

# V CEAM CM Cuenca



## Estadística y Probabilidad Interactivas con Geogebra

Virgilio Gómez Rubio  
Mª José Haro Delicado

**Marzo 2023**

# Descarga de materiales



<https://github.com/becarioprecario/TallerGeoGebra>



# **ESTADÍSTICA DESCRIPTIVA**

# Estadística descriptiva

1. El puntaje de Apgar se usa para evaluar reflejos y respuestas de recién nacidos. A cada bebé un profesional de la medicina le asigna un puntaje y los valores posibles son enteros entre cero y diez. Se toma una muestra de 1000 bebés nacidos en cierto condado y los resultados han sido los siguientes:

0	1	2	3	4	5	6	7	8	9	10
1	3	2	4	25	35	198	367	216	131	18

- Halla la media de los puntajes de Apgar. Halla la desviación típica de la muestra.
- Halla la mediana muestral.
- ¿Cuáles son los cuartiles primero y tercero?

# Resolviendo con Geogebra

- Abrimos la hoja de cálculo
- Introducimos en la 1<sup>a</sup> columna los valores de la variable y en la segunda las frecuencias absolutas
- Creamos sendas listas.

The screenshot shows the GeoGebra interface with two main panes: the Algebra View on the left and the Spreadsheet View on the right.

**Algebra View:** This pane contains a toolbar with various geometric tools at the top. Below the toolbar, there is a section labeled "Vista Algebraica" with a close button. Underneath this, there is a tree view labeled "Lista" which contains two items:

- lista1 = {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
- lista2 = {1, 3, 2, 4, 25, 35, 198, 367, 216, 131, 18}

**Spreadsheet View:** This pane has a header with buttons for  $f(x)$ , **N**, **I**, and other document-related functions. The main area is a table with three columns labeled A, B, and C. The data is as follows:

	A	B	C
1	0	1	
2	1	3	
3	2	2	
4	3	4	
5	4	25	
6	5	35	
7	6	198	
8	7	367	
9	8	216	
10	9	131	
11	10	18	
12			
13			

# Cálculo de las medidas de centralización y dispersión

**Media**[ <Lista de Números>, <Lista de Frecuencias> ] → media =7.14

**Mediana**[ <Lista de Números>, <Lista de Frecuencias> ] → mediana =7

La **moda** sólo se puede calcular para valores sin agrupar Moda[>Lista de números>]

**Varianza poblacional:** Varianza[ <Lista de Números>, <Lista de Frecuencias> ] → 1.71696

**Varianza muestral:** VarianzaMuestral[ <Lista de Números>, <Lista de Frecuencias> ] →1.71867

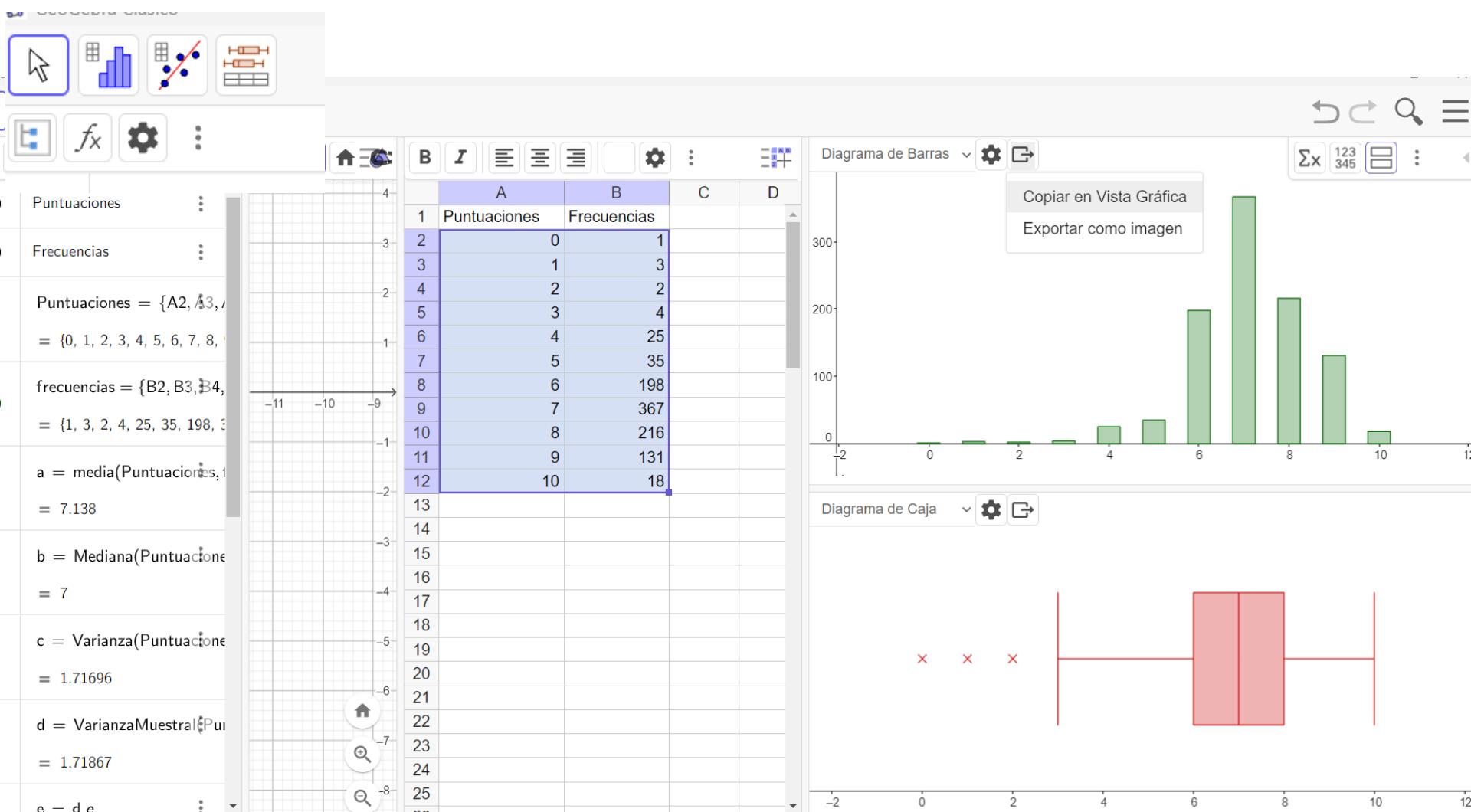
**Desviación típica poblacional:** DE[ <Lista de Números>, <Lista de Frecuencias> ] →1.310328

**Desviación típica muestral:** stdev[ <Lista de Números>, <Lista de Frecuencias> ] →1.31098

**Cuartil1**[lista1, lista2] →6

**Cuartil3**[lista1, lista2] →8

# Análisis de una variable



En la versión Geogebra Classic 5, si hay frecuencias absolutas, éstas, deben formar parte de una lista y pinchando sobre la rueda de la esquina superior derecha, podemos introducirlas. Para hacerlo, seleccionamos la columna que las contenga y pinchamos sobre la mano que aparece encima de las celdas.

