

UNIDAD 1

ESTADÍSTICA Y PROBABILIDAD

Actividades para el aprendizaje y ejemplos

Actividad 1

Describen y comparan distribuciones de datos utilizando representaciones gráficas, calculando, comparando y relacionando indicadores de tendencia central y dispersión.

Ejemplo A

Recoger información en el curso, sobre los aspectos siguientes: edad, estatura, sexo, color de pelo, tipo de música preferida, estatura del padre, estatura de la madre.

- I. Ordenar la información en tablas.
- II. Distinguir variables cuantitativas y cualitativas; señalar para estas últimas los valores que puede tomar cada variable y especificar el rango para los valores de las variables cuantitativas.
- III. Presentar la información utilizando el tipo de gráfico que se considere más adecuado, explicitando las ventajas de los tipos de gráficos seleccionados y las razones para desestimar otros.
- IV. Calcular la media aritmética para las variables cuantitativas, analizar su significado y constatar si es o no necesario el cálculo de otros indicadores para dar una imagen aproximada de la distribución de los datos.
- V. Utilizar porcentaje para cuantificar las variables cualitativas; graficar los resultados.
- VI. Explorar las posibilidades de organización de los datos, cálculo de indicadores y elaboración de tablas y gráficos en una planilla de cálculo.

INDICACIONES AL DOCENTE

Es necesario que los estudiantes ordenen la información recogida ya sea manualmente o utilizando una planilla de cálculo; se sugiere complementar el ejemplo incorporando variables que sean de interés para los alumnos y alumnas, que permitan una descripción del curso o de otro grupo de jóvenes en relación con sus preferencias e intereses.

Es importante que los estudiantes perciban que las distintas formas de representación gráfica son una herramienta que favorece la visualización de comportamientos y relaciones de la variable; asimismo, que distingan entre las representaciones posibles para variables cualitativas y cuantitativas. Se sugiere incentivar a los estudiantes que discutan sobre la relación entre los indicadores calculados y la distribución de los valores de la variable en estudio.

Ejemplo B

Las dos tablas que siguen resumen las notas obtenidas en una misma prueba de matemática, aplicada en dos cursos diferentes.

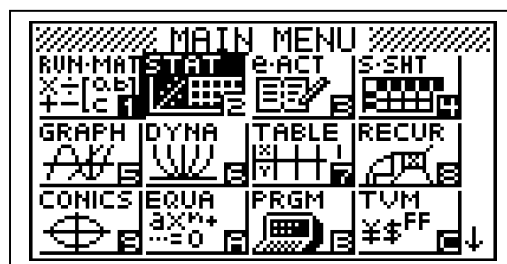
Curso A	
Notas	Frecuencia
7,0	3
6,7	5
6,3	4
6,0	8
4,0	8
3,4	2
3,0	4
	34 alumnos

Curso B	
Notas	Frecuencia
6,0	2
5,5	5
5,3	9
5,2	5
5,1	3
5,0	10
	34 alumnos

Calcular para ambos cursos el valor de la media aritmética y de la desviación estándar; graficar ambas distribuciones en un gráfico de barras. Comparar los gráficos y el valor de los indicadores calculados.

Solución: **Curso A**

En el menú seleccionamos el modo Estadística:



Ingresamos los datos del curso A a la lista 1(Notas) y 2.

SUB	List 1	List 2	List 3	List 4
	Notas	Fréc		
1	7	3		
2	6.7	5		
3	6.3	4		
4	6	8		

7

SUB	List 1	List 2	List 3	List 4
	Notas	Fréc		
5	4	8		
6	3.4	2		
7	3	4		
8				

4

Calculamos la media aritmética, desviación estándar

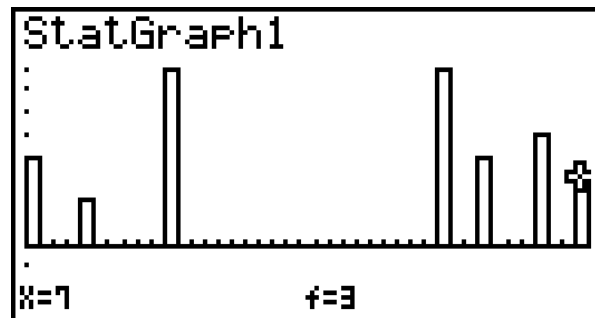
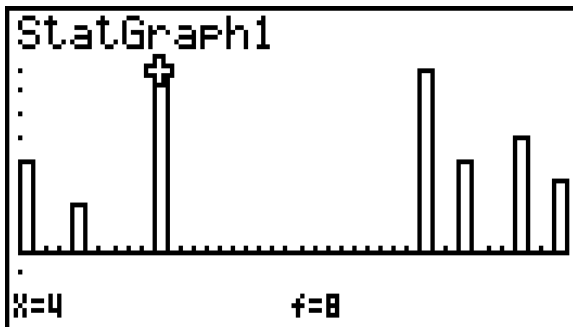
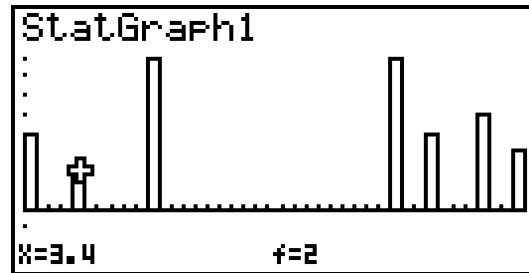
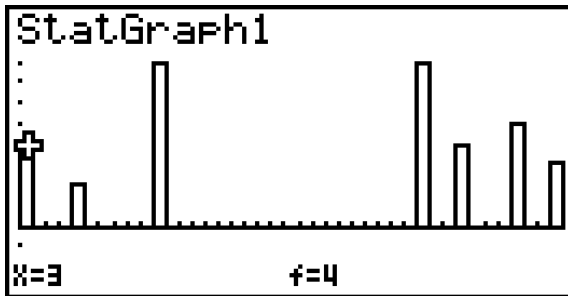
```

1 variable
x̄ =5.25
Σx =178.5
Σx² =1005.33
xσn =1.41634367
xσn-1 =1.43764327
n =34
↓
    
```

Media aritmética

Desviación estándar

Definimos el gráfico de barras (histograma) con la previa definición de las ventanas de visualización.



```

Vent. visualización
Xmin : 3
max : 7.1
scale : 0.1
dot : 0.03253968
Ymin : -3
max : 10
INIT TRIG STD STO RCL
    
```

```

su Ajuste histograma
Start: 3
Width: 0.1
Dibujar: [EXE]
GPH1 GPH2 GPH3 SEL SET
    
```

Solución: **Curso B**

Ingresamos los datos del curso B a la lista 1 y 2.

	LiSt 1	LiSt 2	LiSt 3	LiSt 4
SUB	Notas	Fréc		
1	6	2		
2	5.5	5		
3	5.3	9		
4	5.2	5		

	LiSt 1	LiSt 2	LiSt 3	LiSt 4
SUB	Notas	Fréc		
4	5.2	5		
5	5.1	3		
6	5	10		
7				

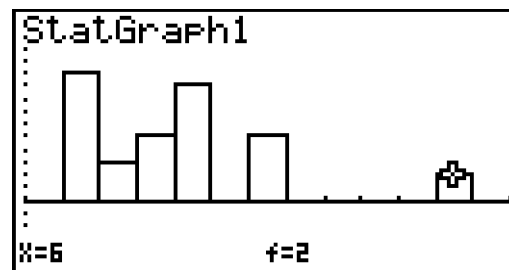
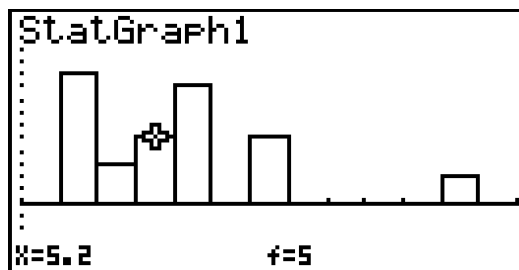
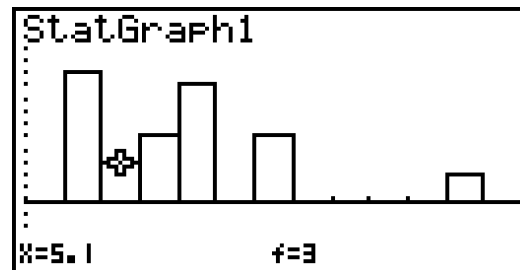
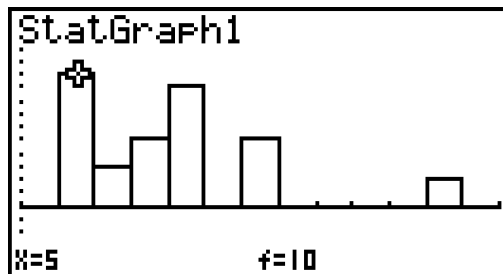
Calculamos la media aritmética, desviación estándar:

1 variable	
\bar{x}	=5.25 ←
Σx	=178.5
Σx^2	=939.29
$s_{x\bar{x}}$	=0.25234197 ←
$s_{x\bar{x}-1}$	=0.2561368
n	=34

Media aritmética

Desviación estándar

Definimos el gráfico de barras (histograma) y sus respectivas ventanas de visualización



Vent. visualización	
Xmin	:4.9
max	:6.2
scale	:0.1
dot	:0.01031746
Ymin	:-5
max	:14

Ajuste histograma	
Start	:4.9
Width	:0.1
Dibujar	: [EXE]

Ejemplo C

Analizar el siguiente cuadro que resume la distribución del ingreso per cápita.

Evolución de la distribución del ingreso monetario según quintiles de ingreso 1987 - 1998

Quintil	1987	1990	1992	1994	1996	1998
I	4,3	4,4	4,6	4,3	4,1	4,1
II	7,9	8,2	8,5	8,2	8,2	8,2
III	11,7	12,3	12,2	12	11,9	11,8
IV	19	18,1	18,4	18,5	19,1	19,1
V	57,2	56,9	56,3	56,9	56,7	56,9
Total	100	100	100	100	100	100

Fuente: MIDEPLAN, Encuestas CASEN.

- I. Investigar sobre el monto de los ingresos per cápita en los años que indica el cuadro y establecer los valores por año y quintil.
- II. Establecer el significado de los quintiles y su aporte como complemento a la media aritmética que es el ingreso per cápita.

Ejemplo D

Pedir a algunos alumnos o alumnas que registren sus pulsaciones, tomadas durante 1 minuto, cinco veces cada uno, antes y después de hacer un ejercicio físico. Buscar formas de organizar y presentar la información recogida de modo que sea posible comparar la información para un mismo alumno antes y después de los ejercicios y comparar también para todos los alumnos.

Solución:

Las siguientes tablas muestran por columnas el número de pulsaciones de un número de 4 alumnos, tomadas en 5 oportunidades y durante un minuto antes de realizar actividad física.

	List 1	List 2	List 3	List 4
SUB	A1 1	A1 2	A1 3	A1 4
1	80	87	77	95
2	83	90	80	98
3	81	86	79	91
4	85	88	81	96
				80.00000

	List 1	List 2	List 3	List 4
SUB	A1 1	A1 2	A1 3	A1 4
3	81	86	79	91
4	85	88	81	96
5	86	89	78	92
6				

A continuación, calculamos el promedio y la desviación estándar de las pulsaciones tomadas al alumno 1, Alumno 2, Alumno 3, Alumno 4 y Alumno 5, respectivamente

Alumno 1	
1 variable	
\bar{x}	=83
Σx	=415
Σx^2	=34471
$s\sigma_n$	=2.28035085
$s\sigma_{n-1}$	=2.54950975
n	=5

Alumno 2	
1 variable	
\bar{x}	=88
Σx	=440
Σx^2	=38730
$s\sigma_n$	=1.41421356
$s\sigma_{n-1}$	=1.58113883
n	=5

Alumno 3	
1 variable	
\bar{x}	=79
Σx	=395
Σx^2	=31215
$s\sigma_n$	=1.41421356
$s\sigma_{n-1}$	=1.58113883
n	=5

Alumno 4	
1 variable	
\bar{x}	=94.4
Σx	=472
Σx^2	=44590
$s\sigma_n$	=2.57681974
$s\sigma_{n-1}$	=2.88097205
n	=5

Luego, consideramos todas las medias de cada alumno ingresadas en la lista 5 y con éstas obtenemos un nuevo promedio.

	List 5	List 6	List 7	List 8
SUB	Medias			
1	83			
2	88			
3	79			
4	94.4			

1VAR 2VAR REG SET

```

1 variable
x̄ =86.1
Σx =344.4
Σx² =29785.36
x̄σn =5.75586657
x̄σn-1 =6.64630223
n =4
↓
  
```

El promedio de todas las pulsaciones de los alumnos tomadas en 5 oportunidades y durante un minuto alcanza a 86,1. Los siguientes cuadros muestran la suma de las pulsaciones totales de cada alumno tomada por lista y donde la suma total de todos los datos ingresados dividido por el total de datos da como resultado el mismo valor de la media obtenido en el cuadro anterior.

```

Sum List 1          415
Sum List 2          440
Sum List 3          395
□
DEL DELA
  
```

```

Sum List 2          440
Sum List 3          395
Sum List 4          472
□
List L→M Dim Fill Seq ▸
  
```

```

(415+440+395+472)÷20
                        86.1
□
JUMP DEL PMAT MATH
  
```

Consideremos todas las desviaciones estándar obtenidas por alumno anteriormente, y agregamos estos 4 valores en una nueva lista. Calculamos para dichas desviaciones la nueva desviación estándar como sigue:

	List 7	List 8	List 9	List 10
SUB	Jésus			
1	2.2803			
2	1.4142			
3	1.4142			
4	2.5768			

2.28035085

1VAR 2VAR REG SET

```

1 variable
x̄ =1.92139942
Σx =7.68559771
Σx² =15.8399999
x̄σn =0.51790368
x̄σn-1 =0.59802366
n =4
↓
  
```

El valor de dicha desviación es de 0,51790368.

Consideremos todos los datos de las 4 listas en una sola lista (lista 6) de los cuadros siguientes y calculemos la desviación estándar para todos esos datos.

SUB	List 6	List 7	List 8	List 9
	Todos			
1	80			
2	83			
3	81			
4	85			

GRAPH CALC TEST INTR DIST

SUB	List 6	List 7	List 8	List 9
	Todos			
5	86			
6	87			
7	90			
8	86			

86
GRAPH CALC TEST INTR DIST

SUB	List 6	List 7	List 8	List 9
	Todos			
9	88			
10	89			
11	77			
12	80			

80
GRAPH CALC TEST INTR DIST

SUB	List 6	List 7	List 8	List 9
	Todos			
13	79			
14	81			
15	78			
16	98			

98
GRAPH CALC TEST INTR DIST

```

1 variable
x̄ =86.1
Σx =1722
Σx² =149006
x̄σn =6.09015599
x̄σn-1 =6.2483682
n =20
    
```

↓

El valor de esta desviación es de 6.09015599, valor distinto del que calculamos en el caso anterior.

