



Facultad de Ciencias Económicas

Universidad Nacional de Jujuy

ESTADÍSTICA
Año: 2022

ESTADÍSTICA DESCRIPTIVA





Problema 8

Con el fin de estimar la cantidad de agua que se necesitara para abastecer una zona rural del norte de la provincia, se registró cuantos miles de litros de agua consume una muestra de familias en el último año. Los resultados son los siguientes:

42	72	81	111	62	92	75	105	55	85
44	122	102	69	49	50	30	90	70	69
59	74	71	86	86	101	53	68	64	49



a) Definir la variable y clasificarla e indicar su escala de medición.

X_i : Cantidad de agua (en miles de litros) que consume una familia de una zona rural del norte de la provincia en el último año.

Clasificación: Variable Cuantitativa Continua

Escala de Medicion: De Razón.

b) Agrupar los datos en intervalos de clase. Graficar el histograma de frecuencias absolutas e interpretar.

REGLA PRÁCTICA DE STURGES (INTERVALOS DE CLASE)

Siguiendo los pasos para construir la distribución de frecuencias:

1. Rango: $r = x_{max} - x_{min} = 122 - 30 = 92$ (amplitud total de la serie)
2. Nº de clases: $k = 1 + 3.3 \log 30 = 5.87 \cong 6$
3. Amplitud de clase : $a = 92/6 = 15,33 \cong 15$
4. Para simplificar la construcción de los intervalos se tomará $a = 15$

Aclaración

A los fines prácticos para este problema, se decide considerar una amplitud de 15 (nº entero y múltiplo de 5) y con ello $k=7$

Limite Inferior (Li) = Valor mas pequeño de la muestra ➡ 30

Limite Superior (Ls) = sera igual al Limite Inferior mas el ancho del intervalo= (Li + a)

c) Realizar una tabla de frecuencias relativas porcentuales y de frecuencias acumuladas. Interpretar dos frecuencias de cada clase.

Tabla de Distribución de Frecuencias

L	Ls	Xm	fi	ri	ri%	Fi ↑	Ri ↑	Ri% mayor	Fi ↓	Ri ↓	Ri% menor	fi.Xm	Xm - X̄	(Xm - X̄)²	(Xm - X̄)². fi
						mayor	mayor		menor	menor					
[30	45)	37,5	3	0,1000	10,00%	30	1,00	100,00%	3	0,10	10,00%	112,50	-35,5000	1.260,25	3.780,75
[45	60)	52,5	6	0,2000	20,00%	27	0,90	90,00%	9	0,30	30,00%	315,00	-20,5000	420,25	2.521,50
[60	75)	67,5	9	0,3000	30,00%	21	0,70	70,00%	18	0,60	60,00%	607,50	-5,5000	30,25	272,25
[75	90)	82,5	5	0,1667	16,67%	12	0,40	40,00%	23	0,77	76,67%	412,50	9,5000	90,25	451,25
[90	105)	97,5	4	0,1333	13,33%	7	0,23	23,33%	27	0,90	90,00%	390,00	24,5000	600,25	2.401,00
[105	120)	112,5	2	0,0667	6,67%	3	0,10	10,00%	29	0,97	96,67%	225,00	39,5000	1.560,25	3.120,50
[120	135)	127,5	1	0,0333	3,33%	1	0,03	3,33%	30	1	100,00%	127,50	54,5000	2970,25	2970,25
			30	1	100,00%							2190,00			15.517,50



Interpretacion

Frecuencias Absolutas:

Seis (6) familias de una zona rural del norte de la provincia consumieron durante el último año 45 mil litros de agua o mas y menos de 60 mil litros.

Solo una familia de una zona rural del norte de la provincia consumió 120 mil litros de agua o mas pero hasta 135 mil litros de agua, durante el ultimo año.

Frecuencias Relativas:

0.13 es la proporción de familias de una zona del norte de la provincia que consumieron 90 mil litros de agua o mas y menos de 105 mil litros en el último año.

La proporción de 0,1667 familias de una zona del norte de la provincia consumieron en el último años 75 mil litros de agua o mas y menos de 90 mil litros.

Frecuencia Relativa Porcentual:

El 10,00% de las familias de una zona rural del norte de la provincia consumió durante el ultimo año 30 mil litros de agua o mas y menos de 45 mil litros.