The screenshot shows the 'Fuente de datos' (Data Source) dialog box in GeoGebra. The dialog has a title bar 'Fuente de datos' and a sub-section 'Análisis de una variable'. It contains two tables:

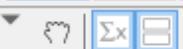
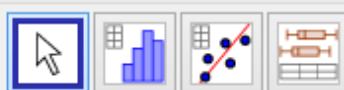
A1:A11		B1:B11	
0		1	
1		3	
2		2	
3		4	
4		25	
5		35	
6		198	
7		367	

A	B
0	1
1	3
2	2
3	4
4	25
5	35
6	198
7	367
8	216
9	131
10	18

At the bottom of the dialog are two buttons: 'Cancela' (Cancel) and 'Analiza' (Analyze). The 'Analiza' button is highlighted with a blue border.

## Análisis de datos

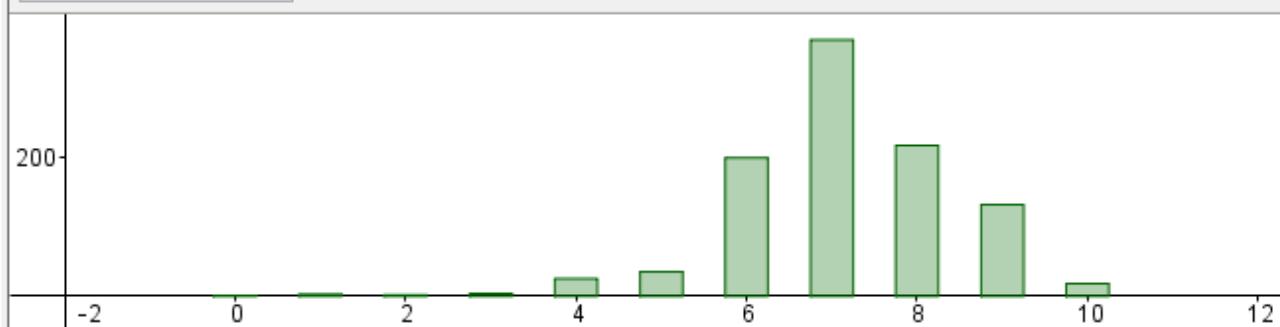
x



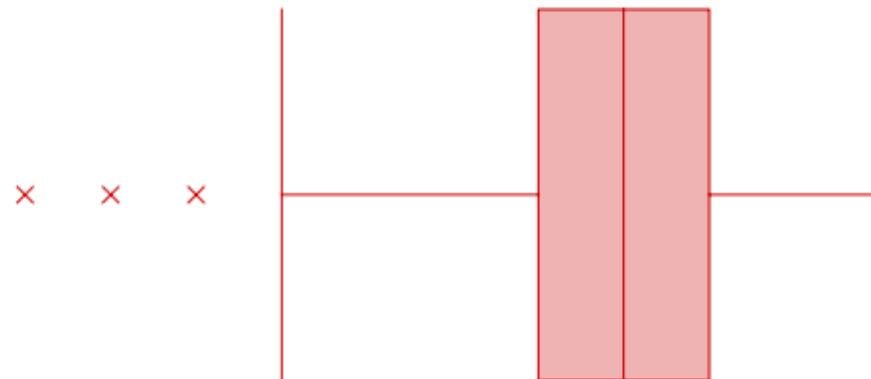
## Estadísticas

n	1000
Media	7.138
$\sigma$	1.3103
s	1.311
$\Sigma x$	7138
$\Sigma x^2$	52668
Mín	0
Q1	6
Mediana	7
Q3	8
Máx	10

## Diagrama de Barras



## Diagrama de Caja



# Algunas representaciones gráficas

2. En un estudio sobre la amnesia postraumática tras una lesión craneal, se estudió el tiempo en días que estuvieron los pacientes en coma. Se recogieron los datos siguientes:

2	8	9	14	16	6	10	8	7
13	12	11	11	11	13	15	10	11
15	12	20						

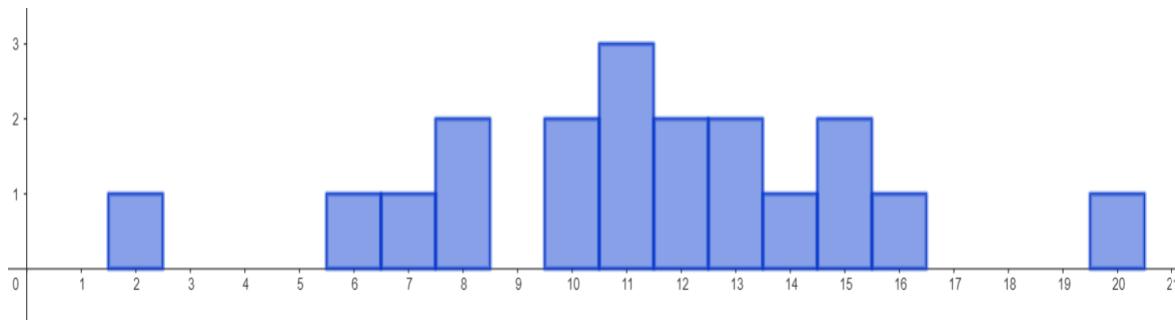
- Construir un diagrama de barras
- Construir un diagrama de tallo y hojas para estos datos. ¿Parecen estar los datos simétricamente distribuidos?
- Construir un diagrama de cajas y bigotes para los datos. ¿Da la misma impresión de simetría que con el diagrama de tallo y hojas?
- ¿Existen datos puntuales que puedan considerarse como atípicos?

# DIAGRAMA DE BARRAS

Barras[ <Lista de Datos en Bruto>, <Ancho de Barras> ]

Si el ancho de barras es 0, se obtiene una especie de agujas a diferentes alturas.

También se puede trabajar con frecuencias y con otras opciones.



En este caso, como los valores de la variable van de 1 en 1, puede ser conveniente que el ancho de las barras sea de longitud 1

# DIAGRAMA DE TALLOS Y HOJAS

DiagramaTalloHojas[ <Lista> ]

0	2	6	7	8	8	9						
1	0	0	1	1	1	1	2	2	3	3	4	5
2	0										5	6

Clave : 3|1 implica 31

Para obtener la parte entera de los diferentes valores hay que multiplicar por 10

# DIAGRAMA DE TALLOS Y HOJAS

DiagramaTalloHojas[ <Lista>, <Ajuste -1|0|1> ]

2	0
3	
4	
5	
6	0
7	0
8	0 0
9	0
10	0 0
11	0 0 0 0
12	0 0
13	0 0
14	0
15	0 0
16	0
17	
18	
19	
20	0

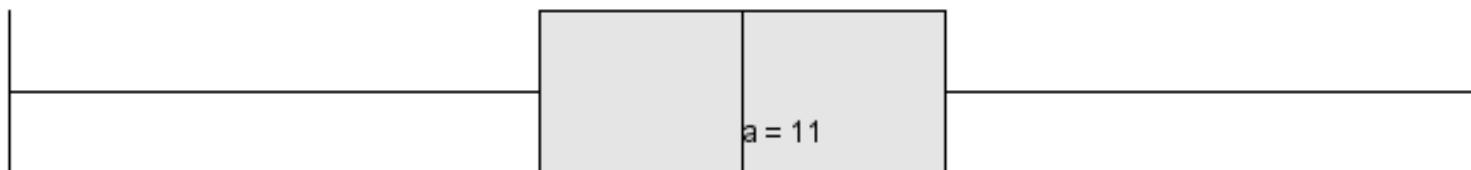
Clave : 3|1 implica 3.1

En este caso el valor del ajuste es -1. Significa que la parte entera del número queda a la izquierda de la barra vertical (el tallo).