Consideremos todas las desviaciones estándar obtenidas por alumno anteriormente, y agregamos estos 4 valores en una nueva lista. Calculamos para dichas desviaciones la nueva desviación estándar como sigue:

SUB	List 7	List 8	List 9	List 10
	desos			
1	2.2803			
2	1.4142			
3	1.4142			
4	2.5768			

2.28035085
IVAR EVAR REG SET

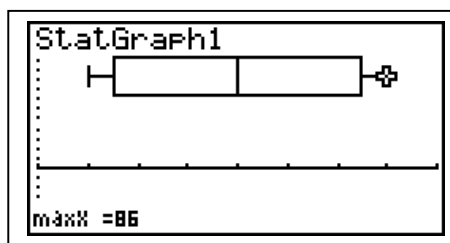
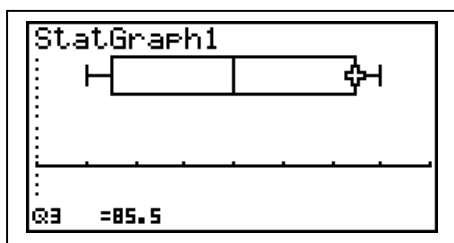
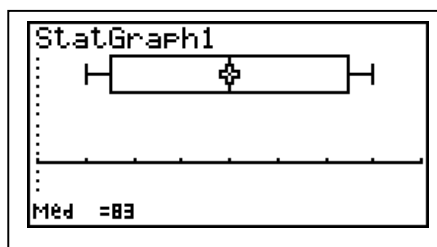
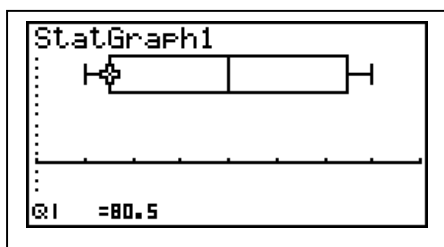
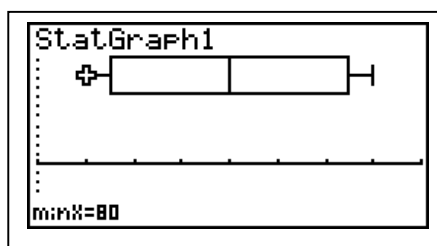
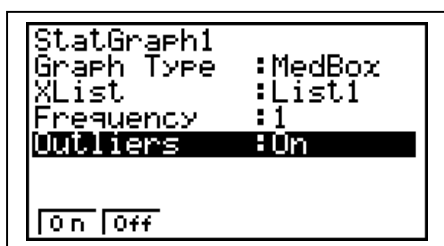
```

1 variable
x̄ =1.92139942
Σx =7.68559771
Σx² =15.8399999
x̄σn =0.51790368
x̄σn-1 =0.59802366
n =4
    
```

↓

El valor de dicha desviación es de 0,51790368 el cual no guarda relación alguna con la desviación estándar del total de datos calculada previamente. En cambio, al calcular las medias por grupos y con cada media obtenida calcular una nueva media, ese valor resultante si es equivalente al de la media del total de alumnos.

Las pulsaciones del alumno 1 serán representadas mediante un diagrama de caja. Estos diagramas son útiles para reflejar aquellos valores atípicos (outliers) que figuran dentro de la muestra de datos.



Ejemplo E

El análisis de las notas de un curso al término del primer y segundo trimestre señala que en ambos trimestres el promedio en matemática es 5.1, la nota máxima es 7 y la mínima es 3.2. Sin embargo, los alumnos tienen la sensación de mejores resultados en un trimestre que en otro. Determinar cuál es ese mejor trimestre y por qué se considera mejor. Las notas en cada trimestre se presentan en los cuadros que siguen.

Primer trimestre									
7	7	6,9	6,8	6,5	6,3	5,8	5,7	5,6	5,6
5,6	5,4	5,4	5,2	5,2	5,2	4,8	4,8	4,8	4,5
4,3	4,3	4,3	4,1	4,1	4,1	4,1	3,2	3,2	3,2

Segundo trimestre									
7	6,4	6,1	6	5,7	5,5	5,4	5,3	5,3	5,3
5,3	5,2	5,2	5,2	5,1	5	5	5	5	5
5	4,9	4,7	4,7	4,7	4,6	4,5	4,5	3,2	3,2

Calcularemos la media y desviación estándar por trimestre. Para ello ingresaremos las notas de cada trimestre con sus respectivas frecuencias. Las listas 1 y 3 contienen las notas del primer y segundo trimestre respectivamente y las lista 2 y 4 sus frecuencias respectivas.

	List 1	List 2	List 3	List 4
SUB	1ero	Frec	2do	Frec
1	7	2	7	1
2	6.9	1	6.4	1
3	6.8	1	6.1	1
4	6.5	1	6	1

	List 1	List 2	List 3	List 4
SUB	1ero	Frec	2do	Frec
5	6.3	1	5.7	1
6	5.8	1	5.5	1
7	5.7	1	5.4	1
8	5.6	3	5.3	4

	List 1	List 2	List 3	List 4
SUB	1ero	Frec	2do	Frec
9	5.4	2	5.2	3
10	5.2	3	5.1	1
11	4.8	3	5	6
12	4.5	1	4.9	1

	List 1	List 2	List 3	List 4
SUB	1ero	Frec	2do	Frec
13	4.3	3	4.7	3
14	4.1	4	4.6	1
15	3.2	3	4.5	2
16			3.2	2

El primer trimestre tiene nota promedio de 5,1 y desviación estándar 1,09

```

1Var XList :List1
1Var Freq  :List2
2Var XList :List1
2Var YList :List2
2Var Freq  :1
LIST
    
```

```

1 variable
x̄      =5.1
Σx     =153
Σx²    =816.24
x̄σn    =1.09453186
x̄σn-1  =1.11324316
n      =30
    
```

El segundo trimestre tiene nota promedio de 5,1 y desviación estándar 0,75

```

1Var XList :List3
1Var Freq  :List4
2Var XList :List1
2Var YList :List2
2Var Freq  :1
LIST
    
```

```

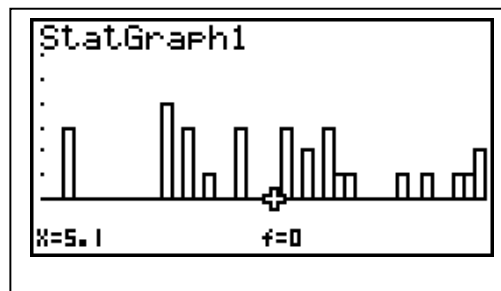
1 variable
x̄      =5.1
Σx     =153
Σx²    =796.98
x̄σn    =0.74565407
x̄σn-1  =0.75840122
n      =30
    
```

La desviación estándar es menor en el segundo trimestre que en el primero, y dado que en ambos trimestres hay igual promedio, las notas del segundo trimestre tienen una menor dispersión respecto a la media. Esto último indica que la muestra del segundo trimestre es más homogénea que la del primero, es decir, hubo un mejor rendimiento global en dicho segundo trimestre.

Los siguientes gráficos de frecuencias (histogramas) muestran la ubicación de las notas de cada trimestre en relación a la media. El gráfico 2, correspondiente a las notas del segundo trimestre, muestra mayor concentración de notas en torno a la media, a diferencia del gráfico 1, correspondiente al primer trimestre, cuyas notas están más dispersas en torno a la media. Como conclusión podemos decir que se observa un mejor rendimiento en el segundo trimestre que en el primero.

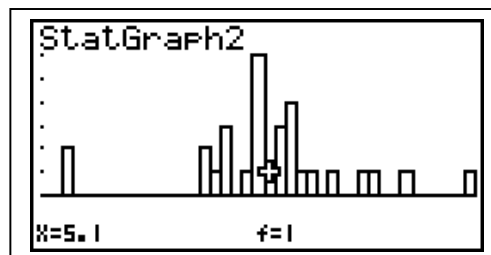
```

Vent. visualización
Xmin :3
max  :7.1
scale:0
dot  :0.03253968
Ymin :-1
max  :6
INIT TRIG STD STO RCL
    
```



```

StatGraph2
Graph Type :Hist
XList      :List3
Frequency  :List4
GPH1 GPH2 GPH3
    
```



Ejemplo adicional:

Un colegio debe inscribir a dos alumnos en el campeonato regional de atletismo en la prueba de 400 metros planos. El tiempo en segundos obtenido por los alumnos de la selección del colegio se presentan en la siguiente tabla.

Nombres	TIEMPO EN SEGUNDOS											
	Alberto	80	79	81	79	79	81	80	79	78	77	78
Vicente	83	77	75	82	83	81	80	79	82	77	81	79
Julio	79	78	80	81	85	79	86	79	75	74	84	78
Manuel	80	79	80	79	80	81	80	80	80	81	80	79

Ingresaremos en las listas 1, 2, 3 y 4 los nombres y tiempos respectivos de cada estudiante

	List 1	List 2	List 3	List 4
SUB	Albert	Vicent	Julio	Manuel
1	76	79	79	80
2	80	80	78	79
3	84	80	80	80
4	81	79	81	79

	List 1	List 2	List 3	List 4
SUB	Albert	Vicent	Julio	Manuel
5	77	79	85	80
6	81	80	79	81
7	84	78	86	80
8	75	80	79	80
				75.0

	List 1	List 2	List 3	List 4
SUB	Albert	Vicent	Julio	Manuel
9	78	79	75	80
10	77	78	74	81
11	76	79	84	80
12	81	80	78	78
				81.0

Calcularemos la media y desviación estándar por estudiante:

Alberto

```
1 variable
x̄ =79.1666666
Σx =950
Σx² =75314
xón =2.96741563
xón-1 =3.09936454
n =12 ↓
```

Vicente

```
1 variable
x̄ =79.25
Σx =951
Σx² =75373
xón =0.72168783
xón-1 =0.75377836
n =12 ↓
```

Julio

```
1 variable
x̄ =79.8333333
Σx =958
Σx² =76630
xón =3.53160335
xón-1 =3.68863939
n =12 ↓
```

Manuel

```
1 variable
x̄ =79.8333333
Σx =958
Σx² =76488
xón =0.79930525
xón-1 =0.8348471
n =12 ↓
```

Sin duda que Alberto posee la menor de las medias en los tiempos, sin embargo, la media de Vicente es levemente mayor que la de Alberto. Si comparamos las desviaciones estándar entre ellos, notamos que la de Alberto es más de cuatro veces mayor que la de Vicente. Esto indica que si bien la media de Alberto es la menor, sus tiempos son muy poco homogéneos y no da seguridad que pueda correr en el día de la competencia un tiempo cercano al promedio. Lo mismo ocurre entre Julio y Manuel. Ambos tienen igual promedio en el tiempo pero Julio sobrepasa a Manuel en más de tres desviaciones estándar. Luego, Manuel es mejor candidato que Julio.

Ejemplo F

En un grupo, los varones obtienen las siguientes notas en Educación Física:

5,6; 5,5; 4,8; 7; 6,4; 5,5; 4,5; 5,6; 5,4; 5,8; 6,0; 5,2

Las damas, a su vez, obtuvieron:

6,6; 6,5; 5,8; 4,8; 5,9; 7

Un niño dijo: nuestro promedio es 5,6; y una niña dijo nuestro promedio es 6,1.

Julio, que estaba escuchando, realizó los siguientes cálculos en su cuaderno:

$11,7 : 2 = 5,85$ y dijo: este grupo tiene promedio 5,9; ¿qué error cometió Julio? Fundamente su respuesta.

Solución:

Verificamos las medias de cada grupo con el uso de listas en la ventana STAT. Ingresamos las notas de los varones a la lista 1 y las notas de las damas a la lista 2 como se puede ver a continuación:

	List 1	List 2	List 3	List 4
SUB	Varon	Dama		
1	5.6	6.6		
2	5.5	6.5		
3	4.8	5.8		
4	7	4.8		

Varon

GRAPH CALC TEST INTB DIST ▸

Calculamos la media de los varones y obtenemos la nota 5,6

```
1Var XList :List1
1Var Freq  :1
2Var XList  :List1
2Var YList  :List2
2Var Freq   :1
LIST
```

```
1 variable
x      =5.60833333
Σx     =67.3
Σx²    =382.31
x̄n     =0.63699598
x̄n-1  =0.66532061
n      =12
```

Calculamos la media de las damas y obtenemos la nota 6,1

```
1Var XList :List2
1Var Freq  :1
2Var XList  :List1
2Var YList  :List2
2Var Freq   :1
LIST
```

```
1 variable
x      =6.1
Σx     =36.6
Σx²    =226.3
x̄n     =0.71180521
x̄n-1  =0.77974354
n      =6
```

Definimos una lista 3 que contiene el total de notas entre damas y varones y calculamos la media de la lista 3.



El promedio del total de damas y varones es 5,8 el cual no coincide con el promedio de 5,9 obtenido por el niño.

En las indicaciones al docente, podemos constatar el problema existente en el cálculo realizado por el niño, el que tiene que ver con obtener promedios considerando universos de **distinto tamaño**, lo cual no tiene sentido.

Designaremos como V_1, V_2, \dots, V_{12} , las notas obtenidas por el primer varón hasta el doceavo varón. Designaremos como D_1, D_2, \dots, D_6 , las notas obtenidas por la primera dama hasta la sexta dama. Calcularemos el promedio de notas de los varones y de las damas y los denominaremos \bar{V} y \bar{D} respectivamente:

$$\frac{\sum_{i=1}^{12} V_i}{12} = \bar{V} \quad , \quad \frac{\sum_{i=1}^6 D_i}{6} = \bar{D}$$

Es evidente que la cantidad total de datos es distinta en cada caso. El cálculo realizado por el niño fue el siguiente: $\frac{\bar{V} + \bar{D}}{2}$, lo que a su parecer representaba el promedio total de todos los datos.

A continuación calcularemos el promedio total de todas las notas obtenidas entre los varones y las damas y lo denominaremos \bar{T} .

$$\bar{T} = \frac{\sum_{i=1}^{12} V_i + \sum_{i=1}^6 D_i}{18}$$

A continuación se puede demostrar que \bar{T} es distinto a $\frac{\bar{V} + \bar{D}}{2}$

$$\frac{\bar{V} + \bar{D}}{2} = \left[\frac{\sum_{i=1}^{12} V_i}{12} + \frac{\sum_{i=1}^6 D_i}{6} \right] \cdot \frac{1}{2} = \left[\frac{\sum_{i=1}^{12} V_i + 2 \sum_{i=1}^6 D_i}{24} \right] \neq \frac{\sum_{i=1}^{12} V_i + \sum_{i=1}^6 D_i}{18} = \bar{T}$$

Si los universos considerados son de igual tamaño, entonces sería correcto considerar los resultados de los promedios parciales, como para obtener un nuevo promedio.

Designaremos como V_1, V_2, \dots, V_n , las notas obtenidas por el primer varón hasta el n-ésimo varón. Designaremos como D_1, D_2, \dots, D_n , las notas obtenidas por la primera dama hasta la n-ésima dama. Calcularemos el promedio de notas de los varones y de las damas y los denominaremos \bar{V} y \bar{D} respectivamente:

$$\frac{\sum_{i=1}^n V_i}{n} = \bar{V} \quad \frac{\sum_{i=1}^n D_i}{n} = \bar{D}$$

Es evidente que la cantidad total de datos es igual en cada caso. El cálculo realizado por el niño sería el siguiente: $\frac{\bar{V} + \bar{D}}{2}$, lo que a su parecer representaría el promedio total de todos los datos.

