acumulada (N_i): Frecuencia absoluta de la clase más las n_i de las clases anteriores.

Frecuencia relativa acumulada (F_i): Frecuencia relativa de la clase más las f_i de las clases anteriores.

Ejemplo 1: Los siguientes datos recogen el número de ausencias en el último semestre de los 100 trabajadores de un centro de trabajo.

2	0	3	1	3	5	3	1	5	2
3	0	3	1	2	5	1	5	4	3
2	0	4	1	2	5	3	5	3	3
2	0	4	4	2	5	4	5	2	4
3	5	4	4	2	4	3	4	1	4
3	5	4	4	1	3	2	3	2	4
3	5	4	4	1	4	2	2	3	5
4	5	1	4	1	3	2	1	4	5
4	3	1	3	1	4	1	3	5	1
0	3	1	3	5	2	1	4	20	1

Datos Primarios.

Tabla de Distribución de Frecuencia.

Variable: # de ausencias de cada trabajador en el semestre.

Clases.	n_i	f_i	N_i	F_i
0	5	0,05	5	0,05
1	18	0,18	23	0,23
2	15	0,15	38	0,38
3	22	0,22	60	0,60
4	23	0,23	83	0,83
5	16	0,16	99	0,99
20	1	0,01	100	1
Total	100	1		

Gráficos:

- **Diagramas de Barras:** Se representa en el eje de ordenadas las Clases y en el eje de las abscisas las frecuencias absolutas o bien, las frecuencias relativas.
- **Diagramas de sectores: (también llamados 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

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

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

Tabla de Distribución de Frecuencias para Variables con muchos valores diferentes. (Datos Agrupados)

Clases: Se representan a partir de intervalos.

Para formar estos intervalos seguiremos los siguientes pasos:

- Decidir el número de clases o intervalos (k) (entre 5 y 20).
- Determinar el Recorrido (r) Recorrido = Max – Min.
- Calcular la amplitud de las clases (h) $h = \frac{r}{k}$. (Se aproxima por exceso)
- Determinar el valor inicial para el intervalo, tomando el valor mínimo o el más cercano de los valores de la variable

Ejemplo 2: Tenemos la talla de los alumnos de un grupo de Ingeniería:

168 160 168 175 175 160 165 154 163 165
168 168 158 150 160 161 162 166 163 159
178 169 158 163 171 170 165 150 167 164
162 165 163 156 174 165 173 172 168 168

Como podemos apreciar las estaturas poseen muchos valores diferentes por lo que acudimos los intervalos:

- Determinamos $k = 6$
- Calculamos el recorrido $r = 178 - 150 = 28$
- Calculamos la amplitud: $h = 28/6 = 4.66 \approx 5$
- Valor inicial 150

Tabla de distribución de frecuencia:

Intervalos	Frec. Absoluta. n_i	Frec. Relativa f_i	Frec. Absoluta Acumulada (N_i)	Frec. Relativa Acumulada (F_i)
150 < 155	3	0.075	3	0.075
155 < 160	4	0.1	7	0.175
160 < 165	11	0.275	18	0.45
165 < 170	14	0.35	32	0.8
170 < 175	5	0.125	37	0.925
175 < 180	3	0.075	40	1
Total	40	1		

Gráficos:

- **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.
- **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.

Estudio Independiente:

L/T. Miller pág. 12. (Estudiar diagrama de Pareto y Diagrama de puntos)

Ejercicios L/T. Miller pág. 22

Ejercicios Guía Tema II-Estadística Descriptiva.

Medidas Descriptivas.

Las Tablas de distribución de frecuencias, los gráficos e histogramas y otros resumen un conjunto de datos de manera analítica y gráfica, así que es posible discernir visualmente el patrón general de variación. Ahora se desarrollarán medidas numéricas para describir un conjunto de datos.