Los valores a cuyo lado no aparece el cero no corresponden a valores de la variable.

# DIAGRAMA DE CAJAS Y BIGOTES

DiagramaCaja[ <Offset\_y>, <Escala\_y>, <Lista de Datos en Bruto> ]



El problema de calcular el diagrama de cajas y bigotes utilizando la instrucción anterior es que los bigotes se extienden hasta los valores máximo y mínimo, con lo cual, no sirve para detectar los valores atípicos.

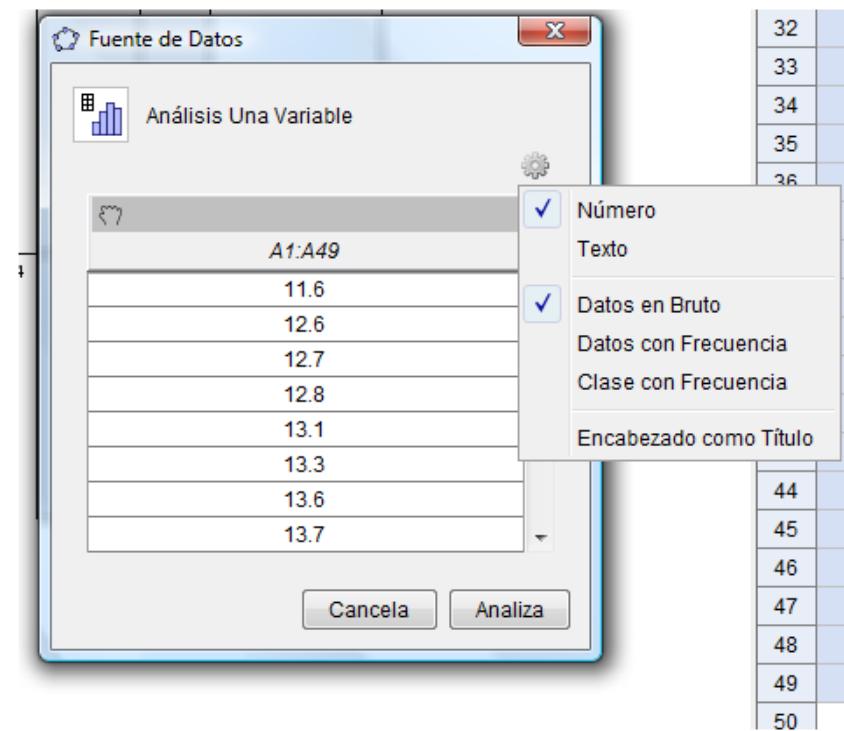
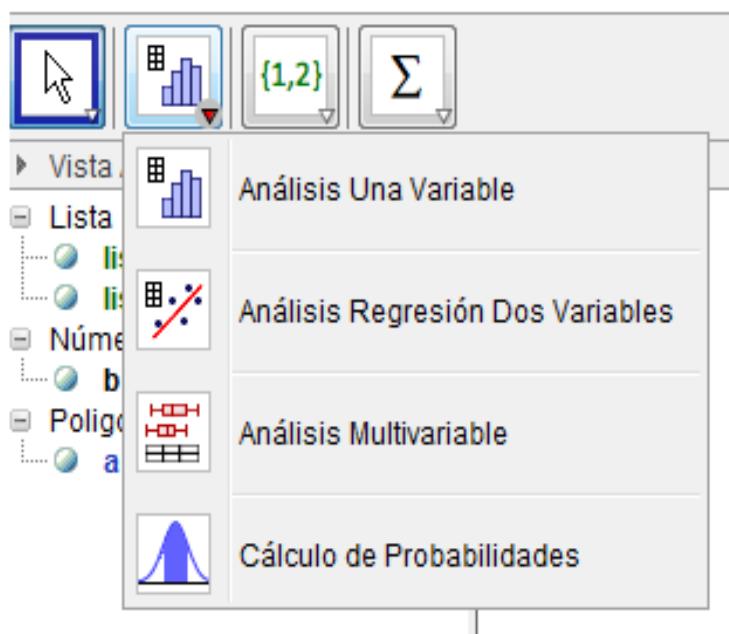
*3. Parte de un estudio de control de calidad tuvo como objetivo mejorar una línea de producción. Se midieron los pesos (en onzas) de 50 barras de jabón. Los resultados son los siguientes, ordenados de menor a mayor.*

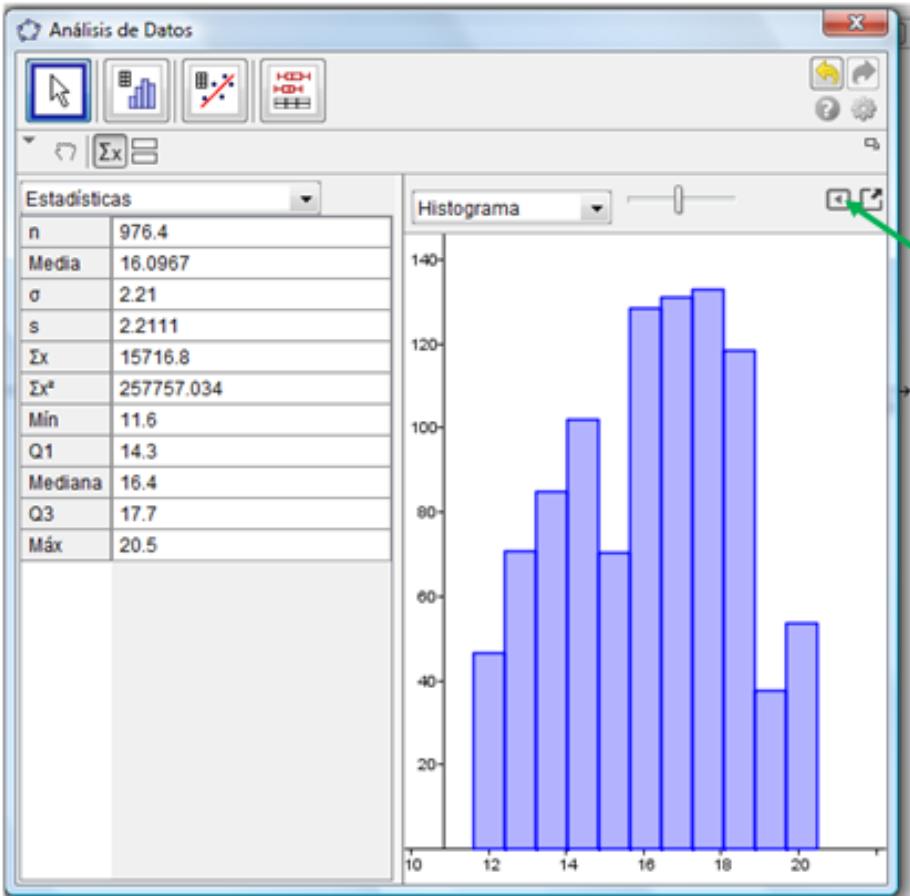
11.6	12.6	12.7	12.8	13.1	13.3	13.6	13.7	13.8	14.1
14.3	14.3	14.6	14.8	15.1	15.2	15.6	15.6	15.7	15.8
15.8	15.9	15.9	16.1	16.2	16.2	16.3	16.4	16.5	16.5
16.5	16.6	17.0	17.1	17.3	17.3	17.4	17.4	17.4	17.6
17.7	18.1	18.3	18.3	18.3	18.5	18.5	18.8	19.2	20.3