A continuación calcularemos el promedio total de todas las notas obtenidas entre los varones y las damas y lo denominaremos \bar{T} .

$$\bar{T} = \frac{\sum_{i=1}^n V_i + \sum_{i=1}^n D_i}{2n}$$

A continuación se puede demostrar que \bar{T} es igual a $\frac{\bar{V} + \bar{D}}{2}$

$$\frac{\bar{V} + \bar{D}}{2} = \left[\frac{\sum_{i=1}^n V_i}{n} + \frac{\sum_{i=1}^n D_i}{n} \right] \cdot \frac{1}{2} = \left[\frac{\sum_{i=1}^n V_i + \sum_{i=1}^n D_i}{2n} \right] = \bar{T}$$

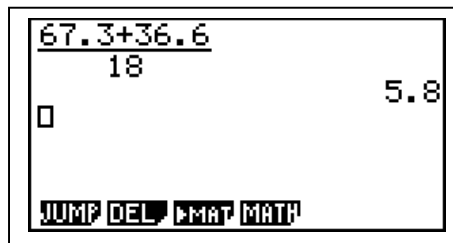
En nuestro caso: $\sum_{i=1}^{12} V_i = 67,3$, $\sum_{i=1}^6 D_i = 36,6$,

Calcularemos el promedio de notas de los varones y de las damas y los denominaremos \bar{V} y \bar{D} respectivamente:

$$\frac{\sum_{i=1}^{12} V_i}{12} = \frac{67,3}{12} = 5,6 = \bar{V} \quad \text{y} \quad \frac{\sum_{i=1}^6 D_i}{6} = \frac{36,6}{6} = 6,1 = \bar{D}$$

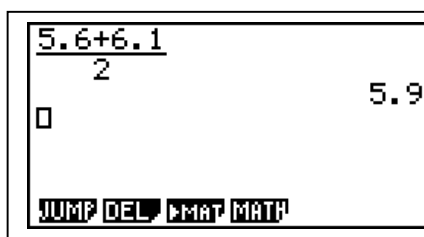
A continuación calcularemos el promedio total de todas las notas obtenidas entre los varones y las damas y lo denominaremos \bar{T} .

$$\bar{T} = \frac{\sum_{i=1}^{12} V_i + \sum_{i=1}^6 D_i}{18} = \frac{67,3 + 36,6}{18} = 5,8$$



Se puede verificar que \bar{T} es distinto a $\frac{\bar{V} + \bar{D}}{2}$, dado que:

$$\frac{\bar{V} + \bar{D}}{2} = \frac{5,6 + 6,1}{2} = 5,9$$



Claramente \bar{T} es distinto a $\frac{\bar{V} + \bar{D}}{2}$.

Propuesta didáctica de la Calculadora Casio fx-9860G en la Unidad de Estadística 4to Medio del programa general

Objetivos Fundamentales

Unidad de Estadística y Probabilidad

- Variable aleatoria: estudio y experimentación en casos concretos. Gráfico de frecuencia de una variable aleatoria a partir de un experimento estadístico.**
- Relación entre la probabilidad y la frecuencia relativa. Ley de los grandes números. Uso de programas computacionales para la simulación de experimentos aleatorios.**
- Resolución de problemas sencillos que involucren suma o producto de probabilidades. Probabilidad condicionada.**

En primer lugar definamos algunos aspectos generales del uso de la calculadora:

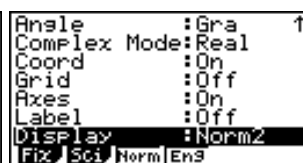
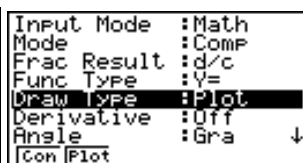
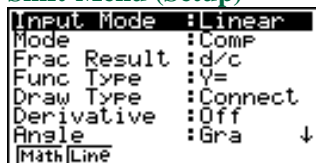
Trabajo con números



MENU-1

Al usar la calculadora podrás seleccionar el Menú 1 de herramientas básicas. En este lugar se hacen los cálculos de suma, resta, multiplicación y división, etc. En estas tareas puedes hacer que la calculadora te entregue decimales o fracciones, para esto último sigue los pasos de a continuación...

Shift-Menu (Setup)



- EXIT

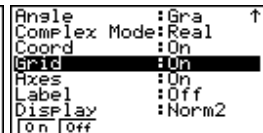
Trabajo con estadística



MENU-2

Cuando apliquemos las herramientas estadísticas usaremos la calculadora en Menú 2. Este lugar permite ingresar datos y graficarlos, para este efecto los gráficos también deben estar configurados, asegúrate de tener las siguientes condiciones...

Shift-Menu (Setup)



EXIT

Actividad 2

Analizan e interpretan información estadística actualizada y comunican las conclusiones de estos análisis.

Ejemplo A

Seleccionar algunas de las variables del cuadro de datos sobre 97 países, que se incorpora al final de esta unidad. Esta información está tomada de internet.

INDICACIONES AL DOCENTE

- El profesor o profesora podrá pedir a los alumnos y alumnas que busquen artículos en la prensa en que se hable de alguno de estos indicadores, expliquen con sus propias palabras la utilidad que pueden tener y averigüen la fuente responsable de esa información.
- En este cuadro se ha usado un código para agrupar los países en función de la zona geográfica y desarrollo económico. Los estudiantes podrían sugerir otras variables de clasificación de los países o añadir otras variables o países.
- Los alumnos y alumnas pueden analizar las ventajas que el diagrama de barras tiene frente a la tabla para visualizar el grupo que tiene mayor o menor número de países. Asimismo pueden elaborar otros gráficos adecuados para representar algunas de las variables elegidas.

I. Analizar el significado de esas variables considerando la definición propuesta al inicio del cuadro.

En esta primera tarea podemos preguntarnos:

¿Qué comentarios se pueden hacer respecto de la natalidad y mortalidad de los países?, ¿cómo se relaciona con el tamaño de población? ¿en algunos casos existen leyes impuestas en el control de la natalidad? Primero comentemos la información que se muestra en la tabla adjunta, luego, averiguamos más de los países extraída desde Internet sobre natalidad y mortalidad.

País	Tasa natalidad	Tasa mortalidad	Diferencia nat/mort	Población (miles)	Procedemos a establecer la diferencia entre la natalidad y la mortalidad, ya que esta responde a una tasa por cada 1000 habitantes. Seleccionando algunos casos de países: 1. Alemania del este , notablemente tiene la misma cifra en natalidad y
Afganistán	4	18.7	21	160	
Alemania	1	12.4	-	613	
Argelia	3	8.3	27	244	
Angola	4	20.2	27	969	
Arabia Saudí	4	7.6	34	135	
Botswana	4	11.6	36	121	
Brasil	2	7.9	20	147	
Chile	2	5.8	17	129	
China	2	6.7	14	110	
España	1	8.2	2	391	

Francia	1	9.4	4	561	<p>mortalidad, se deduce un estancamiento en la población.</p> <p>2. Argelia y Afganistán tienen casi la misma natalidad pero difieren notablemente en la tasa de mortalidad, debido a condiciones sociales y económicas.</p>
India	3	10.2	20	832	
Iraq	4	7.8	34	182	
Turquía	2	8.4	20	548	
U.K.	1	11.5	2	572	
U.S.A.	1	8.1	8	248	

3. Por cercanía de nombres entre **Chile y China**, estamos seguidos en la tabla, sin embargo vemos un país con baja población como el nuestro y el más poblado del mundo. Si dividimos para proporcionar los habitantes $1105067/12980 = 85.1$ esto significa que por cada un chileno existen 85 chinos. A pesar de esta notable diferencia, Chile gana en natalidad, seguramente por las restricciones que tiene el pueblo de China (El Gobierno de China, en 2009 con 1,3 billones de personas, impuso su política de restricción de la natalidad en 1979.).
4. **Botswana**. Tiene entre los elegidos una diferencia entre nacidos y fallecidos la más alta, después le sigue **Iraq** a pesar de las dificultades de la guerra Iraq-Irán desde 1980.

Más información extraída y actualizada de Internet [fecha de la búsqueda dic. 2009].

1. Sobre **Alemania del Este**. [Antecedentes de la UE de post segunda guerra mundial, datos desde 1995-2001]
Esta pequeña diferencia entre natalidad y mortalidad hace que el crecimiento vegetativo de estos países sea muy pequeño, y hay años en que incluso se da un crecimiento negativo en algunos países. Las tasas de Natalidad no pasan en general del 14% de la población y las de Mortalidad del 12%^o En cuanto a la Mortalidad Infantil, no superan el 10%^o lo que se debe al nivel de desarrollo sanitario e higiénico alcanzado. A esto, también se debe que la Esperanza de Vida al nacer, sea de las más altas del mundo, pues los Hombres superan los 71 años, y las mujeres los 77 años, y en algunos países pasan de los 80 años.
2. Sobre **Argelia**. [Material extraído de la dirección <http://ddd.uab.cat/pub/tesis/2001/tdx-0626101-100609/jclu1de1.pdf>]
La tasa de mortalidad global no es más que la suma ponderada de las tasas de mortalidad específicas por edad. En esquema, globalmente da igual que se mueran muchos niños o muchos ancianos. Pero el efecto de las tasas de mortalidad por edad en la esperanza de vida es tremendo: si la mortalidad infantil es alta la esperanza de vida cae en picada, cosa que no sucede si la mortalidad se concentra en edades avanzadas.

3. Sobre **China** [Material extraído de <http://spanish.peopledaily.com.cn/rk2.htm> (2000, Diario del Pueblo en línea, de China)].

A partir de la década de los 70, sobre todo después de la aplicación de la política de reforma y apertura al mundo exterior, China estableció políticas demográficas para controlar el crecimiento poblacional y elevar la cualidad de la población. Desde entonces, la planificación familiar, quedó implantada como política fundamental del Estado. El gobierno chino estimula el matrimonio y la procreación tardíos y aboga por que una pareja tenga un solo hijo y planifique razonablemente el nacimiento de su segundo hijo conforme a leyes y reglamentos. Esta política también se aplica a las minorías étnicas. Todas las provincias, regiones autónomas y ciudades directamente subordinadas al gobierno central han elaborado políticas concretas de acuerdo con la situación local.

II. Construir un diagrama de barras que represente el número de países en los diferentes grupos. A partir del mismo, construir una tabla de frecuencias y discutir el significado de las frecuencias absolutas, relativas y porcentajes.

Contabiliza el número de países de cada grupo (columna de Grupo).

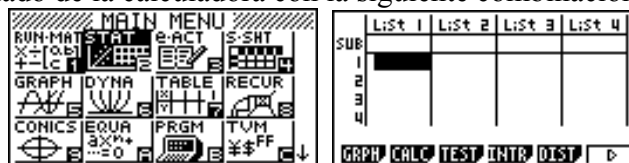
Completa la tabla siguiente:

Comienza contando, cuántos países pertenecen al grupo1 de Europa Oriental y después continúa con los del grupo2 de Iberoamérica, así continúa. Nuevamente esta información se obtienen de la Tabla 1: Fichero de datos del proyecto “Análisis demográfico” final de esta unidad.

Grupo (xi)	Países (ni)
1	11
2	
3	
4	
5	
6	
Total:	

Ahora la idea será ingresar esta tabla en la calculadora, con grupos y frecuencias.

Usa el teclado de la calculadora con la siguiente combinación MENU - 2

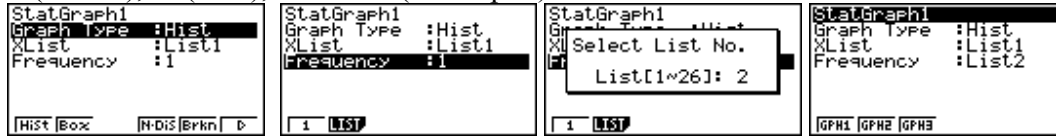


Al subir al título de la lista1 se puede ingresar la palabra GRUPO con la combinación de teclas ALPHA – (letra) al finalizar la tecla EXE para el ingreso.



Así cuando termines podemos graficar el Histograma que lo representa. Entonces presiona la siguiente secuencia de teclas:

F1(GRPH), F6(SET), en GPH1 (StatGrph1)



Estas indican a la calculadora que se trata de un Histograma de barras, los datos están en Lista1 y la frecuencia en la Lista2.

	List 1	List 2	List 3	List 4
SUB	GRUPO	PAISES		
1	1	11		
2	2	12		
3	3	19		
4	4	11		

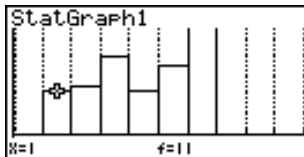
View Window	
Xmin	: 0
max	: 10
scale	: 1
dot	: 0.07936507
Ymin	: -5
max	: 30

Después de EXIT Ingresar los datos de View Window. presiona Shift F3 (V-Window)



Luego EXIT para obtener el gráfico con F1(GPH1)

Además con Shift-F1(Trace)



Puedes desplazarte por los extremos superiores de cada barra del Histograma.

Confirma así las frecuencias por cada grupo de países.
¿En qué grupo se observan menos países? ¿en qué grupo más?

Traza y completa la siguiente tabla de frecuencia en tu cuaderno. Primero, frecuencia relativa y porcentual.

Tabla de frecuencias:



Variab.	Frec		Acumuladas		
	abs.	Frec relat.	FAA	FRA	Porc.Ac
Grupo	Países				
(xi)	(ni)	(fi)	(pi)	(Ni)	(Fi)
1	11	1/9			
2	12	1/8			
3	19	1/5			
4	11	1/9			
5	17	1/6			
6	27	2/7			


Total: 97

Obs. No se trata de que copies la tabla, sino confirmes cada uno de los valores con el apoyo de la calculadora.

- Observa que los f_i son en realidad una buena aproximación de la exacta frecuencia relativa, la cual se obtiene al dividir la Frecuencia absoluta por el Total de países. Es decir el primer caso debiera ser 11/97.
- Podemos pensar que la frecuencia relativa es un concepto, corresponde a imaginar la parte que representa la clase (por ejemplo grupo1) sobre el conjunto total de datos.

En este caso, la frecuencia relativa se puede estudiar con una precisión al segundo decimal. Por ejemplo:

En MENU 1 presiones a^b/c  luego ingresa 



Sin embargo, una fracción reducida y próxima a la fracción 11/97 con precisión al segundo decimal corresponde a 0.11. Verifica esto, ennegrece seleccionando la fracción 11/97 de respuesta, luego con la tecla $F \leftrightarrow D$ quedará expresada en decimales.

Ahora las frecuencias relativas que utilizaremos serán una fracción que represente la precisión al segundo decimal, es por ello que el valor 0.111111 el cual puede ser una mejor representación, esta corresponde a 1/9. Así, podemos afirmar que 1/9 o la “novena parte del total de países” corresponde al grupo1 de Europa Oriental. Continúe de esta manera con el resto de las frecuencias relativas...

Comentario:

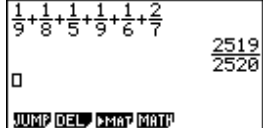
1. La **frecuencia relativa** (f_i). Es la razón descriptiva entre la cantidad de países del grupo n_i respecto del total N , por ejemplo el grupo 1 tiene 11/97 se lee “11 sobre 97” el cual en términos decimales se puede aproximar (o redondear) al segundo decimal. Es así que podemos simplificar estas fracciones, aproximadamente en $\frac{11}{97} \approx \frac{1}{9}$, es una fracción

cercana al valor exacto.

Se llama relativa ya que esta guarda relación con un total 1, el cual como unidad puede ser cualquier cifra.

Esta situación no pierde precisión y permite conceptualizar la parte del grupo respecto del total (representado por 1).

Además, si sumamos todos estos resultados de fracciones con la calculadora $1/9 + 1/8 + 1/5 + 1/9 + 1/6 + 2/7$ se obtiene: 2519/2520.



Y esta solución en decimales es: $\frac{2519}{2520} = 0.999603\dots$ Entonces así confirmamos el valor próximo a 1, el cual equivalente al total de la muestra.

2. El **porcentaje relativo** (p_i). Es un porcentaje que representa al grupo respecto del total de países, esta vez el porcentaje tiene dos dígitos y precisión a un decimal. Podemos notar que en el grupo 6, los de países Africanos, son los más numerosos en términos porcentuales.