El consumo de agua del 30,00% de las familias de una zona rural del norte de la provincia, en el ultimo año, fue de 60 mil litros o mas y menos de 75 mil litros.



Frecuencias acumuladas: (menor que)

9 familias de una zona rural del norte de la provincia consumieron en el ultimo año menos de 60 mil litros de agua.

El 90% de las familias de una zona rural del norte de la provincia tuvieron en el ultimo año un consumo de agua menor a 105 mil litros de agua .

Frecuencias acumuladas: (mayor que)

27 familias de una zona rural del norte de la provincia consumieron durante el ultimo año 45 mil litros de agua o mas.

El 23,33% de las familias de una zona rural del norte de la provincia consumieron en el último año 90 mil litros o mas de agua.

d) Calcular la media, la mediana y la moda e interpretarlas.

Media:
$$\bar{x} = \frac{\sum_{i=1}^k x_{mi} * f_i}{n}$$

Para encontrar la media debemos incorporar a la tabla de frecuencias dos columnas, una que tenga la marca de clase y otra que tenga el producto entre la marca de clase y la frecuencia absoluta de esa clase. La suma de los valores de esta última columna nos dará el numerador de la media.

Intervalo de clase	Xm	fi	fi.Xm
[30 – 45)	37,5	3	112,50
[45 – 60)	52,5	6	315,00
[60 – 75)	67,5	9	607,50
[75 – 90)	82,5	5	412,50
[90 – 105)	97,5	4	390,00
[105 – 120)	112,5	2	225,00
[120 – 135)	127,5	1	127,50
		30	2190,00

$$\bar{x} = \frac{2190,00}{30} = 73$$

Interpretación: *El consumo promedio de agua durante el ultimo año de las familias de una zona rural del norte de la provincia, es de 73 mil litros .*



$$\text{Mediana: } Me = L_i + \frac{\frac{n}{2} - F_{i(i-1)}}{f_i} \cdot a$$

1. $\frac{n}{2} = 15$, luego la primera frecuencia acumulada “menor que” que supera ese valor es $F_i = 18$
2. $L_i = 60$ es el limite inferior del intervalo medial y $L_s = 75$ es el límite superior del intervalo.
3. $f_i = 9$ es la frecuencia absoluta del intervalo medial.
4. $a_i = 15$ es la amplitud del intervalo.
5. $F_{i-1} = 9$ es la frecuencia acumulada del intervalo anterior al que contiene a la mediana.

$$Me = 60 + \frac{15 - 9}{9} * 15 = 70$$

Intervalo de clase	f_i	F_i	
[30 45)	3	3	
[45 60)	6	9	
[60 75)	9	18	> 15
[75 90)	5	23	
[90 105)	4	27	
[105 120)	2	29	
[120 135)	1	30	
Total	30		

Interpretación: El 50% de las familias de una zona rural del norte de la provincia consumió en el ultimo año 70 mil litros o menos de agua , y el 50% restante 70 mil litros o mas.



Moda: $Mo = L_i + \frac{d_1}{d_1 + d_2} \cdot a$

1. El intervalo modal es el que tiene la mayor frecuencia absoluta ($f_i=9$). Es el intervalo [60 - 75).
2. $Li= 60$ es el limite inferior del intervalo modal.
3. $d1= 3$ es la diferencia entre la frecuencia absoluta del intervalo modal y la frecuencia abs. del intervalo premodal.
4. $d2= 4$ es la diferencia entre la frecuencia absoluta del intervalo modal y la frecuencia abs. del intervalo posmodal.
5. $ai= 15$ es la amplitud del intervalo.