- a) *Construye un diagrama de tallos y hojas para estos datos.*
- b) *Construye un histograma para estos datos.*
- c) *Construye un diagrama de cajas para estos datos. ¿Identifica datos atípicos?*

# Usaremos Análisis de una variable

Seleccionamos nuestros datos en la hoja de cálculo y hacemos clic sobre “Análisis de una variable” y después en “analiza”

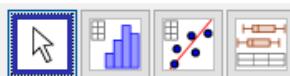




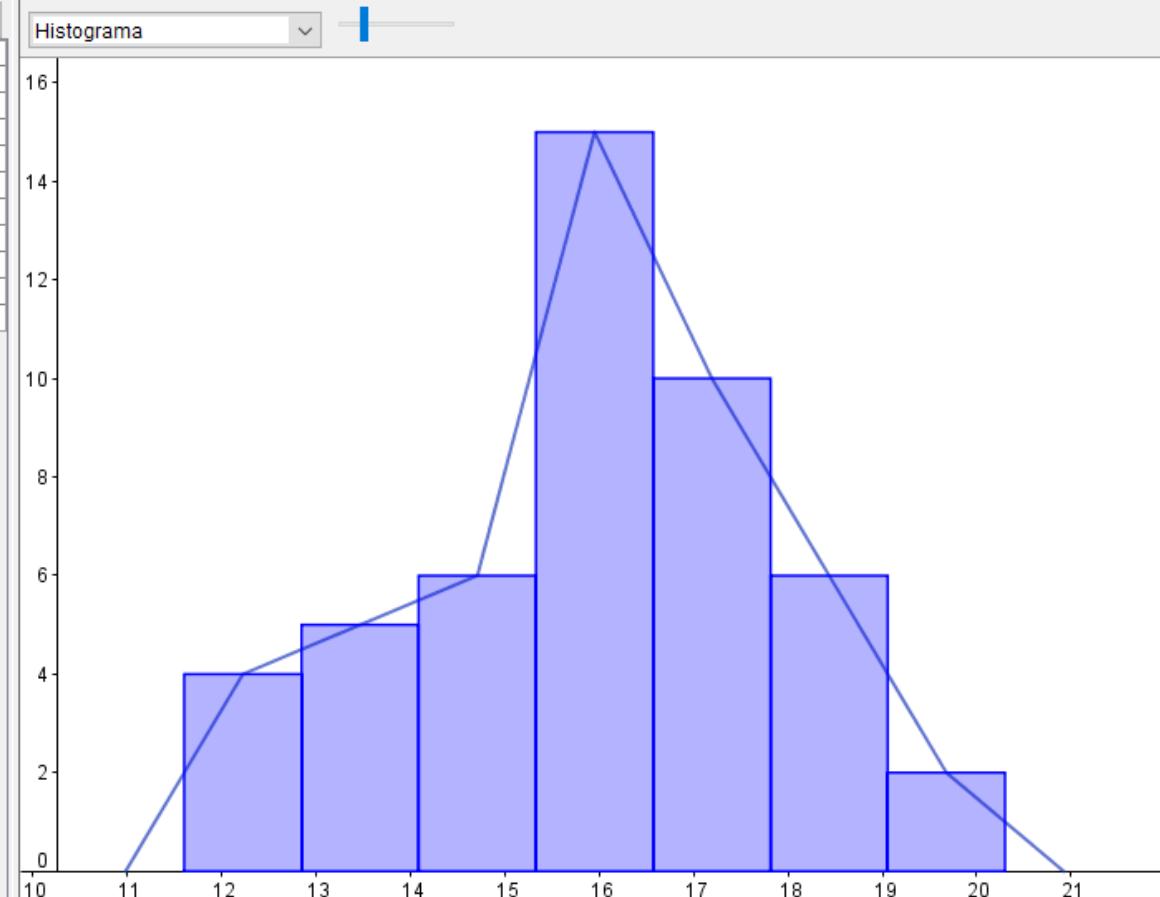
**Si pinchamos sobre el botón que indica la flecha anterior, podemos agregar la tabla de frecuencias y el polígono de frecuencias entre otras cosas. La tabla de la izquierda en la que aparecen las medidas de centralización, posición y dispersión, se obtiene pinchando sobre el botón en el que aparece el signo sumatorio.**

## Análisis de datos - Ejercicio 3.ggb

X



Estadísticas	
n	48
Media	16.0229
$\sigma$	1.9252
s	1.9456
$\Sigma x$	769.1
$\Sigma x^2$	12501.13
Mín	11.6
Q1	14.7
Mediana	16.2
Q3	17.4
Máx	20.3