Continúa con la tabla siguiente. El primer valor porcentual lo obtienes con la fracción multiplicando por 100 en cada caso. Continúa a un decimal de precisión.

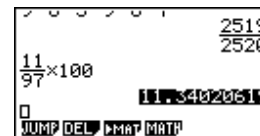


Tabla de frecuencias:

Variab.	Frec			Acumuladas		
	abs.	Frec rel.		FAA	FRA	Porc.Ac
Grupo (xi)	Países (ni)	(fi)	(pi)	(Ni)	(Fi)	(Pi)
1	11	1/9	11,3%			
2	12	1/8				
3	19	1/5				
4	11	1/9				
5	17	1/6				
6	27	2/7				

Total: 97

Responde: ¿cuál es el porcentaje respecto del total del grupo más numeroso? ¿y del menos?

Frecuencias acumuladas y su interpretación

N_i = Frecuencia absoluta acumulada FAA. Corresponde a la suma de la frecuencia absoluta de la clase (en este caso la de cada grupo) con el N_{i-1} anterior, es decir **N_i = N_{i-1} + n_i**. Si es el primer caso y no existe el N₀ anterior porque se trata del primero N₁ entonces es solo hacer N₁=n₁. El último será el mismo valor total de países N₆=97.

F_i = Frecuencia relativa acumulada. Corresponde a la fórmula **F_i = N_i/97** De esta manera el último será **F₆ = 97/97 = 1**

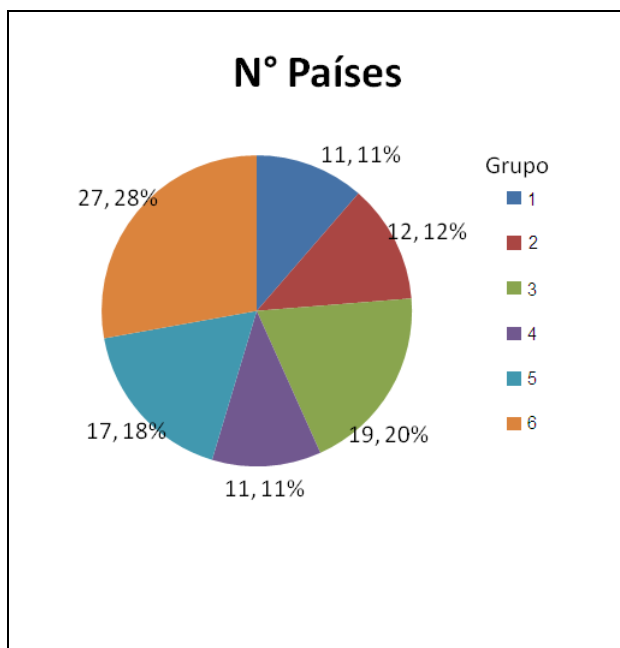
P_i = Porcentaje de la frecuencia relativa acumulada. $P_i = \frac{N_i}{97} \times 100$

Tabla de frecuencias
(¡terminada!):

Variab.	Frec		Acumuladas			
	abs.	Frec rel.	FAA	FRA	Porc.Ac	
Grupo	Países					
(xi)	(ni)	(fi)	(pi)	(Ni)	(Fi)	(Pi)
1	11	1/9	11,3%	11	1/9	11,3%
2	12	1/8	12,4%	23	1/4	23,7%
3	19	1/5	19,6%	42	3/7	43,3%
4	11	1/9	11,3%	53	5/9	54,6%
5	17	1/6	17,5%	70	5/7	72,2%
6	27	2/7	27,8%	97	1	100,0%

Total: 97

Frecuencia relativa



Es posible confeccionar un gráfico de “Torta” con las frecuencias relativas.

La circunferencia total se comprende con un giro de 360° y al multiplicar por el porcentaje de la frecuencia relativa podemos determinar el ángulo del sector circular.

Por ejemplo:

$$p\% f_i = 11,3\% = 0,113 \quad 0,113 \cdot 360 = 40,68$$

Entonces son: 40.68 ° para esta frecuencia.

Utiliza un transportador y la calculadora para confeccionar el gráfico de la izquierda.

¿Cuál crees que es la principal utilidad de este tipo de gráficos? ¿A tu juicio este gráfico, es un buen instrumento descriptivo?

Frecuencia relativa acumulada

La interpretación de las estadísticas acumuladas radica en de toda la muestra de países, en este caso respecto de la variable en estudio sobre pertenencia a los grupos del 1 al 6 y considerando ese orden. Habrá que distinguir entre consultas relacionadas con la frecuencia relativa y la frecuencia acumulada, tales como:

- a) ¿Cuál es la frecuencia de países para un acumulado del 43.3 %?

Esto significa, responder respecto del número de países que están por debajo o igual al 43.3 % del total, la respuesta será entonces $N_3=42$ países.

- b) ¿Cuántos países cubren el 55% de la muestra?
- c) ¿Cuántos países están por sobre el 72,2% del total?

Ejemplo B

Elaborar un gráfico estadístico y calcular la media aritmética, la mediana y la moda de la variable población para cada grupo de países del cuadro que se incluye al final de la unidad.

INDICACIONES AL DOCENTE

La clase puede dividirse en equipos de trabajo para calcular estos promedios y para explicar lo que representa cada uno. Se puede pedir a los alumnos y alumnas que señalen las principales diferencias entre los gráficos y que decidan cuál de los promedios acentúa más las diferencias, explicando la razón.

Parámetros estadísticos: Son datos que resumen el estudio realizado en la población. Pueden ser de dos tipos:

- **Parámetros de centralización.** Son datos que representan de forma global a toda la población. Entre ellos vamos a estudiar la media aritmética, la moda y la mediana.
- **Parámetros de dispersión.** Son datos que informan de la concentración o dispersión de los datos respecto de los parámetros de centralización. Por ejemplo el recorrido (el rango de la variable), la varianza (diferencias al cuadrado respecto de la mediana) y la desviación típica (para obtener el intervalo por donde distribuyen los datos).

En esta ocasión desarrollaremos conceptos para los **parámetros de centralización**.

La Media o Promedio

Para continuar, consideremos la tabla de datos en la columna de Población, por ejemplo, de cada país del grupo1.

Grupo 1	
País	Población
Albania	3204
Alemania Este	61337
Bulgaria	9001
Checoslovaquia	15641
Hungría	10587
Polonia	38061
Rumania	23148
URSS	287664
Yugoslavia	23707

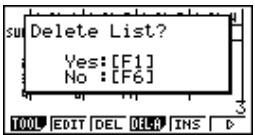

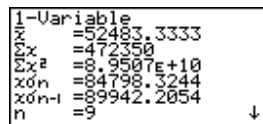
En cada grupo de países podemos obtener parámetros de centralización. Por grupo hacer el mismo trabajo en diferido, con el resto del curso.

Comenzaremos por obtener la **Media** (o promedio), denotada \bar{x} . Muy utilizada para obtener un valor de población que representa al grupo.

$$Media\ G1 = \frac{1}{N_{G1}} \cdot Suma_{G1} = \frac{1}{N_{G1}} \cdot \sum_{i=1}^{N_{G1}} población_i$$

Cualquiera de estas notaciones intenta señalar “la sumar todos los números de población y a este resultado dividirlo por la cantidad total de países del grupo”.

Para obtener la Media (\bar{x}) con la calculadora se procede: (MENU 2)

<p>Si avanzas en el menú de estadística, con F4 DEL-A puedes borrar la columna</p>  <p>Presiona F1</p>	<p>Luego ingresar las poblaciones de cada país del grupo</p> 	<p>Por último puedes pedir todas las estadísticas con las teclas: F2 CALC y F1 1VAR</p> 
--	--	---

Observamos que $\bar{x} = 52483,3333 \dots$ Habrá que notar, que ningún país tiene este número de población, sino que es un valor representativo del grupo. Es decir la mayoría de los países del grupo1 tienden a tener una población del orden de 52483 miles de habitantes.

La Mediana

La mediana consiste en una medida central respecto de clases ordenadas, corresponde al valor de los datos que están precisamente al centro de los datos. Ordenamos los datos de menor a mayor y en forma longitudinal elegimos el valor que está en el centro. En muchos casos es mejor usar como estadístico de tendencia central la mediana que la media, ya que la mediana describe mejor los valores al 50% de los datos.

Por ejemplo: Si estudiamos la esperanza de vida de hombres y mujeres del grupo2. Podemos requerir la mediana de la Esperanza de vida en mujeres.

País	Grupo	Esperanza vida hombre	Esperanza vida mujer
Argentina	2	65,5	72,7
Bolivia	2	51,0	55,4
Brasil	2	62,3	67,6
Colombia	2	63,4	69,2
Chile	2	68,1	75,1
Ecuador	2	63,4	67,6
Guayana	2	60,4	66,1
México	2	62,1	66,0
Paraguay	2	64,4	68,5
Perú	2	56,8	66,5
Uruguay	2	68,4	74,9
Venezuela	2	66,7	72,8

A partir de la tabla de la izquierda, ingresamos la última columna en la calculadora. Utilizamos nuevamente MENU-2. Si tienes datos del trabajo anterior puedes borrar las columnas con DEL-A o F4. Ingresamos los datos.

List 1	List 2	List 3	List 4
MUJ			
10	66.5		
11	74.9		
12	72.8		
13			

Esta lista puede ser ordenada con TOOL y luego SRT-A:

List 1	List 2	List 3	List 4
MUJ			
10	72.8		
11	74.9		
12	75.1		
13			

Ahora, ¿Cuál es la mediana? Son 12 los de datos ordenados. Respuesta: La mediana va a estar entre los valores en posición 6 y 7, los que corresponden a 67.6 y 68.5 obsérvalos en la calculadora.

Observación: Al ser una secuencia con N= par número de datos, debemos considerar los dos datos del centro, si hubieran sido 11 o un número N=impar no tendríamos problema en elegir el dato del centro. En nuestro caso se utiliza como valor mediano al promedio de estos dos, es decir $(67.6 + 68.5)/2 = 68.05$

Este dato Med=68.05 lo podemos obtener con la calculadora, con solo apretar dos teclas, primero EXIT para llegar a...

	List 1	List 2	List 3	List 4
SUB	MUJ			
5	67.6			
6	67.6			
7	68.5			
8	69.2			
				67.6

Luego CALC y 1VAR, con flecha abajo

1-Variable	
x̄n	=5.0856552 ↑
x̄n-1	=5.31179362
n	=12
minX	=55.4
Q1	=62.5
Med	=68.05 ↓

Ejemplo C

En los gráficos siguientes se representa la esperanza media de vida en hombres y mujeres, para los 97 países.

INDICACIONES AL DOCENTE

Es necesario observar que el cálculo de los años de esperanza media de vida al nacer, para los 97 países, es un promedio ponderado.

Otra situación, muy común que se presenta, es determinar las medidas de centralización con los datos en rangos o intervalos. En este caso aparece la frecuencia por intervalo.

Por ejemplo, en la tabla de datos anterior. Observamos en los hombres esperanzas de vida que fluctúan entre 50 a 70 años. Con intervalos de igual ancho, llamamos c_i a este ancho, igual para todos $c_i=5$. Además consideremos un valor representativo del intervalo: 52.5 ; 57.5 ; 62.5 ; 67.5 respectivamente.

Contando los países cuando pertenecen a alguno de los intervalos, obtenemos:

Intervalo	Rep. de clase	Frec frecuencia	Frec Acum	% FA
50-55	52,5	1	1	8%
55-60	57,5	1	2	17%
60-65	62,5	6	8	67%
65-70	67,5	4	12	100%

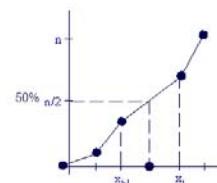
Es así que la Mediana será el valor central. Es decir el que está al 50% de los datos.

Pero la situación ha cambiado, podemos utilizar una fórmula que considera los valores extremos del intervalo; o simplemente el valor de Representación de la clase x_i y su frecuencia f_i para obtener la mediana cuando %FA supera el 50%.

La fórmula que obtiene la mediana considerando extremos de intervalos es:

$$\text{Mediana} = x_{i1} + \left(\frac{(N_M/2) - N_{i-1}}{f_i} \right) \cdot (x_{i2} - x_{i1})$$

Su interpretación geométrica:



Como se trata de rangos, primero debemos encontrar el intervalo donde debiera estar la mediana, en este caso la frecuencia acumulada porcentual ayuda a decir que es el intervalo con el 67%, este es el intervalo de la media, llamada M.

Además en la fórmula identificamos $x_{i1}=60$ y $x_{i2}=65$ y donde $(x_{i2}-x_{i1})=c_i=5$ corresponde al ancho o amplitud del intervalo. $N_M=8$ es la frecuencia acumulada del intervalo M; $N_{i-1}=2$ es la frecuencia acumulada anterior al intervalo M; $f_i=6$ es la frecuencia relativa del intervalo M.

En este ejemplo, el resultado es: $Mediana = 60 + \left(\frac{(8/2) - 2}{6} \right) \cdot 5 = 61.667$

Ahora con calculadora podemos obtener un valor aproximado, usando sólo las representaciones de clases y su frecuencia. En MENU-2 puedes borrar con F6; DEL-A.

Luego ingresa la Rep.de clase y la frecuencia:

Sub	Hom	frec
1	52.5	1
2	57.5	1
3	62.5	6
4	67.5	4

- CALC - SET -

1Var XList	:List1
1Var Freq	:List2
2Var XList	:List1
2Var YList	:List2
2Var Freq	:1

- EXIT-1VAR-

1-Variable	
x̄n-1	=4.50168318 ↑
n	=12
minX	=52.5
Q1	=62.5
Med	=62.5
Q3	=67.5 ↓

Resultando Med = 62.5. Debemos considerar que el resultado más certero corresponde al usar la fórmula, sin embargo mientras estemos usando intervalos siempre tendremos aproximaciones al valor real, que se obtiene con los datos sin intervalos. Es cuestión de realizar el cálculo con los datos verdaderos ingresados en la lista 3 (Liste 3), en este caso obtenemos:

Sub	Hom	frec
1	52.5	1
2	57.5	1
3	62.5	6
4	67.5	4

- 1VAR - 2VAR - REG -

1Var XList	:List3
1Var Freq	:1
2Var XList	:List1
2Var YList	:List2
2Var Freq	:1

- LIST -

1-Variable	
x̄n-1	=4.92552873 ↑
n	=12
minX	=51
Q1	=61.25
Med	=63.4
Q3	=66.1 ↓

Mediana = 63.4

Entonces, las tablas con intervalos son convenientes para resumir grandes cantidades de datos.

Enfoque geométrico para el cálculo de la Moda

Técnicamente, la Moda en Estadística consiste en señalar cuál es el valor, de entre un conjunto de datos, el que más se repite.

Por ejemplo si un médico decide caracterizar a 50 pacientes respecto de sus edades, los puede agrupar en intervalos de a 5 años, si estos comienzan de 20 años se pueden agrupar entre 20 y 25, luego entre 25 y 30 y así sucesivamente. De esta manera, podrá contar cuantos de los pacientes están en estos intervalos, entre 20 y menor a 25 años, luego entre 25 y menor a 30 años, etc. A estos serán nuestras frecuencias de intervalos.

A continuación, la tabla, con 6 intervalos y sus respectivas frecuencias:

i	Edad (LI – LS)	Frecuencia (n)
1	20 – 25	6
2	25 – 30	8
3	30 – 35	10
4	35 – 40	15
5	40 – 45	8
6	45 – 50	3

Notación, de otra literatura:

i = número del intervalo. **LI** = Límite inferior del intervalo.
n = frecuencia o repeticiones. **LS** = Límite superior del intervalo.