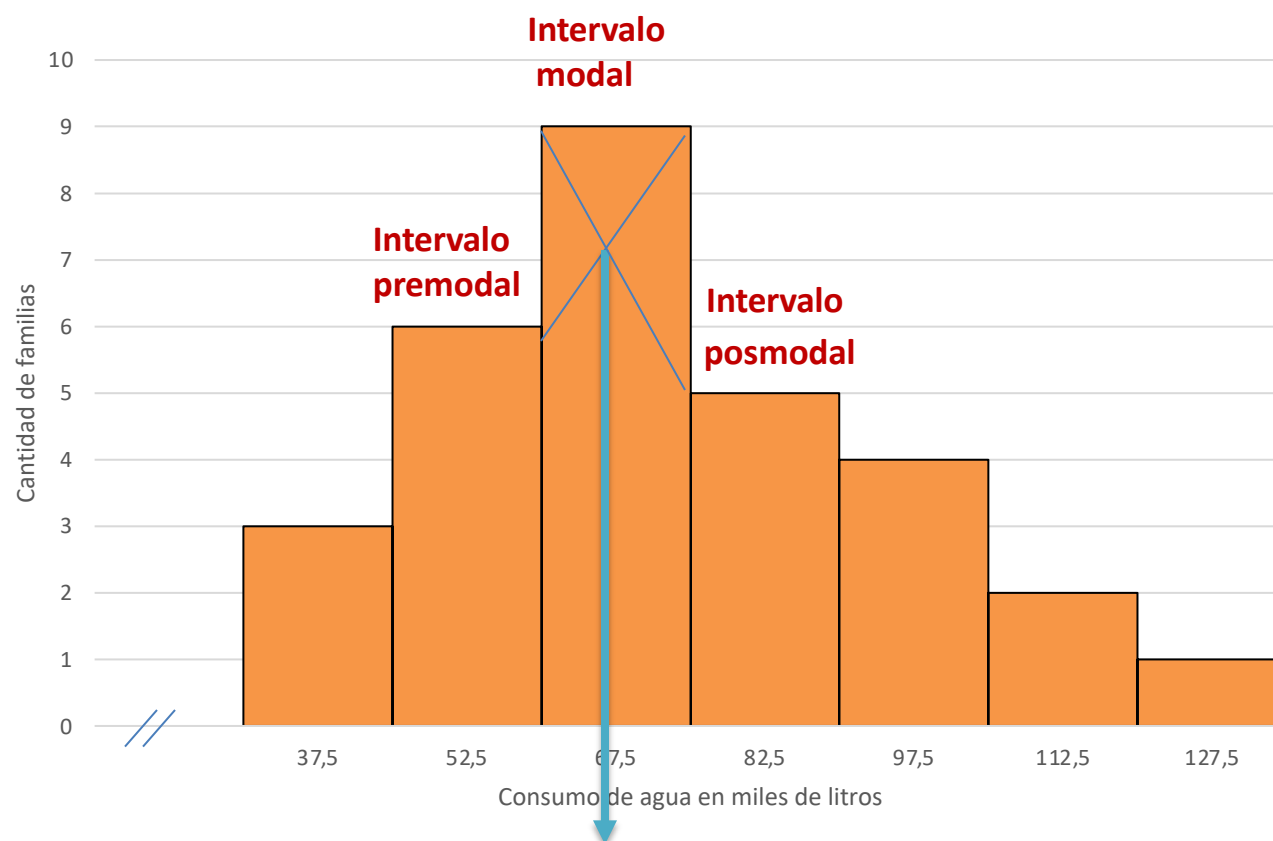
Intervalo de clase	f_i
[30 – 45)	3
[45 – 60)	6 Frecuencia abs. premodal
[60 – 75)	9 Frecuencia abs. modal
[75 – 90)	5 Frecuencia abs. posmodal
[90 – 105)	4
[105 – 120)	2
[120 – 135)	1
Total	30

$$Mo = 60 + \frac{3}{3 + 4} \cdot 15 = 66,4286$$

Interpretación: El consumo de agua mas frecuente *entre las familias de una zona rural del norte de la provincia, en el último año, es de 66,4286 mil litros de agua.*

e) Encontrar gráficamente la moda e interpretar.

Histograma de Frecuencias absolutas del consumo de agua en miles de litros que tuvo una muestra de familias de una zona rural del norte de la provincia.



Mo: 66,43

- Como todos los intervalos tienen el mismo ancho, el rectángulo de mayor altura corresponde al intervalo modal.
- Se unen los vértices superiores del intervalo modal con los vértices superiores del intervalo premodal y del intervalo posmodal.
- Del punto resultante de la intersección de los segmentos se traza una línea vertical hasta que corta en el eje de las abscisas.
- El valor del eje es el valor que corresponde a la moda.

f) Calcular e interpretar el 5° decil y el 75° percentil. ¿Con qué valores coincide?