Medidas de tendencia central

Se llaman “medidas de tendencia central” a valores numéricos que describen un conjunto de datos primarios y que generalmente ocupan una posición “central” dentro de este; en el sentido de que una parte de los valores están situados por encima y otra parte por debajo de ellos, no son valores “extremos”, sino por el contrario valores “típicos” o “característicos”

Entre ellas se encuentran:

Media aritmética (\bar{X})

Es el promedio o media del conjunto de valores.
$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n}$$

- Está situada entre el valor mínimo y máximo del conjunto.
- Resulta muy afectada por valores extremos.
- Si n_1 datos tienen media \bar{X}_1 , n_2 datos tienen media \bar{X}_2 , ..., n_l datos tienen media \bar{X}_l , la media de los

$$n = n_1 + n_2 + \dots + n_l \text{ datos es } \bar{X} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2 + \dots + n_l \bar{X}_l}{n} = \frac{\sum_{i=1}^l n_i \bar{X}_i}{\sum_{i=1}^l n_i}$$

Ejemplo 3:

Dados los siguientes valores de la variable X:

X_i	987,54	993,70	989,48	992,78	993,99	986,35	985,92	986,50	989,54	991,24	988,61	992,41
----------------------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

La media aritmética es:

$$\bar{X} = \frac{987,54 + 993,7 + \dots + 992,41}{12} = \frac{11878,06}{12} = 989,8383$$

Mediana

Es el valor “central” de los datos ordenados.

- Ordenar los datos primarios ascendentemente (o descendentemente)
- Determinar el valor central o los valores centrales.
- Calcular la mediana.

Si **n** es impar $Me = X_{\frac{n+1}{2}}$

Si **n** es par $Me = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$

- Está situada entre el valor mínimo y máximo del conjunto.
- Divide a los datos primarios en dos partes iguales (50% por encima, 50 % por debajo)
- No resulta afectada por valores extremos.
- No es necesario ordenar **todos** los datos para el cálculo, solo hasta los valores centrales.

Ejemplo:

En la muestra seleccionada: La mediana sería el valor medio del 6º y el 7º datos, después de ordenar estos

X_i	985,92	986,35	986,5	987,54	988,61	989,48	989,54	991,24	992,41	992,78	993,7	993,99
----------------------	--------	--------	-------	--------	--------	---------------	---------------	--------	--------	--------	-------	--------

$$Me = \frac{989,48 + 989,54}{2} = 989,51$$

Moda

No es propiamente una “medida de tendencia central”, puesto que puede no ocupar una posición “central”, sería más apropiado clasificarla entre las llamadas “medidas de posición o localización”, valores también característicos del conjunto de datos primarios, pero no “centrales”

M_o: es el valor más frecuente dentro del conjunto de datos primarios.

Puede no existir y si existe puede no ser única.

Ejemplos:

1) 3 4 1 1 2 2 2 5

M_o = 2

2) 4,4 4 3,7 3,6 4,1 4,2

No existe moda

3) 7,1 4,1 3 5,5 5,5 2,7 5,5 2,7 2,7 4 4

M_o = 2,7 ; 5,5

Otras medidas de posición

Cuartiles: 3 valores que dividen al conjunto de datos primarios en 4 partes iguales (25% de los datos se encuentran entre cada dos de ellos, Q₁, Q₂ = Me, Q₃)

Deciles: 9 valores que dividen al conjunto de valores en 10 partes iguales (10% de los datos se encuentran entre cada dos de ellos, D₁...D₉)

Percentiles: 99 valores que dividen al conjunto de datos en 100 partes iguales (1% de los datos se encuentran entre cada dos de ellos, P₁, P₂, ..., P₉₉)

En el caso de una v.a los percentiles se pueden determinar en base a la función de distribución de la variable, p. ejemplo, $F_X(P_{95}) = P(X < P_{95}) = 0,95$ $Q_1 = P_{25}$; $Q_3 = P_{75}$, etc

Medidas de dispersión

Veamos el ejemplo siguiente, que contiene dos conjuntos de datos primarios diferentes.

A	2	3	4	5	6	7	8
B	4	4	4,5	5	5,5	6	6

$\bar{X}_A = \bar{X}_B = 5$
 $Me_A = Me_B = 5$
 Mo_A no existe
 $Mo_B = 4; 6$

Puede observarse que existen diferencias importantes entre los dos conjuntos, a pesar de que la media aritmética y la mediana tienen el mismo valor, la principal es que los datos “difieren más” en el primer conjunto, están “más dispersos” alrededor del valor medio en ese conjunto, la media es “más representativa” del conjunto de datos en el segundo caso. ¿Cómo medir esa característica del conjunto de datos?