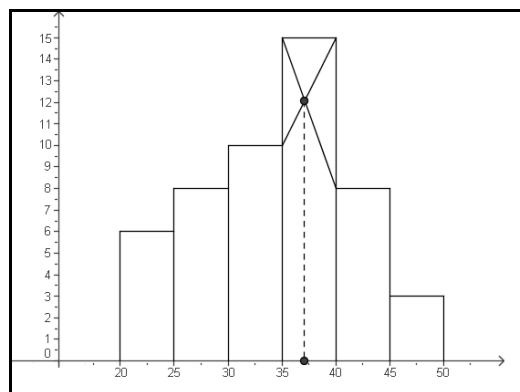
En este gráfico se observa que el cuarto intervalo, entre 35 y 40 años, hay 15 pacientes, la mayor frecuencia. La fórmula estadística para la moda con intervalos corresponde a:

$$M_o = LI_i + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} \cdot c_i$$

Nuevamente c_i corresponde a la amplitud (tamaño) del intervalo donde está la mayor frecuencia.

La fórmula de la Moda en este caso se ve tan compleja como la de la Mediana, el siguiente ejercicio nos permitirá recordarla fácilmente, haremos una interpretación geométrica de la Moda.

Aunque las edades de los pacientes dentro del intervalo con mayor frecuencia pueden estar concentrados en los 39 años, estos están siendo representados por un intervalo entre 35 – 40, la inferencia estadística tomará en cuenta entonces solo el intervalo. La Moda debe ser un único valor, es por ello que sin conocer la forma de cómo se distribuyen los datos, se acepta una razón proporcional entre la diferencia de frecuencias de los intervalos con las frecuencias contiguas.



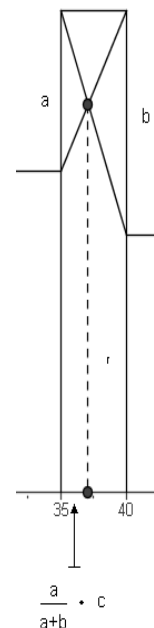
Geoméricamente podemos observar que las diferencias de frecuencias se expresan por segmentos, sus diagonales permiten dividir el segmento del intervalo $35 - 40$ en la misma proporción de esta diferencia. Entonces a 35 le debemos sumar la primera proporción respecto de la amplitud del intervalo.

Es decir, la amplitud es de $c_i = 40 - 35 = 5$, entonces a 35 (el primer valor del intervalo) se suma una proporción de 5 respecto de la razón $\frac{a}{a+b}$, esto se obtiene multiplicando esta razón por 5.

El valor de $LI_i=35$; $a = 15 - 10 = 5$; $b = 15 - 8 = 7$. Entonces $\frac{a}{a+b} = \frac{5}{5+7}$

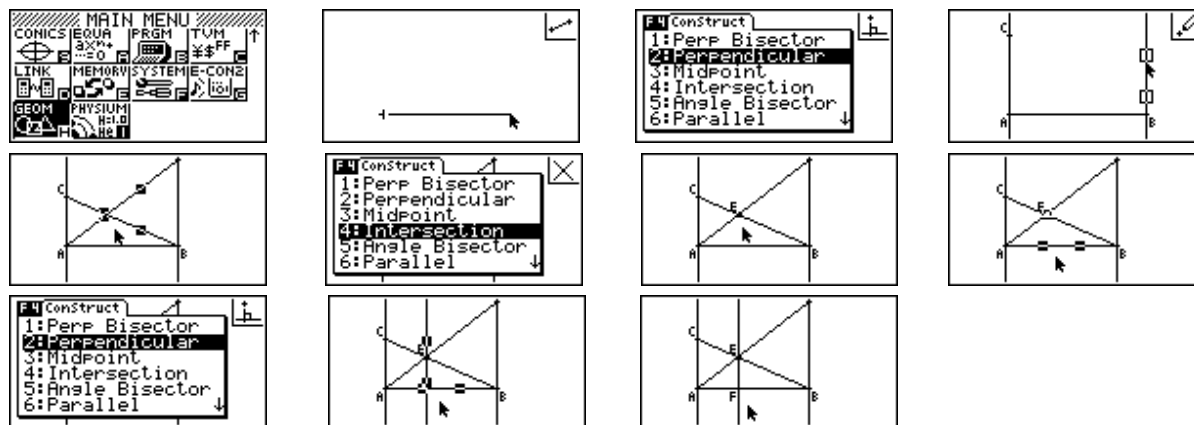
En definitiva la moda se obtiene aplicando la fórmula:

$$35 + \frac{5}{5+7} \cdot 5 = 35 + 2.083 = 37.083$$



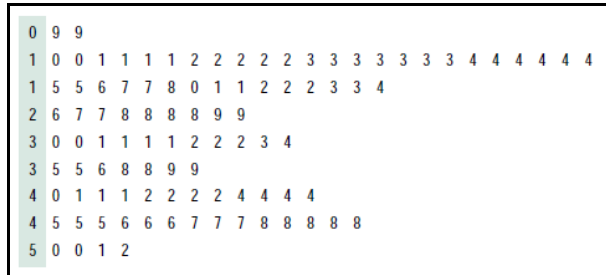
Esta situación numérica obedece a un estudio geométrico, el cual se relaciona con semejanza de triángulos. La Moda está contenida en el intervalo entre 35 y 40 años. Si representamos geoméricamente la situación, desde un segmento de largo 5 unidades dos perpendiculares en sus extremos, de largos respectivos a las diferencias de frecuencias relativas $a=5$ y $b=7$.

Luego:



Ejemplo D

El gráfico que sigue corresponde a las tasas de natalidad de los 97 países; el tallo está definido por la cifra de las decenas del indicador y las hojas por la cifra de las unidades, sin hacer aproximación.



Construir un gráfico similar al anterior, para cada grupo de países y caracterizar cada grupo de acuerdo a este indicador.

INDICACIONES AL DOCENTE

A partir de las gráficas realizadas por los estudiantes, se puede investigar qué países tienen una tasa de natalidad atípica respecto a su grupo. Esta información se puede complementar con gráficos de caja (también conocidos como Cajas y Bigotes, box and whiskers):

Tallos, Hojas y Gráficos de caja

Iniciemos la actividad con un diagrama de Tallos y Hojas para el grupo 3. Esta es una técnica que representa rápidamente todos los datos.

País	Grupo	Tasa natalidad
Alemania (Oeste)	3	11,4
Austria	3	14,9
Bélgica	3	12,0
Canadá	3	14,5
Dinamarca	3	12,4
España	3	10,7
Finlandia	3	13,2
Francia	3	13,6
Grecia	3	10,1
Holanda	3	13,2
Irlanda	3	15,1
Italia	3	9,7
Japón	3	9,9
Noruega	3	14,3
Portugal	3	11,9
Suecia	3	14,5

Si volvemos a anotar la tasa de natalidad de los 19 países, estos número redondeados (o aproximando el decimal) quedarían en una lista como sigue: 11; 15; 12; 15; 12; 11; 13; 14; 10; 13; 15; 10; 10; 14; 12; 15; 13; 14; 17.

Podemos elegir solo un número de tallo, el 1, ya que todos son decenas. Las unidades de cada decena serán las hojas.

Tallo	Hojas
1	1 5 2 5 2 1 3 4 0 3 5
1	0 0 4 2 5 3 4 7

Luego los ordenamos y ya!, terminamos el diagrama.

Tallo	Hojas
1	0 0 0 1 1 2 2 2 3 3 3
1	4 4 4 5 5 5 5 7

Si se ve muy fácil..., estás en lo cierto. Hacemos lo mismo con el grupo 4 para comparar. Entonces sobre los datos: 42; 28; 23; 43; 43; 22; 39; 27; 32; 46; 29.

Suiza	3	12,5
U.K.	3	13,6
U.S.A.	3	16,7

Completa el diagrama:

Hojas Grupo4	Tallo	Hojas Grupo3
	1	0 0 0 1 1 2 2 2 3 3 3
	1	4 4 4 5 5 5 5 7
	2	
	3	
	4	

Efectivamente, los datos así resumidos ofrecen mucha información. Además está la posibilidad de relacionarlos con gráficos de cajas.

Para esto, recordemos a la Mediana como valor de centralización el cual nos indicaba aquel valor que estaba al 50% de los datos. Pues, de igual manera, podemos requerir del valor que está al 25% de los datos, o el valor del 75% de los datos. El conjunto de estos se llaman Cuartiles Q1, Q2, Q3 y Q4, estos número que representan los datos ordenados al 25%, 50%, 75% y 100% respectivamente. Es decir, un gráfico descriptivo puede ser: el mínimo valor (el más pequeño); luego entre Q1 y Q3 se grafica una caja con Q2 (la mediana) trazada al centro; y el máximo valor Q4.

Usemos la calculadora. MENU-2; F6 DEL-A y borrar la tres listas del ejercicio anterior. Continuamos con:

<p>a) En la primera lista ingresar los datos del grupo 4.</p>		<p>b) Luego con la secuencia: EXIT; GRPH; SET y luego seleccionamos GPH2, aquí en la segunda línea y con F6 (o flecha derecha) seleccionamos Box.</p>	
<p>c) Luego EXIT y GPH2, logramos realizar la gráfica</p>		<p>d) Observa qué ocurre si ejecutas F1 1VAR. Tendrás las siguientes estadísticas (flechas abajo).</p>	

Tallo	Hojas Grupo4
2	2 3 7 8 9
3	2 9
4	2 3 3 6

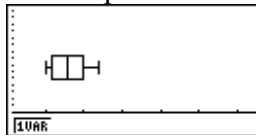
Ahora volvamos al gráfico de Tallo-Hoja del grupo 4.

- Las estadísticas señalan que la Mediana es 32, los datos son impares luego esto es fácil de comprobar, en las hojas están los 11 datos, el valor del 50% está en el tallo 3 primera hoja, es efectivamente el 32. En el gráfico de caja la mediana está desplazada a izquierda.
- Además, Q1(25%) es el 27, y Q3(75%) es el 43.
- Ya sabemos que el mínimo es el 22 y el máximo es 46

Tallo	Hojas Grupo3
	0 0 0 1 1 2 2 2 3 3
1	3
1	4 4 4 5 5 5 5 7

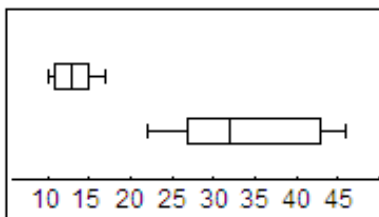
(Q4).

Cambiar la Lista1 a Lista2 con los datos del Grupo3 y graficar para comparar.



```

1-Variable
n      =13
minX   =10
Q1     =11
Med    =13
Q3     =15
maxX   =17
  
```



Luego de esta manera, grafica en una sola tabla ambos gráficos en tu cuaderno, con sub-intervalos de 5 en 5. También prueba con SEL:

```

StatGraph1 :DrawOn
StatGraph2 :DrawOn
StatGraph3 :DrawOff
  
```

Podrás notas las diferencias de los grupos 3 y 4. Observa esta situación con el comando SEL de GRPH

Entonces ¿Cuál es la importancia entonces del uso de los gráficos de caja?

En particular, los gráficos de caja vinculan los conceptos de mediana, cuartiles, valor mínimo y máximo que se manejan individualmente pero no en forma global.

Algunas preguntas podrían ser las siguientes:

¿Qué porcentaje de los datos está representado por la caja?, ¿Qué porcentaje representa cada uno de los bigotes?, ¿Puede ser un bigote más largo que otro?. ¿Cuál es el significado?, ¿Se encuentra la mediana siempre en el centro de la caja?

Respuestas: Los bigotes tienen un límite de prolongación, de modo que aquellos valores atípicos que se separan del cuerpo principal de datos se indican individualmente. A diferencia de otros métodos de presentación de datos, los gráficos de caja muestran los valores atípicos de la variable. Llamaremos **valores atípicos** de la variable a aquellos que están tan apartados del cuerpo principal de los datos que bien pueden representar los efectos de causas extrañas, como algún error de medición o registro. Su eliminación no se justifica, ya que el propósito del gráfico de caja consiste en brindarnos un mayor conocimiento de la forma en que se distribuyen los datos.

Tukey (citado por Hildebrand, 1997) introduce un criterio para fijar los extremos de los bigotes. Para esto calcula 4 barreras, dos interiores y dos exteriores. Considere primero que RIC (Recorrido Intercuartílico) es igual a la diferencia entre el Tercer cuartil y el Primero: $RIC = Q3 - Q1$ (ancho de la caja).

Barrera interior inferior = Primer cuartil - 1,5 . RIC
 Barrera interior superior = Tercer cuartil + 1,5 . RIC
 Barrera exterior inferior = Primer cuartil - 3 . RIC
 Barrera exterior superior = Tercer cuartil + 3 . RIC

```

StatGraph1
Graph Type :MedBox
XList     :List1
Frequency  :1
Outliers  :On
  
```

Los datos que están fuera de estas barreras son valores atípicos. Prueba en la calculadora GRPH-SET la opción On para Outliers.

Ejemplo E

Según estudios realizados por la FAO, la disponibilidad de agua por persona ha descendido bruscamente en un lapso de aproximadamente 50 años.

La siguiente tabla señala la disponibilidad de agua en miles de metros cúbicos:

		1950	2000
1	ÁFRICA	17,8	4,8
2	ASIA	7,6	2,9
3	EUROPA	5,9	4,5
4	AMÉRICA NORTE	32,4	17,6
5	AMÉRICA LATINA	72,1	22,8
6	EX URSS	24,1	14,8
7	OCEANÍA	159,5	65,6

Fuente: FAO.

- I. Hacer el gráfico de barras que permite comparar la disponibilidad de agua en ambos años.
- II. Calcular el porcentaje de descenso para cada región.
- III. ¿Qué explicación le daría Ud. a estos descensos en la cantidad de agua per cápita?
- IV. ¿Por qué cree Ud. que en algunos continentes o regiones este descenso es mayor que en otras?
- V. ¿Qué cree Ud. que sucederá en los próximos 50 años con respecto al agua disponible per cápita en el mundo? (seguirá disminuyendo, se mantendrá o subirá).
- VI. Conversar sobre la siguiente aseveración: “Las futuras guerras serán por el control de las fuentes de agua”.

INDICACIONES AL DOCENTE

El desarrollo de esta actividad es una invitación a interpretar la información y reflexionar sobre las reservas de agua dulce. Este ejemplo se puede complementar considerando la siguiente información.

Estos valores son promedios nacionales lo que significa que pueden existir marcadas diferencias al interior de cada país.

Actividad con la calculadora

Según las instrucciones de este ejemplo, primero procedemos a graficar los datos. Ingresamos en Lista1 solo números consecutivos, luego en Lista2 la columna de disponibilidad de agua en 1950 y en Lista3 la del año 2000.

Graficando los datos:

<p>a)</p>	<p>b) GRPH-SET-GPH1</p> <p>EXIT-SET</p>
<p>c)</p>	<p>d) EXIT</p> <p>GPH1</p>

Con estas instrucciones obtendrás un gráfico...

Pero los gráficos pueden quedar muy distorsionados, es por ello puedes cambiar el tamaño de la

ventana de visualización con SHIFT-F3

Nuevamente GPH1

GPH1

SHIFT-Trace(F1)

Ahora graficamos los datos del 2000 con GPH2. Claramente disminuidos.