Cálculo de 5° decil:
$$D_5 = L_i + \frac{\frac{5n}{10} - F_{(i-1)}}{f_{(i)}} \cdot a_i$$

1. $\frac{5n}{10} = \frac{30}{2} = 15$, luego la primera frecuencia acumulada “menor que” que supera ese valor es $F_i = 18$
2. $L_i = 60$ es el limite inferior del intervalo del quinto decil
3. $f_i = 9$ es la frecuencia absoluta del intervalo
4. $a_i = 15$ es la amplitud del intervalo.
5. $F_{(i-1)} = 9$ es la frecuencia acumulada del intervalo anterior al que contiene al quinto decil.

$$D_5 = Me = 60 + \frac{15 - 9}{9} \cdot 15 = 70$$

Intervalo de clase		f_i	F_i
[30	45)	3	3
[45	60)	6	9
[60	75)	9	18
[75	90)	5	23
[90	105)	4	27
[105	120)	2	29
[120	135)	1	30
Total		30	



Interpretación: *El 50% de las familias de una zona rural del norte de la provincia consumieron en el ultimo año 70 mil litros de agua o menos, y el 50% restante consumieron 70 mil litros o más.*



Podemos observar que el valor del quinto decil coincide con la mediana, por lo que su interpretación es la misma.

Cálculo de 75° percentil:
$$P_{75} = L_i + \frac{\frac{75n}{100} - F_{(i-1)}}{f_{(i)}} \cdot a_i$$

1. $\frac{75n}{100} = \frac{3n}{4} = 22,5$, luego la primera frecuencia acumulada "menor que" que supera ese valor es $F_i = 26$.
2. $L_i = 75$ es el limite inferior del intervalo del 75° percentil.
3. $f_i = 5$ es la frecuencia absoluta del intervalo del 75° percentil.
4. $a_i = 15$ es la amplitud del intervalo.
5. $F_{(i-1)} = 18$ es la frecuencia acumulada del intervalo anterior al que contiene a la del 75° percentil.

$$P_{75} = Q_3 = 75 + \frac{22,5 - 18}{5} \cdot 15 = 88,50$$

Li	Ls	fi	Fi menor
[30	45)	3	3
[45	60)	6	9
[60	75)	9	18
[75	90)	5	23
[90	105)	4	27
[105	120)	2	29
[120	135)	1	30
Total		30	

f) Calcular e interpretar el 5° decil y el 75° percentil. ¿Con qué valores coincide?

Cálculo de 75° percentil:
$$P_{75} = L_i + \frac{\frac{75n}{100} - F_{(i-1)}}{f_{(i)}} \cdot a_i$$

Como $\frac{75n}{100} = \frac{3n}{4}$, El valor del 75° percentil coincide con el del 3° cuartil, por lo que su interpretación es la misma:

$$P_{75} = Q_3 = 75 + \frac{22,5 - 18}{5} \cdot 15 = 88,50$$

Interpretación: El 75% de las familias de una zona rural del norte de la provincia tuvo un consumo de agua en el ultimo año de 88,50 mil litros de agua o menos, y el 25% restante tuvo consumos de 88,50 mil litros o mas de agua.

g) Calcular el desvío estándar e interpretarlo.

Desvío Estándar:
$$S = \sqrt{\frac{\sum_{i=1}^k (x_m - \bar{x})^2 \cdot f(i)}{n-1}}$$



$$S = \sqrt{\frac{15517,50}{29}} = 23,13$$

Intervalo de clase	Xm	fi	Xm - $\bar{X}$	(Xm - $\bar{X}$) ²	(Xm - $\bar{X}$) ² · fi
[30 – 45)	37,5	3	-35,5000	1.260,25	3.780,75
[45 – 60)	52,5	6	-20,5000	420,25	2.521,50
[60 – 75)	67,5	9	-5,5000	30,25	272,25
[75 – 90)	82,5	5	9,5000	90,25	451,25
[90 – 105)	97,5	4	24,5000	600,25	2.401,00
[105 – 120)	112,5	2	39,5000	1.560,25	3.120,50
[120 – 135)	127,5	1	54,5000	2970,25	2970,25
Total		30			15.517,50

Interpretación

La dispersión promedio en el consumo de agua que realizó una muestra de familias de una zona rural del norte de la provincia, con respecto a la media es de 23,13 mil litros de agua .



h) Interpretar el rango y el rango intercuartílico.

1. Rango: $r = x_{max} - x_{min}$

$$r = 122 - 30 = 92$$

El recorrido total de la variable cantidad de agua que consumen en el ultimo una familia de una zona rural del norte de la provincia es de 92 mil litros de agua.