Histograma  Gráfica

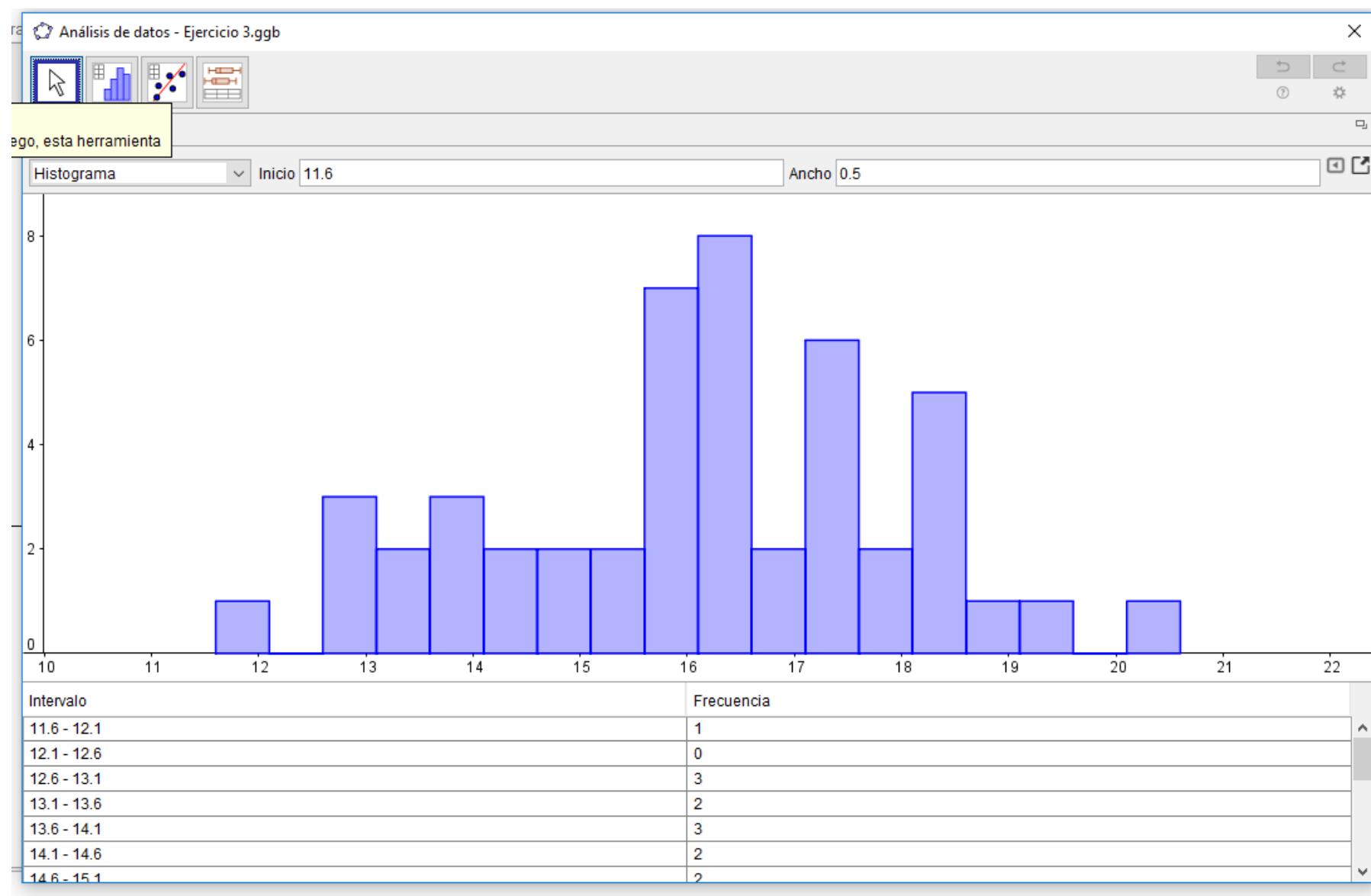
Clases  Establecer las clases manualmente

Reglas de clase:   $\leq x <$    $< x \leq$

Tipo de frecuencia  Acumulada  Cuenta  Relativo  Normalizado

Muestra  Histograma  Tabla de frecuencias  Polígono de frecuencias  Curva Normal

# Creación manual de los intervalos

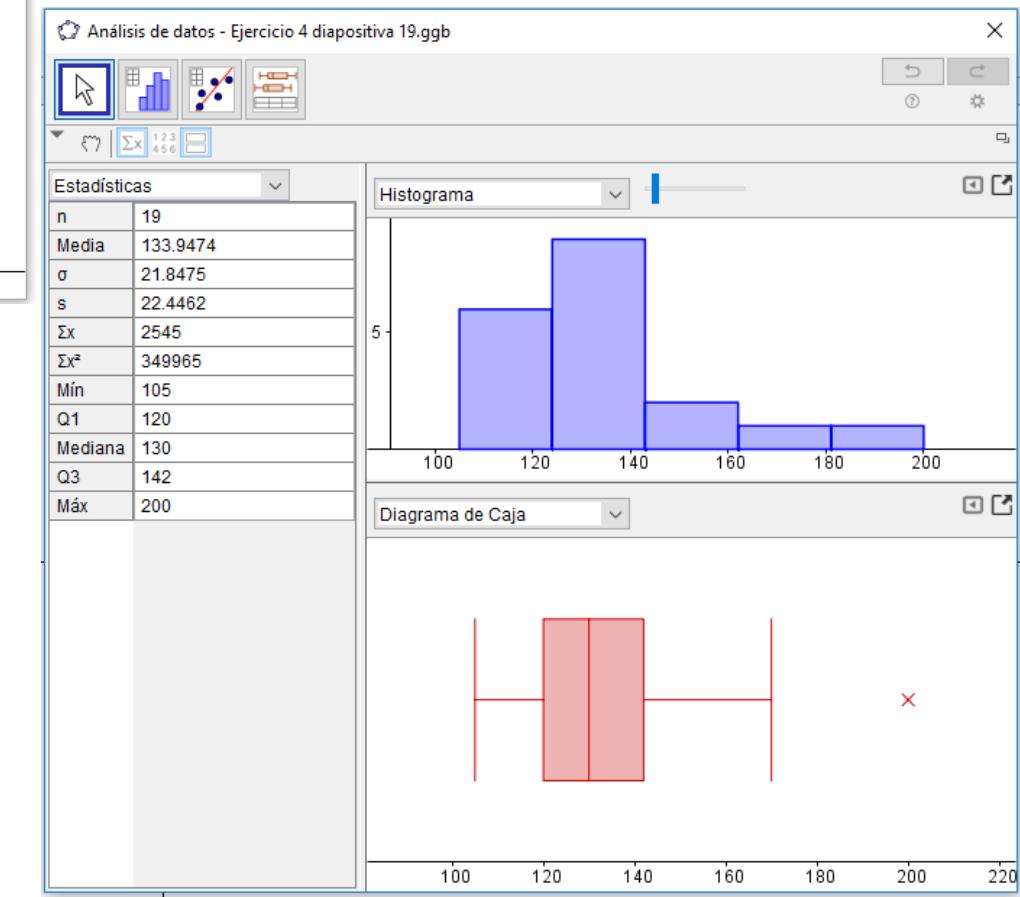
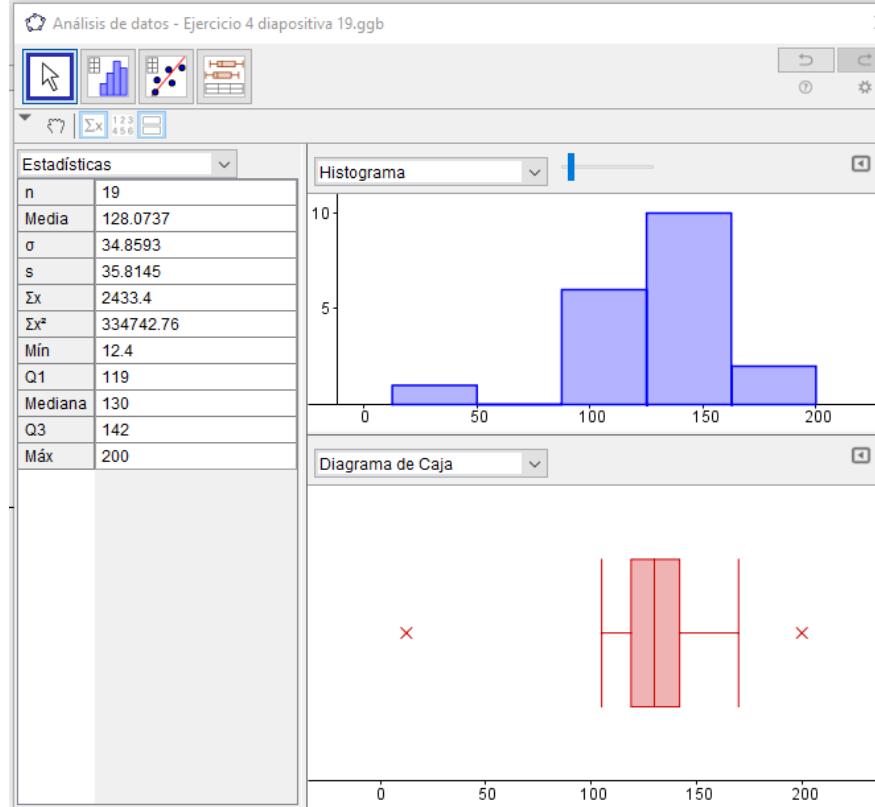