Parece claro que en el indicador a utilizar deben aparecer las diferencias entre cada valor y la media o la mediana, de modo que mientras más grandes sean estas el indicador tenga un valor mayor. Vamos a trabajar con la media en lugar de con la mediana, lo cual se justifica más adelante.

Podría considerarse $\frac{\sum_{i=1}^n (X_i - \bar{X})}{n}$, solo que... la suma que aparece en el numerador SIEMPRE tiene valor cero! (¿por qué?)

$\frac{\sum_{i=1}^n (X_i - \bar{X})}{n} = \frac{\sum_{i=1}^n X_i - n\bar{X}}{n} = \frac{\sum_{i=1}^n X_i - n \frac{\sum_{i=1}^n X_i}{n}}{n} = \frac{0}{n} = 0$. Existen valores por encima de la media que originan diferencias positivas, otros por debajo que originan diferencias negativas, estas se compensan unas con otras para producir una suma nula.

Esta situación puede resolverse evitando la aparición de diferencias negativas:

Varianza

Si los datos primarios pertenecen a una muestra:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} = \frac{1}{n - 1} \left[\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n} \right]$$

En el denominador aparece **n-1** en lugar de **n** por razones que veremos después.

Si los datos corresponden a toda la población, se representa por σ^2 y en el denominador se utiliza **n** en lugar de **n-1**, o sea,

$$\sigma^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} = \frac{1}{n} \left[\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n} \right] = \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2$$

Ejemplo (Conjunto A)

X_i	X_i²
2	4
3	9
4	16
5	25
6	36
7	49
8	64
$\sum X_i = 35$	$\sum X_i^2 = 203$

$$S^2 = \frac{1}{n - 1} \left[\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n} \right] = \frac{1}{6} \left[203 - \frac{35^2}{7} \right]$$

$$= \frac{1}{6} [203 - 175] = \frac{28}{6} = 4$$

Desviación media

$$DM = \frac{\sum_{i=1}^n |X_i - \bar{X}|}{n}$$

Es más usada la varianza, pues posee propiedades matemáticas “mejores” a las cuales haremos referencia más adelante.

Desviación típica o estándar

$S = \sqrt{S^2}$ A partir de la varianza ofrece una medida de dispersión expresada en las mismas unidades de medida que los datos originales (obsérvese que la varianza queda expresada en unidades cuadradas)

Recorrido

$R = \text{Max} - \text{Min}$

Es de muy fácil cálculo, pero solo toma en cuenta los valores extremos del conjunto de datos primarios

Amplitud intercuartílica (IQR, interquartile range)

$$IQR = Q_3 - Q_1 = P_{75} - P_{25}$$

Es relativamente de fácil cálculo también, pero no toma en cuenta los valores extremos y puede no cambiar al cambiar valores entre los cuartiles

Coefficiente de variación

Supongamos dos conjuntos A y B de datos, con los siguientes valores de media y varianza:

$$\bar{X}_A = 100; S_A^2 = 36$$

$$\bar{X}_B = 6; S_B^2 = 9$$

¿Cuál de los dos conjuntos presenta una mayor variabilidad?

¿En cuál de los dos conjuntos la media es menos típica o representativa del conjunto?

Si se observan los valores de las varianzas y los de las desviaciones típicas correspondientes podríamos responder que en el conjunto A, porque tiene mayor varianza.

Sin embargo, EN RELACIÓN AL VALOR MEDIO DE LOS DATOS, DE MODO RELATIVO, la varianza del conjunto A es menor

$$CV = \frac{S}{\bar{X}} \cdot 100\%$$

$$\bar{X}_A = 100; S_A^2 = 36; S_A = 6; CV_A = 6\%$$

$$\bar{X}_B = 6; S_B^2 = 9; S_B = 3; CV_B = 50\%$$

Cuando se comparan dos conjuntos de datos de valores medios similares, esta comparación puede hacerse tomando como referencia el valor de la varianza o la desviación típica; cuando se trata de conjuntos con muy diferentes valores medios o en diferentes unidades de medida debe hacerse utilizando el coeficiente de variación.