Tenemos una opción más, graficar juntas las dos variables:

SEL

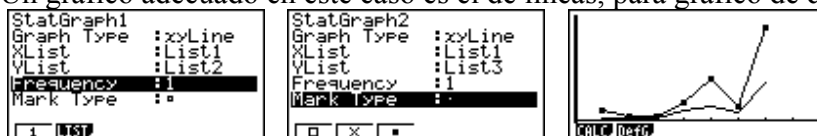
Procedemos a compararnos. Desde aquí

Cambiamos a

luego graficamos ambos casos con DRAW

El gráfico muestra diferencias, pero no es tan visible. Probemos con otro tipo de gráficos.

Un gráfico adecuado en este caso es el de líneas, para gráfico de dispersión. Gráfico xyLine.

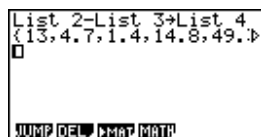


Diferencias porcentuales

En nuestro ejercicio las diferencias son la clave. Es por ello que procedemos a obtener estas diferencias con una sola instrucción entre listas.

a) Primero creamos una lista con las diferencias:

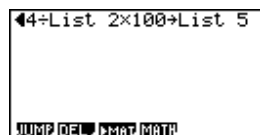
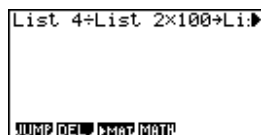
MENU-1 borramos con DEL; DEL-A (ALL);
SHIFT-List 2 luego SHIFT-List 3 asignamos a SHIFT-List4



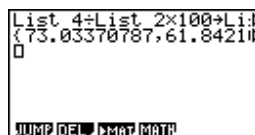
ahora volvamos a MENU-2

	List 1	List 2	List 3	List 4
SUB	1	1950	2000	13
1	1	17.6	4.8	13
2	2	7.6	2.9	4.7
3	3	5.9	4.5	1.4
4	4	32.4	17.6	14.8

De igual manera calculamos los porcentajes en la Lista5, ingresando:



EXE



Volvemos:

	List 2	List 3	List 4	List 5
SUB	1950	2000		
1	17.6	4.8	13	73.0337
2	7.6	2.9	4.7	61.8421
3	5.9	4.5	1.4	23.7228
4	32.4	17.6	14.8	45.6791

	List 2	List 3	List 4	List 5
SUB	1950	2000		
5	72.1	22.8	49.3	68.3771
6	24.1	14.8	9.3	38.5889
7	159.5	65.6	93.9	58.8711
8				

Hemos logrado entonces tener las diferencias y los porcentajes de descenso respectivos.

b) Conclusiones

Por último, si relacionamos estos datos con el Agua dulce usada diariamente por persona en labores domésticas y supondremos que la disponibilidad de agua es en alguna medida un factor cultural que incide en el consumo. Es por ello que:

Senegal (de Africa)..... 30 litros
Chile (de América Latina) 300 litros
EEUU (de América del Norte) 700 litros

Datos promedios per cápita de consumo al año, señalados con anterioridad.

Este consumo por persona cambiará drásticamente. Si las situaciones climáticas continúan en 50 años más, para el 2050 Inevitablemente los habitantes de:

- Senegal, deberán consumir $30*(1-0.73)=8.1$ litros de agua.
- Chile, consumirán $300*(1-0.68)=96$ litros de agua.
- EEUU, consumirán $700*(1-0.59)=287$ litros de agua.

Ejemplo F

Según EMOS (Empresa Metropolitana de Obras Sanitarias), el consumo promedio de agua, en m^3 , en una familia de 5 integrantes es:

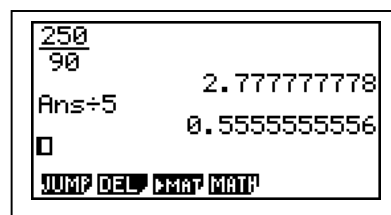
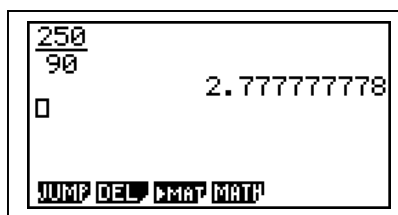
Uso	invierno	verano
Duchas	250	350
Aseo en lavatorios	50	60
Descargas WC	300	300
Comidas y lavado vajilla	80	90
Lavado general	150	185
Riego	5	165
Total diario	835	1.150
Total mensual	25.050	34.500

Fuente: Emos. Interpretar esta información y responder las preguntas siguientes:

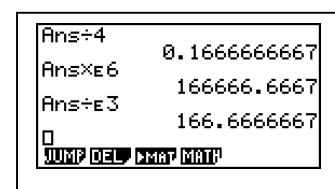
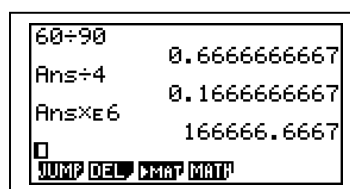
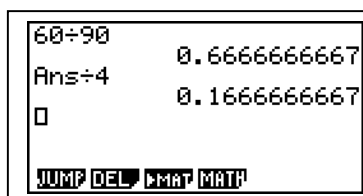
- I. ¿Qué indican y qué ocultan estos valores promedios?
- II. ¿Que significaría dividir por 5 cada uno de los valores de la tabla dada?
- III. Analizar el tema de la escasez de agua y su mal uso.
- IV. Investigar en la propia realidad familiar y encontrar indicadores que permitan compararlos con los datos anteriores.

Solución:

- I. Por ejemplo, los 250 m³ de agua consumidos por el ítem ducha en invierno, tan solo indica el volumen total empleado en duchas durante tres meses del año (21 de junio al 20 de septiembre). Si consideramos aproximadamente un total promedio de 90 días durante el invierno, podríamos obtener el consumo promedio diario en duchas, el cual aparece en la ventana siguiente y es de casi 2,8 m³ diarios. Ahora, si dicho resultado se divide por 5, obtenemos el consumo promedio diario de m³ de agua utilizado por cada integrante (0,56 m³), suponiendo que cada uno de ellos se dio una ducha diaria durante cada uno de los 90 días del invierno. Dado que no hay información respecto a si todos los integrantes se ducharon diariamente durante dicho período, es que carece de validez el promedio de 0,56 m³ de agua consumidos diariamente por cada individuo. Tampoco sabemos si hay personas que se dan duchas más largas en contraposición con los que lo hacen en tiempos muy breves, por lo que el consumo por persona variaría.



Respecto a los 300 m³ de descarga del WC, tan solo indica que tanto en invierno como en verano se mantiene el mismo número de integrantes en la familia y dado que el estanque llena a una misma capacidad independientemente del número de descargas, entonces, sería fácil calcular el total de agua en cada descarga por persona. Suponiendo un total de 4 descargas diarias por persona, cada descarga sería de aproximadamente 0,17 m³ lo que traducido a cm³ sería un total de 166.666 cm³, es decir, 167 litros por descarga. Este último valor no tiene sentido, ya que un estanque de WC tiene aproximadamente 7 litros de capacidad. Lo más probable es que existan más descargas, aquellas producidas cuando se realiza la limpieza del WC o al verter algún otro líquido para luego realizar la descarga o simplemente que cada integrante frecuente el WC más de 4 veces diarias.



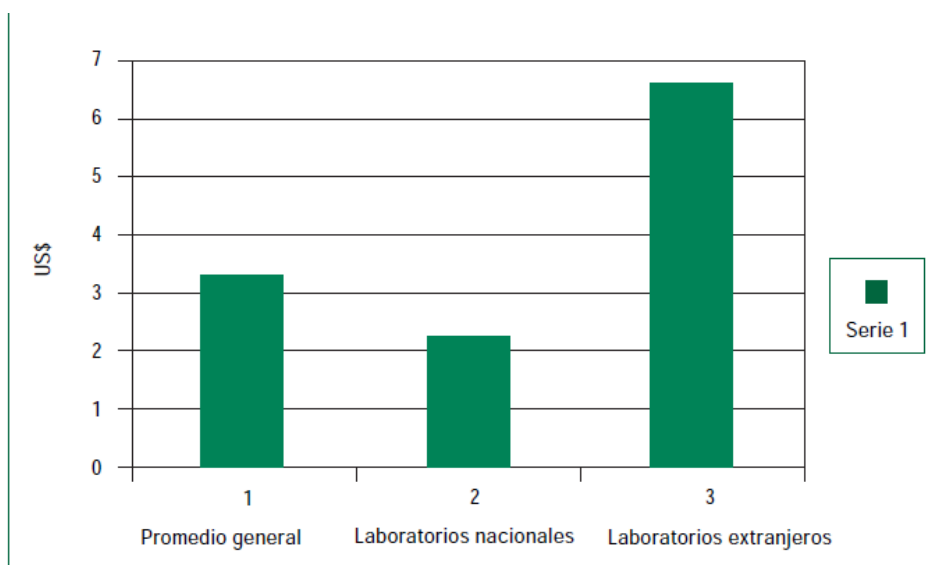
II. Obtendríamos el consumo promedio de agua utilizado por cada individuo durante todo el invierno y el verano en cada uno de los ítemes de la tabla, por ejemplo, en las descargas del WC, sabríamos que cada individuo utiliza no más de 60 m³ de agua en el uso del WC durante todo el invierno, y en el verano es la misma cantidad. En el lavado de vajilla, 16 m³ de agua se utilizan durante todo el invierno para el lavado de la vajilla de un solo integrante de la familia. En el caso del riego, no tiene sentido dividirlo por 5, ya que el volumen de agua utilizado no depende de los integrantes sino del tamaño de área verde del hogar.

III. Para el alumno.

IV. Para el alumno.

Ejemplo G

El gráfico y la tabla que se presentan a continuación corresponden al precio promedio de remedios en Chile según su origen de elaboración.



Fuente: Asociación Industrial de Laboratorios Farmacéuticos de Chile. (4º trimestre de 1999).

Laboratorios	Promedio US\$
Nacionales	2,30
Extranjeros	6,54
Promedio general	3,30

- I. Interpretar el significado de “precio promedio” de los remedios en Chile.
- II. Explicar por qué el promedio general de los remedios no corresponde a la semi-suma de 2,30 y 6,54.

INDICACIONES AL DOCENTE

Aquí se trata de una situación de cálculo del promedio ponderado que se explica por la información siguiente: las ventas totales de esta industria durante el año 2000 en farmacias fueron de 165 millones de unidades valorizadas en US\$ 552 millones. Sobre esa cifra la industria nacional tiene una participación del 77% en unidades vendidas, pero en valores ello representa sólo el 54% del mercado, por lo tanto, el 46% restante corresponde a los laboratorios extranjeros que venden menos cantidad de medicamentos pero a un mayor precio.

Solución

- I. Del total 165 millones de remedios vendidos en Chile durante el año 2000, incluyendo laboratorios nacionales y extranjeros, valorizados en 552 millones de dólares, el precio promedio por remedio en dicho año fue de 3,34 US\$. Esto no significa que la mayoría de los remedios se vendieron a 3,34 US\$ la unidad, sino que en promedio se alcanzó dicho valor. Más aun, es posible que ningún medicamento se haya vendido en dicho valor, dado que los precios por medicamento pudieran haber estado por encima o debajo del precio promedio.
- II. Este problema es similar al expuesto en el ejemplo F de la Actividad 1. Revise dicho ejemplo.

Actividad 3

Ejemplo A

I. Para estimar el número de peces que hay en un lago, se realizó lo siguiente:

- se capturó una muestra al azar de peces, se les marcó y fueron devueltos al agua.
- un breve tiempo después, se capturó una nueva muestra, se registró la proporción de peces marcados versus el total de peces de la muestra.

Si las muestras fueron efectivamente aleatorias, entonces se espera que la frecuencia relativa de peces marcados en la segunda muestra sea aproximadamente la misma que la de peces marcados en la población. Suponer que en el primer proceso se capturan y marcan 120 peces. Posteriormente se capturan 100 peces de los cuales 22 están marcados. Estimar el número de peces del lago.

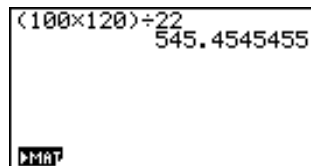
II. Aplicar el procedimiento anterior para estimar el número de bolitas que hay en una bolsa.
Solución:

I. Podemos estimar el número total de peces mediante una proporción entre el número total de peces y los peces marcados.

Supongamos que n es el número total de peces del lago, entonces:

$$\frac{n}{120} = \frac{100}{22}$$

Calculando n



```
(100*120)/22
545.4545455
```

Nos da que el número estimado de peces en el lago es aproximadamente de 545.

II. Supongamos que tenemos una bolsa con bolitas azules y amarillas. Si solo se sabe que el número total de bolitas amarillas es de 55, y se hace 5 veces el siguiente experimento se sacan 20 bolitas de la bolsa al azar y se observa cuantas de ellas son amarillas, obteniéndose

	muestra	Bolitas amarillas
Extracción 1	20	14
Extracción 2	20	11
Extracción 3	20	16
Extracción 4	20	8
Extracción 5	20	10

¿Cómo es posible obtener la mejor estimación del número de bolitas azules? Analice el método de estimación planteado.

Solución:

Basándonos en el caso anterior tenemos que si n es el número total de bolitas en la bolsa, se deduce que

	Proporción
Extracción 1	$\frac{n}{55} = \frac{20}{14}$
Extracción 2	$\frac{n}{55} = \frac{20}{11}$
Extracción 3	$\frac{n}{55} = \frac{20}{16}$
Extracción 4	$\frac{n}{55} = \frac{20}{8}$
Extracción 5	$\frac{n}{55} = \frac{20}{10}$

Si llevamos nuestros datos a una planilla Excel de la calculadora, en donde la primera columna representa el número de bolitas amarillas en cada extracción y la segunda columna representa la estimación del total de bolitas con el número de bolitas amarillas correspondiente, se tiene

Número de bolitas amarillas en cada extracción

SHEE	A	B	C	D
1	14	78.571		
2	11	100		
3	16	68.75		
4	8	137.5		
5	10	110		
		=1100÷A5		
1VAR	2VAR	REG		SET

$$\frac{20 * 55}{A1}$$

Utilizando el modo estadística en nuestra planilla Excel

```
1-Variable
x̄ =98.9642857
Σx =494.821428
Σx² =51906.2818
x̄n =24.2348206
x̄n-1 =27.0953531
n =5 ↓
```

Podemos determinar que la mejor estimación del número de bolitas total es de 99 bolitas, de las cuales $99-55=44$ son azules aproximadamente.

Podemos concluir que mientras más extracciones hagamos mejor será nuestra aproximación del número total de bolitas.