# A PRACTICAR

4. Se está estudiando un medicamento, amantadina (A), para combatir el virus de la gripe. Se ha administrado por vía oral una dosis única de 100 mg a adultos sanos. La variable estudiada es  $T_{max}$ , tiempo requerido en minutos para alcanzar la concentración máxima de plasma. Se obtuvieron los datos siguientes: (Basado en la información publicada en Gordon Douglas Jr. "Drug Therapy", New England Journal of Medicine, 1990).

<b><math>T_{max}</math> (A)</b>			
105	123	12.4	126
108	134	120	112
130	119	132	130
133	136	142	145
156	170	200	

- Construye un diagrama de cajas para el conjunto de datos e identifica los datos atípicos
- Calcula  $\bar{x}$  y  $s^2$  para los datos del conjunto A.
- Supongamos que el dato atípico del conjunto A es el resultado de un punto decimal mal colocado. Corregir el error borrando el decimal y observar qué cambios produce esto en el diagrama de cajas. Volver a calcular  $\bar{x}$  y  $s^2$ , utilizando los datos correctos y comparar los resultados con los del apartado b).



# **ESTADÍSTICA DESCRIPTIVA PARA DOS VARIABLES**

1. Se llevó a cabo un estudio acerca de la cantidad de azúcar refinada mediante un cierto proceso a varias temperaturas diferentes. Los datos se codificaron y registraron en el siguiente cuadro.

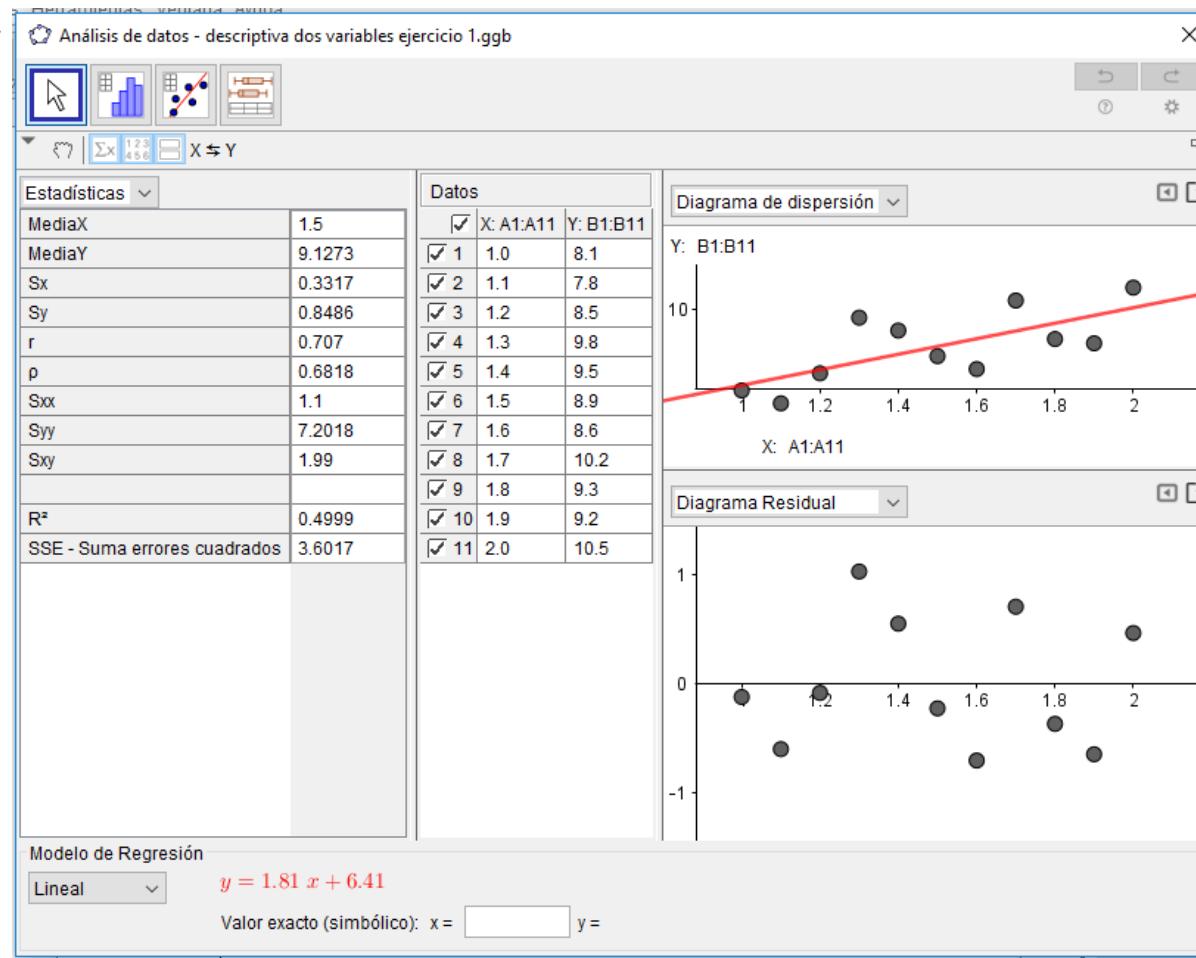
<i>Temperatura, x</i>	<i>Azúcar transformada, y</i>
1.0	8.1
1.1	7.8
1.2	8.5
1.3	9.8
1.4	9.5
1.5	8.9
1.6	8.6
1.7	10.2
1.8	9.3
1.9	9.2
2.0	10.5

¿Es correcto relacionar los datos mediante una recta de regresión?