Veamos un ejemplo integrador:

Consideremos el conjunto de datos correspondiente al contenido (mL) de 40 envases de detergente líquido producidos en un turno de una empresa de la industria ligera en una semana.

985,74	993,7	986,40	995,49	990,10	988,41	988,47	990,94	990,26	991,22
989,58	989,37	990,97	989,22	987,54	993,61	989,25	988,83	994,67	990,92
993,7	991,85	990,37	989,14	990,54	987,24	986,69	989,39	983,90	989,64
991,93	985,83	991,61	989,74	989,65	990,71	994,68	987,88	996,72	993,06

Tabla de Distribución de frecuencia:

Construcción de los intervalos de clases:

1. Determinar el máximo y el mínimo de los datos y a partir de ellos el recorrido $R = \text{Max} - \text{Min}$.
Aquí $R = 996,72 - 983,90 = 12,82$
2. Dividir R entre la cantidad de intervalos deseados (usualmente entre 5 y 15, dependiendo de la cantidad de datos). En este caso tomemos 5 intervalos ($k = 5$), de modo que la amplitud de cada intervalo C sería
$$C = \frac{R}{k} = \frac{12,82}{5} = 2,564 \approx 3 \text{ (redondeo por exceso)}$$
3. Determinar valor inicial (983)
4. Formar los intervalos, en este caso quedarían así:

Intervalos de clase
[983; 986[
[986; 989[
[989; 992[
[992; 995[
[995; 998[

- Frecuencia absoluta de la clase (n_i): Es el número de datos perteneciente a la clase i . Un valor igual al extremo derecho del intervalo se incluye en la clase siguiente.

Intervalos de clase	n_i
[983; 986[3
[986; 989[8
[989; 992[21
[992; 995[6
[995; 998[2
	$n = 40$

- Frecuencia relativa de la clase (f_i):

Intervalos de clase	n_i	f_i
[983; 986[3	$3/40=0,075$ (7,5%)
[986; 989[8	$8/40=0,20$ (20%)
[989; 992[21	$21/40=0,525$ (52,5%)
[992; 995[6	$6/40=0,15$ (15%)
[995; 998[2	$2/40=0,05$ (5%)
	$n = 40$	1 (100%)

- Frecuencia acumulada de la clase (N_i): $N_i = \sum_{j=1}^i n_j$

Intervalos de clase	n_i	f_i	N_i
[983; 986[3	3/40=0,075 (7,5%)	3
[986; 989[8	8/40=0,20 (20%)	11
[989; 992[21	21/40=0,525 (52,5%)	32
[992; 995[6	6/40=0,15 (15%)	38
[995; 998[2	2/40=0,05 (5%)	40
	n = 40	1 (100%)	

- Frecuencia acumulada relativa (F_i): $F_i = \frac{N_i}{n} = \sum_{j=1}^i f_j$

Intervalos de clase	n_i	f_i	N_i	F_i
[983; 986[3	3/40=0,075 (7,5%)	3	3/40=0,075 (7,5%)
[986; 989[8	8/40=0,20 (20%)	11	11/40=0,275 (27,5%)
[989; 992[21	21/40=0,525 (52,5%)	32	32/40=0,8 (80%)
[992; 995[6	6/40=0,15 (15%)	38	38/40=0,95 (95%)
[995; 998[2	2/40=0,05 (5%)	40	40/40=1 (100%)
	n = 40	1 (100%)		

- Marca de clase (m_i). Es el punto medio del intervalo. Se asume que es el valor medio de los datos contenidos en el intervalo.