Ejemplo B

Seis amigos conversan sobre el número de hermanos que tiene cada uno. Llegan a la información que se resume en la tabla siguiente:

Nº de hermanos	Frecuencia
1	2
2	1
3	2
4	1
Total	6

- I. Calcular el promedio de hermanos del grupo.
 - II. Para experimentar en relación con las muestras, formar todos los dúos de amigos y para cada dúo calcular el promedio de hermanos.
 - III. Hacer el gráfico de la distribución del promedio de hermanos de todas las muestras, calcular la desviación estándar de esta distribución y compararla con el promedio y la distribución estándar del número de hermanos del grupo de amigos.
 - IV. Formar todos los tríos de amigos y proceder a hacer los mismos cálculos.
 - V. Comparar con los resultados obtenidos en relación con los promedios calculados.
 - VI. Constatar la relación $s = \sigma / \sqrt{n}$ en que s es la desviación estándar de la distribución de todas las muestras, σ es la de la población y n es el número de elementos de la muestra.
- Solución:

- I) Calculemos el promedio de hermanos del grupo

	List 1	List 2	List 3	List 4
SUB	N her	Frec		
1	1	2		
2	2	1		
3	3	2		
4	4	1		
			Frec	

GRAPH CALC TEST INTR DIST

1-Variable	
\bar{x}	=2.33333333
Σx	=14
Σx^2	=40
$s\sigma_n$	=1.10554159
$s\sigma_{n-1}$	=1.21106014
n	=6

↓

II) Formemos los dúos de amigos y calculemos para cada uno su promedio de hermanos.
Para esto supongamos que

amigos	Número de hermanos
A	1
B	1
C	2
D	3
E	3
F	4

El número de combinaciones de 6 amigos elegidos de a dos está dado por

$$C_2^6 = \frac{6!}{4! 2!} = 15$$

Luego los dúos de amigos y sus respectivos promedios para cada dúo son:

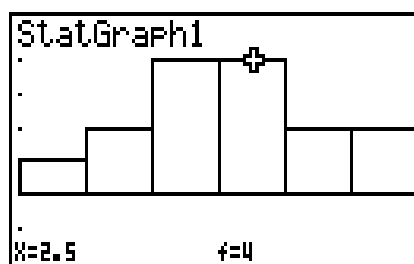
AB (1)	AC (1,5)	AD (2)	AE (2)	AF (2,5)
	BC (1,5)	BD (2)	BE (2)	BF (2,5)
		CD (2,5)	CE (2,5)	CF (3)
			DE (3)	DF (3,5)
				EF (3,5)

III) Ingresems el promedio de hermanos y su frecuencia en una tabla

SUB	List 1 Pr Hér	List 2 Fréc	List 3	List 4
1	1	1		
2	1.5	2		
3	2	4		
4	2.5	4		

SUB	List 1 Pr Hér	List 2 Fréc	List 3	List 4
4	2.5	4		
5	3	2		
6	3.5	2		
7				

Y luego, grafiquemos



La cual tiene desviación estándar y promedio

```

1-Variable
x̄ = 2.33333333
Σx = 35
Σx² = 89
x̄n = 0.69920589
x̄n-1 = 0.72374686
n = 15
    
```

Si comparamos los resultados obtenidos con el promedio y la distribución estándar del número de hermanos del grupo de amigos, tenemos:

	Solos	Dúos
Promedio	2.3333	2.3333
Desviación estándar	1.1055	0.6992

De lo que se deduce que el promedio o media aritmética en este caso es el mismo y la desviación estándar es menor a la original

IV) El número de combinaciones de 6 amigos elegidos de a tres está dado por

$$C_3^6 = \frac{6!}{3! 3!} = 20$$

Los tríos de amigos están dados en la siguiente tabla:

ABC (1.3333)	ACD (2)	ADF (2.6666)	BCF (2.3333)	CDE (2.6666)
ABD (1.6666)	ACE (2)	AEF (2.6666)	BDE (2.3333)	CDF (3)
ABE (1.6666)	ACF (2.3333)	BCD (2)	BDF (2.6666)	CEF (3)
ABF (2)	ADE (2.3333)	BCE (2)	BEF (2.6666)	DEF (3.3333)

Y su frecuencia se representa en

	List 1	List 2	List 3	List 4
SUB	Pr	Hér	Frec	
1	1.3333		1	
2	1.6666		2	
3		2	5	
4	2.3333		4	

1VAR 2VAR REG SET

	List 1	List 2	List 3	List 4
SUB	Pr	Hér	Frec	
4	2.3333		4	
5	2.6666		5	
6		3	2	
7	3.3333		1	

1VAR 2VAR REG SET

Con desviación estándar y promedio

1-Variable	
\bar{x}	=2.3333
Σx	=46.666
Σx^2	=113.774622
$s_{x\bar{x}}$	=0.49441098
$s_{x(n-1)}$	=0.50725496
n	=20

↓

V) Si comparamos los resultados obtenidos con el promedio y la distribución estándar del número de hermanos del grupo de amigos, tenemos:

	Solos	Dúos	Tríos
Promedio	2.3333	2.3333	2.3333
Desviación estándar	1.1055	0.6992	0.4944

De lo que se concluye que la media aritmética es un estimador insesgado, no así la desviación estándar. Además a medida que la muestra aumenta de tamaño la desviación estándar disminuye.

VI)

Verifiquemos la

relación $s = \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}}$, con los datos obtenidos anteriormente

	Solos	Dúos	Tríos
Promedio	2.3333	2.3333	2.3333
Desviación estándar	1.1055	0.6992	0.4944

Para pares de amigos

$$s = \frac{1.10554149}{\sqrt{2}} * \sqrt{\frac{6-2}{6-1}} = \frac{1.10554149}{\sqrt{2}} * \sqrt{\frac{4}{5}} = 0.6992058312$$

```
(1.10554149÷√2)×(√(4÷
5))
0.6992058312

```

Para tríos de amigos

$$s = \frac{1.10554149}{\sqrt{3}} * \sqrt{\frac{6-3}{6-1}} = \frac{1.10554149}{\sqrt{3}} * \sqrt{\frac{3}{5}} = 0.4944131847$$

```
(1.10554149÷√3)×(√(3÷
5))
0.4944131847

```

Ejemplo C

Se dispone de una bolsa con 100 fichas numeradas distribuidas como indica la tabla:

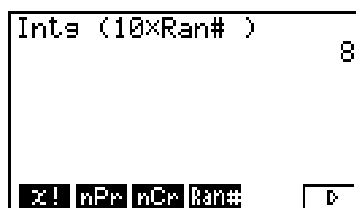
Nº de la ficha	1	2	3	4	5	6	7	8	9	10
cantidad	10	10	10	10	10	10	10	10	10	10

Se pide:

- I. Obtener muestras al azar de tamaño 10 y calcular para cada una de ellas la media de los valores de las fichas como también su desviación estándar.
- II. Obtener muestras al azar de tamaño 20 y calcular para cada una de ellas la media de los valores de las fichas como también su desviación estándar.
- III. Obtener muestras al azar de tamaño 30 y calcular para cada una de ellas la media de los valores de las fichas como también su desviación estándar.
- IV. Comparar los valores de las medias y desviaciones estándar obtenidos en los experimentos anteriores.
- V. Realizar inferencias sobre el valor de la media poblacional a partir de algunas de las muestras anteriores.

Solución:

- I) Para obtener muestras de tamaño 10, usaremos 10 veces la función random de nuestra calculadora



Obteniendo la muestra deseada: 8, 6, 2, 4, 6, 9, 10, 4, 5, 3. Luego organizamos nuestros datos considerando su frecuencia para ingresarlos en una tabla

	List 1	List 2	List 3	List 4
SUB	Dato	Frec		
1	2	1		
2	3	1		
3	4	2		
4	5	1		
2				
[1VAR] [2VAR] [3EQ] [SET]				

	List 1	List 2	List 3	List 4
SUB	Dato	Frec		
5	6	2		
6	8	1		
7	9	1		
8	10	1		
6				
[1VAR] [2VAR] [3EQ] [SET]				

Y obtenemos la media aritmética y su desviación estándar.

```

1-Variable
x̄ = 5.7
Σx = 57
Σx² = 387
x̄n = 2.49198715
x̄n-1 = 2.6267851
n = 10
↓

```

II) Análogamente si la muestra es de tamaño 20, con elección random al azar obtenemos:
2, 5, 9, 6, 4, 5, 7, 4, 8, 10, 6, 9, 3, 1, 5, 6, 3, 8, 9, 10.

	List 1	List 2	List 3	List 4
SUB	Datos	Frec		
1	1	1		
2	2	1		
3	3	2		
4	4	2		
1				
[TOOL] [EDIT] [DEL] [DEL] [INS] [D]				

	List 1	List 2	List 3	List 4
SUB	Datos	Frec		
7	7	1		
8	8	2		
9	9	3		
10	10	2		
2				
[TOOL] [EDIT] [DEL] [DEL] [INS] [D]				

Con media y desviación estándar

```

1-Variable
x̄ = 6
Σx = 120
Σx² = 858
x̄n = 2.6267851
x̄n-1 = 2.69502465
n = 20
↓

```

III) Análogamente si la muestra es de tamaño 20, con elección random al azar obtenemos:
5, 8, 6, 7, 2, 3, 6, 9, 10, 4, 5, 3, 6, 10, 7, 1, 2, 5, 6, 9, 4, 7, 7, 3, 6, 1, 9, 3, 4, 6.

	List 1	List 2	List 3	List 4
SUB	Datos	Frec		
1	1	2		
2	2	2		
3	3	4		
4	4	3		
				2
[IVAR] [ZVAR] [REG]				[SET]

	List 1	List 2	List 3	List 4
SUB	Datos	Frec		
7	7	4		
8	8	1		
9	9	3		
10	10	2		
				2
[IVAR] [ZVAR] [REG]				[SET]

Con media y desviación estándar

1-Variable	
\bar{x}	=5.46666666
Σx	=164
Σx^2	=1088
s_{x_n}	=2.52630604
$s_{x_{n-1}}$	=2.56949388
n	=30
	↓

IV) Comparemos los valores de las medias y desviaciones estándar obtenidos en los tres casos

	n=10	n=20	n=30
\bar{x}	5,7	6	5,466
σ_n	2,491	2,626	2,526
σ_{n-1}	2,626	2,695	2,694

Conclusiones:

- 1) Existe diferente variabilidad entre las medias obtenidas en las muestras y la media del valor de las fichas de la población (promedio =5,5). Esta variabilidad depende solo de la muestra elegida.
- 2) La desviación estándar de los valores obtenidos, se aproxima a la desviación estándar de la población (2,582), a medida de que la muestra va creciendo.

Actividad 4

Ejemplo A

En su columna dominical un periodista plantea la siguiente pregunta a sus lectores: ¿donaría usted sus órganos?

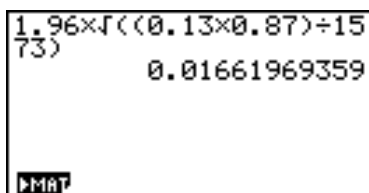
De las casi 5000 cartas recibidas, aproximadamente el 70% dijo que no. Sin embargo, una investigación del Instituto de Estadísticas señala que de 1573 personas encuestadas, un 87% dijo que sí donaría sus órganos.

¿Qué explicaciones da usted a esta diferencia?

Solución:

Supongamos que se cumple que el 70% no donaría sus órganos, entonces podemos suponer que $p = 0,70$, entonces

$$0,13 \pm Z_{0,975} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = 0,13 \pm 1,96 \cdot \sqrt{\frac{0,13 \cdot 0,87}{1573}} = 0,13 \pm 1,96 \cdot \sqrt{\frac{0,1131}{1573}} = 0,13 \pm 0,01661$$



```
1.96*sqrt((0.13*0.87)/1573)=1573
0.01661969359
MATH
```

Luego

$$0,113387 \leq p \leq 0,146619$$

Conclusión:

- 1) Como consecuencia de nuestros cálculos, sabemos que nuestra hipótesis es falsa y que es más factible que el 87% done sus órganos.
- 2) Se debe señalar que las cartas tienen una intencionalidad mientras que la encuesta del INE es de una muestra totalmente al azar, por lo cual tiene un más alto grado de credibilidad.

Ejemplo B

En una encuesta se pregunta, ¿debiera haber una legislación que prohíba el trabajo infantil?
Las respuestas fueron: 43% a favor; 48% en contra; 9% indecisos.

En una segunda encuesta se reformuló esta pregunta: ¿debiera haber una legislación que proteja el derecho a la educación y recreación de los niños?

Las respuestas fueron: 62% a favor; 27% en contra; 11% indecisos.

- I. Analizar las diferencias de las preguntas y su posible incidencia en los porcentajes de respuesta.
- II. Redactar dos o más formas de hacer una pregunta orientada a la búsqueda de la misma información, con las que suponen obtendrían distintos resultados en una encuesta.
- III. Experimentar con las preguntas que se consideren más adecuadas.

Es una pregunta conceptual, no es posible utilizar la calculadora.

Ejemplo C

El candidato A que postula a la presidencia afirma: ya tengo ganada esta elección de acuerdo a la encuesta publicada hoy en los diarios.

B, el otro candidato dice: de acuerdo a los resultados de la encuesta publicada hoy en los diarios, hay claramente un empate.

En los diarios se informa que una encuesta con un 2% de margen de error y alto nivel de confianza indica que para el candidato A hay un 38,7% de intención de voto y para el candidato B, esta intención de voto llega a un 35,3%.

I. Según la encuesta, ¿cuál de los dos candidatos ganaría la elección?

II. ¿Cuál es el significado de 2% de error en los resultados?

Solución:

Ambos candidatos tienen el mismo margen de error que es del 2%, luego

$$\text{Candidato A} \quad 0,387 - 0,02 \leq p \leq 0,387 + 0,02$$

$0.387 - 0.02$	0.367
$0.387 + 0.02$	0.407
▶MAT	

$$\Rightarrow 0,367 \leq p \leq 0,407$$

$$\text{Candidato B} \quad 0,353 - 0,02 \leq p \leq 0,353 + 0,02$$

$0.353 - 0.02$	0.333
$0.353 + 0.02$	0.373
▶MAT	

$$\Rightarrow 0,333 \leq p \leq 0,373$$

Como estos intervalos se interceptan es imposible predecir cuál de los dos candidatos ganaría las elecciones

Actividades para la evaluación y ejemplos

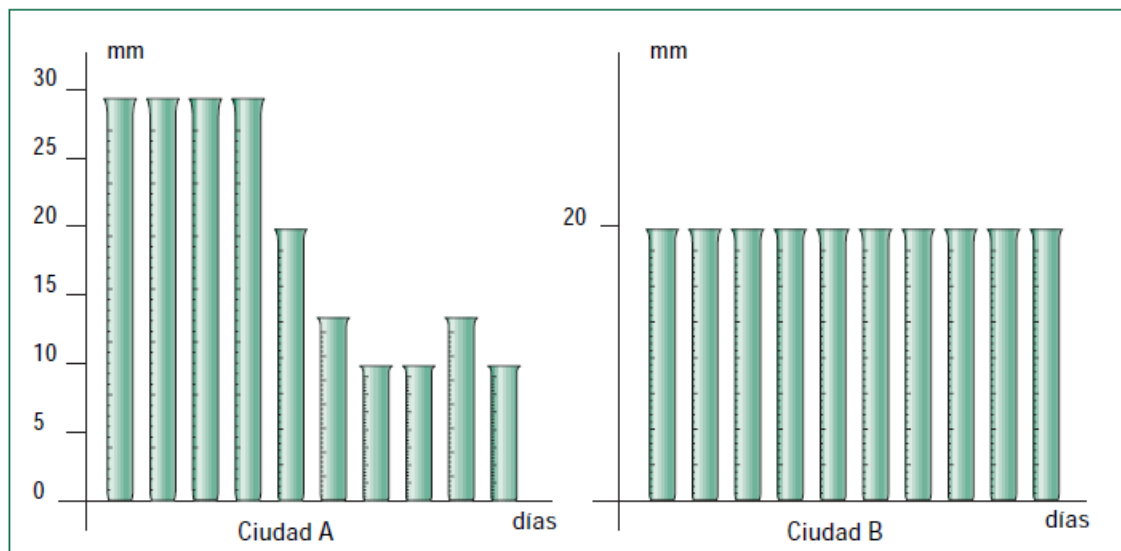
Las actividades que se proponen a continuación se complementan con algunos ejemplos. Para cada ejemplo se propone un conjunto de indicadores que importa tener en cuenta para evaluar el logro de los aprendizajes esperados por el alumno o alumna.

Actividad 1

Analizan distribuciones de datos a partir de gráficos, indicadores de tendencia central y de dispersión.