2. Rango intercuartílico:

$$RI = Q_3 - Q_1$$

$$Q_3 = L_i + \frac{\frac{3n}{4} - F_{(i-1)}}{f_{(i)}} \cdot a_i = 88,50 \quad (\text{calculado en el inciso f})$$

$$Q_1 = L_i + \frac{\frac{1n}{4} - F_{(i-1)}}{f_{(i)}} \cdot a_i = 45 + \frac{7,5 - 3}{6} \cdot 15 = 56,25$$

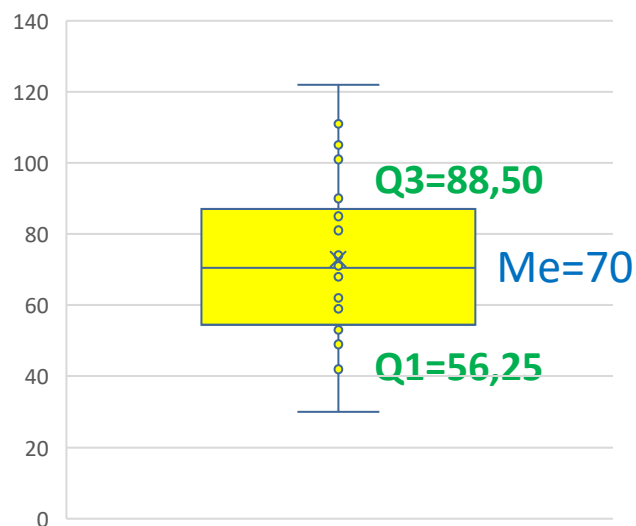
$$RI = 88,50 - 56,25 = 32,25$$

Cálculos:
 $\frac{1 \cdot n}{4} = \frac{30}{4} = 7,5$

Interpretación: La variabilidad del 50% central del consumo de agua es de 32,25 mil litros de agua.

i) Construir el diagrama de caja e interpretarlo.

Diagrama de caja del consumo de agua en miles de litros realizado por una muestra de familias de una población rural del norte de la provincia.



INTERPRETACION

En el diagrama de Boxplot o Caja podemos observar que el 50% central de los datos (la caja) no se encuentra en el centro de la distribución, sino un leve desplazamiento hacia los valores más bajos de la variable, con lo cual tiene una moderada asimetría positiva.



j) Calcular el coeficiente de asimetría y compararlo con el histograma y el diagrama de caja.

Coeficiente de Asimetría: $As = \frac{\bar{x} - Mo}{s} \rightarrow As = \frac{73 - 66,43}{23,13} = 0,28$

Interpretación: La distribución presenta una asimétrica positiva moderada.

Aclaración

El coeficiente de asimetría concuerda con lo observado en el Boxplot y en el gráfico de las frecuencias absolutas.

k) Calcular el coeficiente de Curtosis e interpretarlo.

Coeficiente de Curtosis: $K = \frac{\frac{1}{2}(Q_3 - Q_1)}{P_{90} - P_{10}}$

Cálculos:
 $\frac{90.n}{100} = \frac{90 \cdot 30}{100} = 27$

$\frac{10.n}{100} = \frac{10 \cdot 30}{100} = 3$

$$P_{10} = L_i + \frac{\frac{10.n}{100} - F_{(i-1)}}{f_{(i)}} \cdot a_i = 45 + \frac{3 - 3}{6} \cdot 15 = 45$$

$$P_{90} = L_i + \frac{\frac{90.n}{100} - F_{(i-1)}}{f_{(i)}} \cdot a_i = 105 + \frac{27 - 27}{2} \cdot 15 = 105$$

$$P_{90} - P_{10} = 60$$

➔ $K = \frac{\frac{1}{2}(32,25)}{60} = 0,26875$

La distribución es aproximadamente mesocurtica .



Coeficiente de Variación. Interpretación

Coeficiente de Variación: $CV = \frac{S}{X} = \frac{23,13}{73} \times 100 = 31,69\%$

Interpretación: La variabilidad en el consumo de agua (en miles de litros) que realizó una muestra de familias de una zona rural del norte de la provincia, en el ultimo año, es de una 31,69% con relación a la media.

La Utilidad de esta medida radica en la posibilidad de comparar la dispersión o variabilidad de dos o mas grupos.



Facultad de Ciencias Económicas

Universidad Nacional de Jujuy

ESTADÍSTICA
Año: 2022

MUCHAS GRACIAS

**Nos encontramos en el FORO
y en la Clase de Consulta.**