Si lo es, calcúlala y obtén la cantidad promedio de azúcar refinada que se produce cuando la temperatura codificada es de 1.75

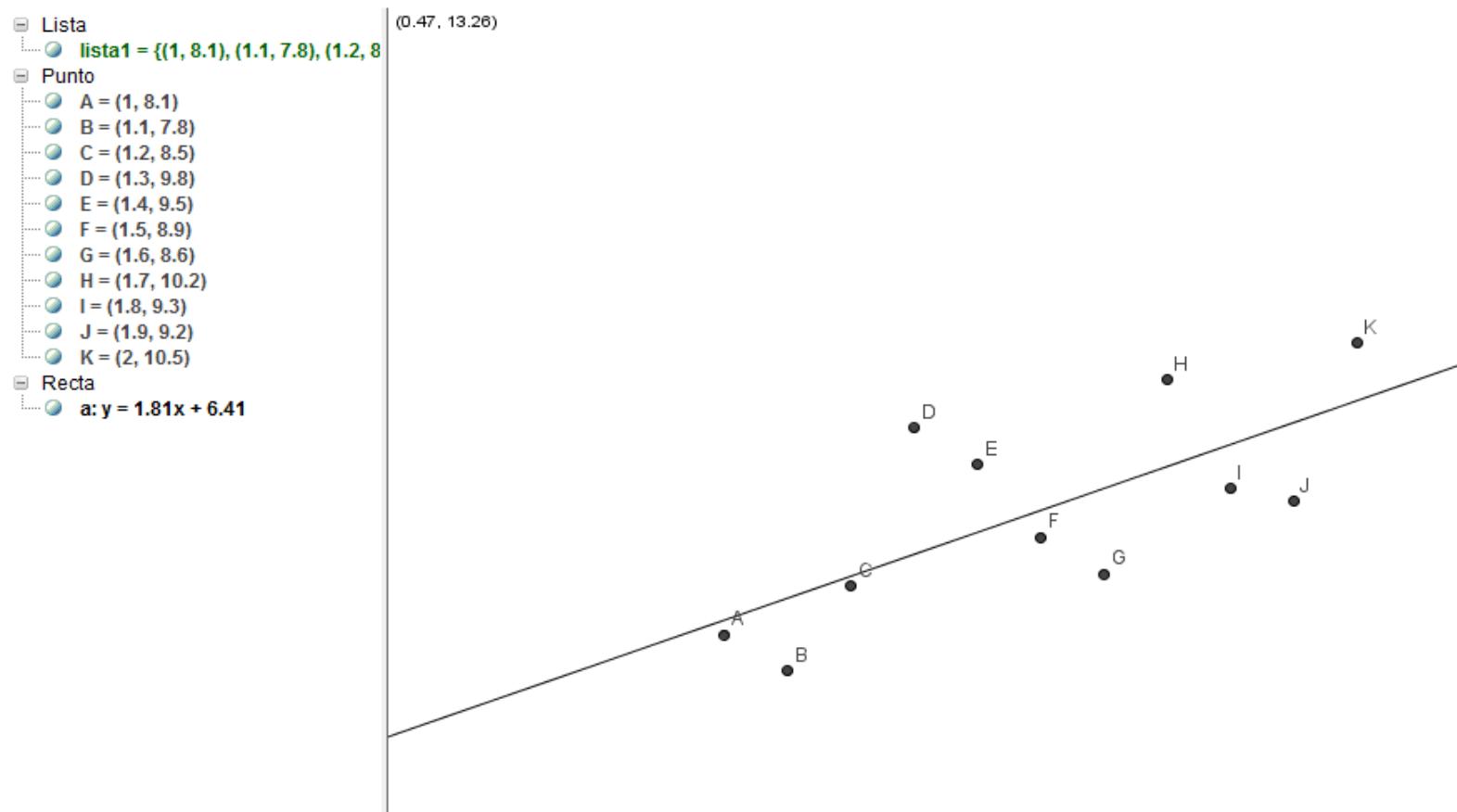
Introducimos los datos en dos columnas y usamos la opción **Análisis regresión de dos variables**. De esta forma, obtenemos tanto el diagrama de dispersión, como el diagrama de residuos y un resumen estadístico de la relación entre las dos variables. Se realizan diversos tipos de ajustes y se pueden usar para predecir valores de la variable dependiente.

Se pueden intercambiar las variables, pasando la variable dependiente a ser independiente.

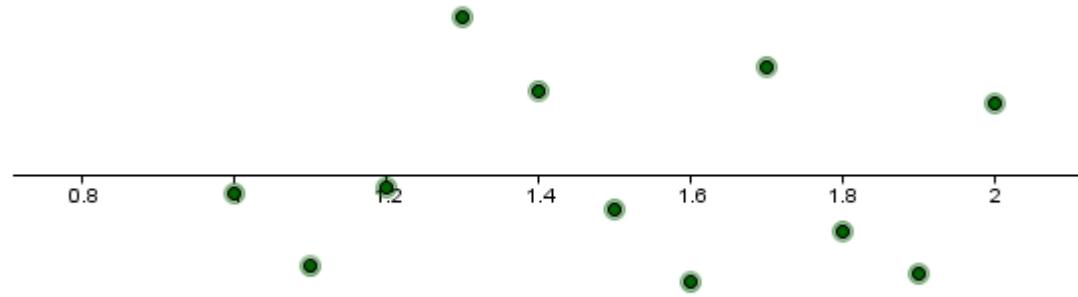


También se puede realizar usando comandos seleccionando previamente las dos columnas de datos y eligiendo la opción “crea lista de puntos”

## AjusteLineal[ <Lista de Puntos> ]



El diagrama de residuos lo podemos obtener con la opción  
**DiagramaResidual[ <Lista de Puntos>, <Función> ]**  
introduciendo en la opción Función, la ecuación de la recta obtenida previamente.



Para obtener el coeficiente de correlación, usariamos  
**Coeficientecorrelación[<Lista de Puntos>]**

También se puede obtener el coeficiente de correlación de Spearman → **Spearman[ <Lista de Puntos> ]**