Ejemplo A

Considerar los siguientes gráficos de barra, que representan los mm de agua caída en dos ciudades diferentes, durante los primeros días de un mes.



I. A partir de los gráficos, ¿qué se puede afirmar en relación con la cantidad de agua caída en esos días en las dos ciudades?

En ambos gráficos se aprecia el mismo promedio de un total de los 10 primeros días en agua caída. Por simple inspección, podemos calcular un promedio de 20 mm de agua caída en el gráfico de la Ciudad A. En el caso del gráfico de la ciudad B, es evidente que no hubo variación diaria en el agua caída ninguno de los 10 días, por lo que el promedio de 20 mm de agua caída está a simple vista. La desviación estándar de la ciudad B es de 0, dado que todos los datos están agrupados en torno de la media. En el caso de la Ciudad A, la desviación estándar debe estar por sobre las 5 unidades, dado que los primeros 4 días el total de agua caída está por encima de 10 unidades respecto a la media y el agua caída en los últimos cuatro días están por debajo de la media.

II. ¿Cuál es la media aritmética de mm de agua caída en cada ciudad?

Observar qué cálculos realizan y cómo éstos reflejan la interpretación de ambos gráficos.

La media de la ciudad A se calcula por simple extrapolación de datos:

$$4 \times 30 + 20 + 15 + 2 \times 10 + 15 + 10 = 200$$

$$\text{Media} = \frac{200}{10} = 20 \text{ mm}$$

La media de la ciudad B es de 20 mm, dado que todos los días hubo la misma cantidad de agua caída.

Ejemplo B

En una reunión de alumnos se escucha el siguiente diálogo:

Juan: Mi curso obtuvo promedio 5 en lenguaje.

Andrea: El mío obtuvo promedio 5,1, por lo tanto es un curso mejor.

Mateo: Momento, ¿me pueden dar la desviación estándar de las notas de cada curso?

Juan: Al calcularla resulta 0,2.

Andrea: A mí me dio 0,6.

De acuerdo a estos datos, ¿cuál es su opinión respecto a ambos cursos?

Observar si relacionan los indicadores (media aritmética y dispersión) para fundamentar la opinión.

Solución:

Diseñaremos dos listas de notas. La lista 1, con las notas del curso de Juan y la lista 2, con las notas del curso de Andrea. Cada lista tiene un total de 15 notas. La calculadora permite ir calculando la media y desviación estándar a medida que vamos ingresando notas. Es posible ir probando qué notas son las que cumplirán con las condiciones expuestas en el enunciado. Las tablas siguientes muestran las notas que satisfacen las condiciones de media y desviación estándar para los cursos de Juan y Andrea.

	List 1	List 2	List 3	List 4
SUB	Juan	Andrea		
1	5	6		
2	4.9	4.5		
3	5.1	3.8		
4	4.8	5.6		

	List 1	List 2	List 3	List 4
SUB	Juan	Andrea		
5	5.1	5.5		
6	4.9	4.3		
7	5.4	5.3		
8	4.7	4.5		
				4.7

	List 1	List 2	List 3	List 4
SUB	Juan	Andrea		
9	5.2	5.5		
10	4.8	4.9		
11	5.1	5.7		
12	5.4	4.3		
				5.4

	List 1	List 2	List 3	List 4
SUB	Juan	Andrea		
13	5	5.2		
14	5.2	5.2		
15	5	5.9		
16				

A continuación calculamos la media y desviación estándar para cada curso

```

1Var XList :List1
1Var Freq  :1
2Var XList :List1
2Var YList :List2
2Var Freq  :1
LIST
    
```

```

1 variable
x̄      =5.04
Σx     =75.6
Σx²    =381.62
x̄σn    =0.19933221
x̄σn-1  =0.20632844
n      =15
    
```

```

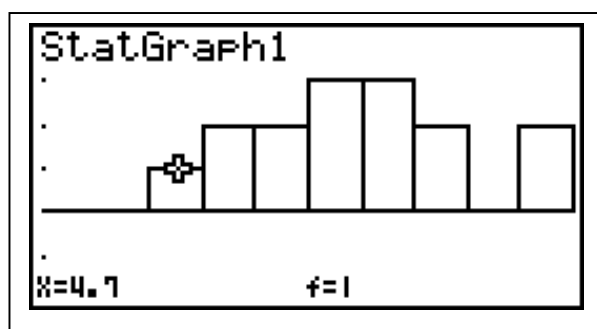
1Var XList :List2
1Var Freq  :1
2Var XList :List1
2Var YList :List2
2Var Freq  :1
LIST
    
```

```

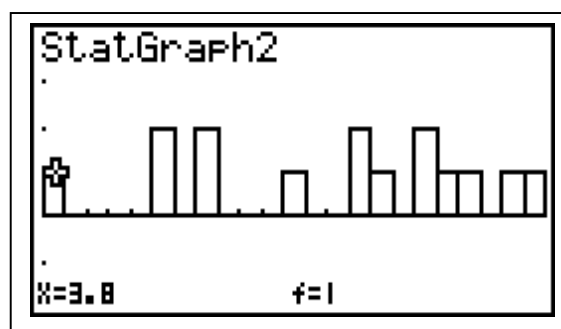
1 variable
x̄      =5.08
Σx     =76.2
Σx²    =393.26
x̄σn    =0.64104082
x̄σn-1  =0.66354028
n      =15
    
```

El curso de Juan tiene una media de 5,04, es decir, se aproxima a 5,0 y una desviación de 0,199 \approx 0,2. El curso de Andrea tiene una media de 5,08 \approx 5,1 y una desviación de 0,64 \approx 0,6. Observemos los gráficos de barras para cada curso.

Curso de Juan



Curso de Andrea



Las barras están más dispersas en el curso de Andrea, esto significa que las notas son heterogéneas respecto a la media y el rendimiento no es parejo en general. En cambio el curso de Juan, muestra una clara tendencia de notas homogéneas y cercanas a la media. Dado que la desviación estándar en el curso de Andrea es tres veces menor que la del curso Juan y las medias difieren en tan solo una décima, podemos concluir que el rendimiento del curso de Andrea es más parejo y mejor.

Ejemplo C

Según el Informe sobre Desarrollo Mundial (1994) las áreas naturales protegidas a nivel mundial representan, en promedio, el 5,4% de la superficie de los países. Chile, en cambio, cuenta con un 18,3% de su superficie total protegida.

(<http://www.infor.cl/webinfor/producserv/inforestad/estadisticas.htm>)

Pais	Superficie protegida (% de superficie del pais)
Brasil	3,3
Uruguay	0,2
México	5,1
Argentina	3,4
Puerto Rico	4,0
Nueva Zelanda	10,7
España	6,9
Canadá	5,0
Irlanda	0,6
Francia	9,8
EEUU	10,5
Australia	10,6
Singapur	2,6
Finlandia	2,5
Noruega	5,0
Suecia	6,6
Promedio	5,4
Chile	18,3

De acuerdo a estos datos, responder las siguientes preguntas:

- ¿Está el dato de Chile incluido en el cálculo del promedio?
Fundamentar la respuesta.
- ¿Cómo interpreta el valor 5,4%?
- Observando los valores de la tabla anterior, ¿le parecen a Ud. muy dispersos? Explicar.
- Representar gráficamente los datos de la tabla.

Observar qué cálculos realizan, si hacen referencia al referente del % y cómo explican el que corresponde al promedio; si tienen nociones sobre la dispersión; si muestran competencia para hacer el gráfico.

Solución:

- a) A simple vista es evidente notar que Chile no está incluido en este promedio, ya que el % máximo de superficie protegida es de 10,7 % y dado que el promedio es de 5,4%, es imposible que el 18,3% de superficie protegida para Chile esté incluido en el promedio.

Otra forma de fundamentar esta respuesta sería generar una lista con los porcentajes de los 17 países de la lista (incluido Chile) y calcular la media. Iniciamos dicha lista con Brasil para concluir con Chile de la forma siguiente.

SUB	List 8	List 9	List 10	List 11
1	3.3			
2	0.2			
3	5.1			
4	3.4			

1VAR 2VAR REG SET

SUB	List 8	List 9	List 10	List 11
6	10.7			
7	6.9			
8	5			
9	0.6			

0.6000000000
1VAR 2VAR REG SET

SUB	List 8	List 9	List 10	List 11
10	9.8			
11	10.5			
12	10.6			
13	2.6			

2.6000000000
1VAR 2VAR REG SET

SUB	List 8	List 9	List 10	List 11
14	2.5			
15	5			
16	6.6			
17	18.3			

18.3000000000
1VAR 2VAR REG SET

```

1 variable
x̄ = 6.18235294
Σx = 105.1
Σx² = 987.07
x̄n = 4.45437462
x̄n-1 = 4.59146426
n = 17
    
```

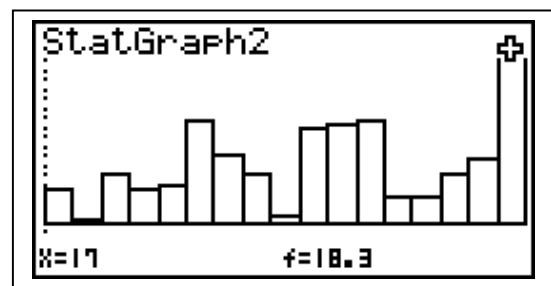
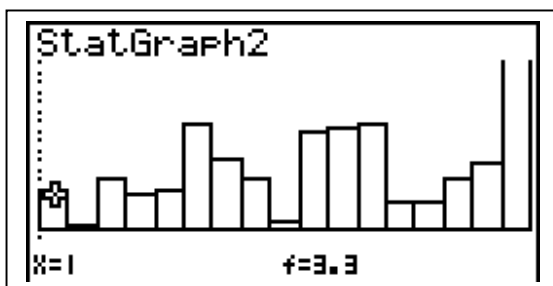
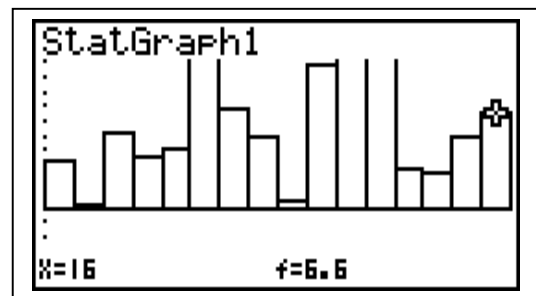
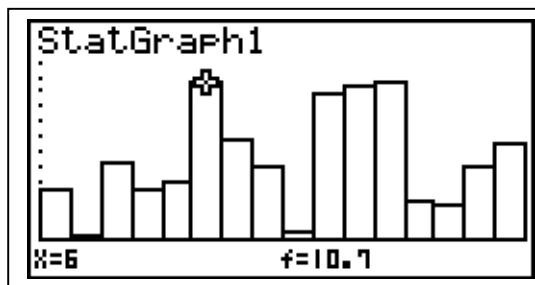
La media obtenida es de 6,18%, la que difiere del 5,4% mencionado.

```

1 variable
x̄ = 5.425
Σx = 86.8
Σx² = 652.18
x̄n = 3.36609937
x̄n-1 = 3.47649248
n = 16
    
```

Si calculamos la media con tan solo los primeros 16 países excluyendo a Chile obtenemos el valor de 5,4 %.

- b) El valor de 5,4 % promedio de superficie protegida de cada país se interpreta como si de cada 100 km² de superficie de un país, 5,4 km² están protegidos. En el caso de Chile, cuya superficie insular es de aproximadamente 756.600 km², el 18,3 % de dicha superficie está protegida, es decir, 138.458 km² de superficie total protegida. En nuestro caso contamos con 18,3 km² protegidos por cada 100 km², o sea, 12,9 km² más que la media.
- c) Si consideramos tan solo los 16 primeros países (excluyendo a Chile) a simple vista hay 4 países (N Zeland, EEUU, Australia y Francia) que duplican o casi duplican la media. Por otro lado hay 4países (Uruguay, Irlanda, Sigapur y Finlandia) que aparecen por debajo de la mitad de la media. Los 8 países restantes están más próximos a la media, lo que hace en general una visión dispersa de los datos. Si agregamos a Chile, más disperso estarían los datos.
- d) Esto último se puede constatar con el gráfico de barras proporcionado por la calculadora. Los dos primeros gráficos (StatGraph1) muestran los datos de todos los países exceptuando a Chile; el último dato corresponde al 16-avo país (Suecia) con un 6,6% de superficie protegida. Los dos segundos gráficos (StatGraph2) muestran los datos del total de 17 países, siendo Chile el último país.



Actividad 2

Establecen diferencias entre estadística descriptiva e inferencial.

Ejemplo A

Clasificar los estudios siguientes en estadística descriptiva o inferencia estadística:

- a) Pedro predice que el candidato Belaire va a ganar la elección presidencial con un 53% de los votos a partir de los resultados de 45 comunas.
- b) El ecologista Dr. Valverde dice que la carne de los peces del lago Rapel contiene un promedio de 400 unidades de mercurio.
- c) En el Colegio Fuente Nueva, el promedio de la PAA Verbal fue de 550.
- d) Se prevé 25 accidentes de tránsito el próximo fin de semana largo.
- e) El año pasado 72% de los trabajadores de la fábrica de zapatos Tacones perdieron, al menos, un día de trabajo.

Solución:

La estadística descriptiva describe información, analiza e interpreta datos. Realiza el estudio sobre la población completa, observando una característica de la misma y calculando medidas de tendencia central.

La estadística inferencial infiere, predice, induce sobre una muestra de población. Realiza el estudio descriptivo sobre un subconjunto de la población llamado muestra y, posteriormente, extiende los resultados obtenidos a toda la población.

- a) Inferencial, ya que predice un resultado a partir de los resultados obtenidos por una muestra de 45 comunas de un total desconocido de comunas (universo poblacional)
- b) Inferencial, ya que predice a partir de una muestra de peces del lago.
- c) Descriptiva, dado que se conoce el universo total de alumnos que rindió la PAA y entrega una medida de tendencia central.
- d) Inferencial, ya que predice a partir de estadísticas anteriores.
- e) Descriptiva, ya que ese porcentaje es obtenido a partir de un universo conocido de trabajadores de la fábrica de zapatos.

Ejemplo B

Indicar por qué las muestras que se proponen a continuación no son adecuadas:

- a) Para tener una información sobre si una obra de teatro fue o no del agrado del público se encuestó a diez personas familiares de los actores.
- b) Para tener información sobre las preferencias de los electores para la próxima elección presidencial se hizo una encuesta a 50 personas que trabajan en la minería del cobre.
- c) Para tener información sobre la lectura de una revista entre los jóvenes se encuestó a 30 estudiantes de la carrera de diseño de una universidad.

Observar si hacen referencia a la representatividad de las muestras en relación con la población que debiera considerarse.

Solución:

- a) No hay información si las personas encuestadas asistieron a dicha obra de teatro. Para obtener información fidedigna se debe encuestar directamente a quienes asistieron a la obra.
- b) No es una muestra representativa de la población del país, tan solo refleja una muestra de la población de un sector productivo del país.
- c) La muestra no representa a la población total de jóvenes que leen revistas de un país, tan solo se enfoca en una determinada carrera de una universidad específica. No considera otras carreras ni otras universidades.

Actividad 3

Interpretan los márgenes de error y los grados de confianza señalados en las investigaciones y encuestas de opinión.

Ejemplo A

Considerar encuestas que se publican en la prensa, constatar si indican o no los márgenes de error y el nivel de confianza. De acuerdo a esa información opinar sobre los resultados.

Ejemplo B

Conocer investigaciones hechas por instituciones responsables y analizar los resultados considerando los índices de error y los grados de confianza.