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

Intervalos de clase	n_i	f_i	N_i	F_i	m_i
[983; 986[3	3/40=0,075 (7,5%)	3	3/40=0,075 (7,5%)	984,5
[986; 989[8	8/40=0,20 (20%)	11	11/40=0,275 (27,5%)	987,5
[989; 992[21	21/40=0,525 (52,5%)	32	32/40=0,8 (80%)	990,5
[992; 995[6	6/40=0,15 (15%)	38	38/40=0,95 (95%)	993,5
[995; 998[2	2/40=0,05 (5%)	40	40/40=1 (100%)	996,5
	n = 40	1 (100%)			

Pueden obtenerse a partir de la primera, sumando sucesivamente la amplitud de intervalo (3)

Salida del Software STAT GRAPHICS

Para obtener una tabla de frecuencias en este software, primeramente se llena la hoja de datos y luego se va a la opción Describe\Numeric Data\One-Variable Analysis, cuyo resultado sería:

Frequency Tabulation for Contenido en ml

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		983,0		0	0,0000	0	0,0000
1	983,0	986,0	984,5	3	0,0750	3	0,0750
2	986,0	989,0	987,5	8	0,2000	11	0,2750
3	989,0	992,0	990,5	21	0,5250	32	0,8000
4	992,0	995,0	993,5	6	0,1500	38	0,9500
5	995,0	998,0	996,5	2	0,0500	40	1,0000
above	998,0			0	0,0000	40	1,0000

Mean = 990,224 Standard deviation = 2,80077

Propiedades e interpretación

Como se observa en el ejemplo y se deduce a partir de las definiciones se tienen como propiedades básicas:

- La suma de las frecuencias absolutas coincide con la cantidad de datos **n**.
- La suma de las frecuencias relativas es igual a 1 o 100% (salvo errores de redondeo)
- La frecuencia acumulada de la primera clase coincide con la absoluta y la última con **n**.
- La frecuencia acumulada de la última clase coincide con **n**.

Es importante la interpretación de los elementos de la tabla, en los dos sentidos, es decir, describir los datos a partir de la tabla como interpretar valores específicos de esta.

Por ejemplo:

a) Interpretar n_2 , f_3 y N_4

N_2 : 8 envases tienen un contenido entre 983 mL y 986 mL

f_3 : el 52,5 % de los envases tienen un contenido entre 989 mL y 992 mL

N_4 : 38 envases tienen un contenido menor de 995 mL

b) ¿Qué parte de los envases tienen menos de 992 mL? R/ El 80%

¿Cuántos envases tienen al menos 995 mL? R/ 2

c) Suponiendo que la tabla presenta la siguiente forma, completarla

Intervalos de clase	n_i	f_i	N_i	F_i	m_i
[983; 986[3	$3/40=0,075$ (7,5%)	3		
[986; 989[11		
[989; 992[21	$21/40=0,525$ (52,5%)	32		
[992; 995[15%			
[995; 998[
	$n = 40$	1 (100%)			

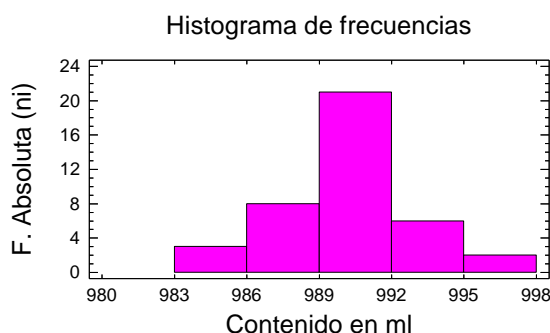
R/

$$n_2 = N_2 - n_1 = 11 - 3 = 8 \quad n_4 = h_4 \cdot n = 0,15 \cdot 40 = 6 \quad n_5 = n - \sum_{i=1}^4 n_i = 40 - 38 = 2$$

Representación gráfica

Histograma de frecuencias

Consta de una serie de rectángulos, uno por cada intervalo. La base del rectángulo es el propio intervalo y la altura el valor de la frecuencia absoluta.



Cálculo de las medidas estadísticas:

$$\text{Media: } \bar{X} = \frac{X_1 + X_2 + \dots + X_{40}}{40} = 990,22$$

$$\text{Mediana: } Me = \frac{X_{20} + X_{21}}{2} = 989,92$$

$$\text{Moda: } Mo = 993,7$$

$$\text{Varianza: } S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = 7,84$$

$$\text{Desviación típica: } S = \sqrt{S^2} = 2,8$$

Estudio Independiente:

Ejercicios L/T. Miller pág. 36. (Estudiar diagrama de Pareto y Diagrama de puntos)

Ejercicios L/T. Wallpole pág. 30

Ejercicios Guía Tema II-Estadística Descriptiva.