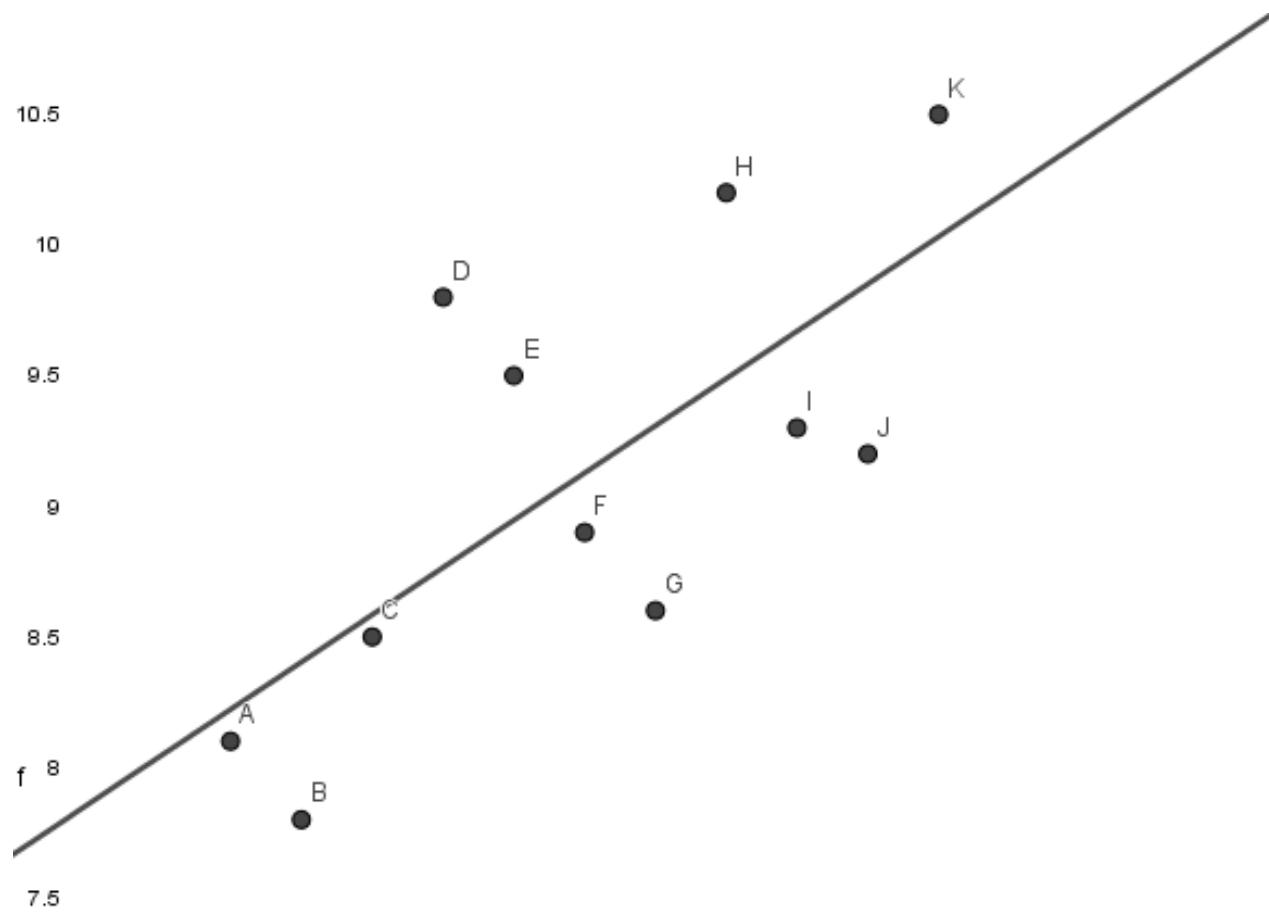
También hay opciones para calcular las medias, varianzas y desviaciones típicas marginales, así como para calcular la covarianza.

# Diagnóstico de la regresión lineal

- Linearidad
- Normalidad de los residuos
- Variabilidad constante

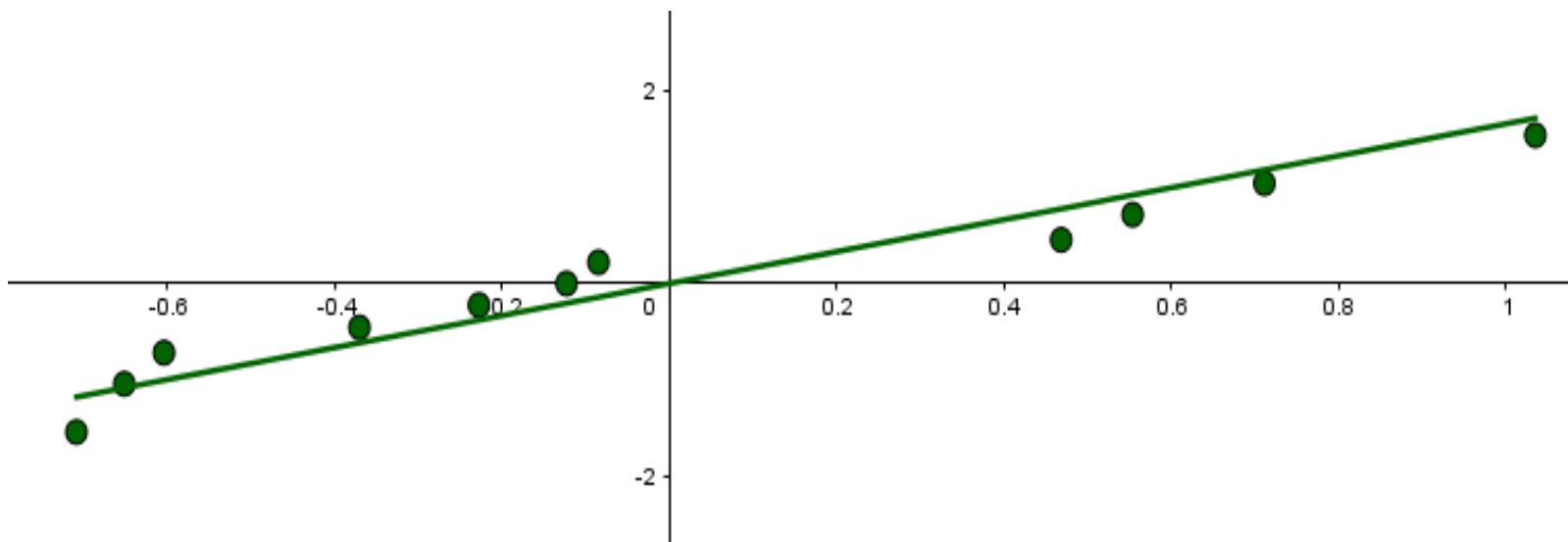
# Linearidad

## Diagrama de dispersión



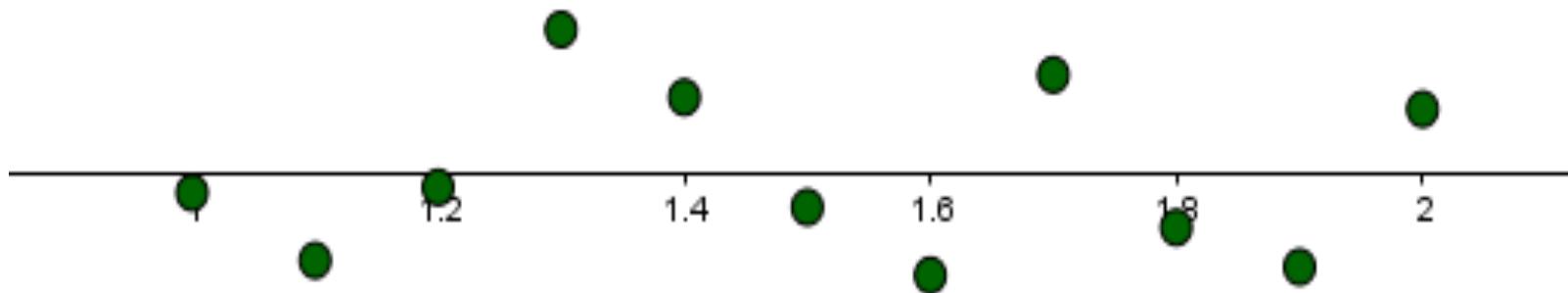
# Normalidad de los residuos

## Gráfico QQ[lista de residuos]



# Variabilidad constante de los residuos

DiagramaResidual[ <Lista de Puntos>, <Función> ]



# Contraste de la regresión lineal simple

Es la inferencia estadística la que debe validar la recta de regresión obtenida. Lo hace mediante un test de hipótesis basado en un análisis de la varianza.

*Contraste de  $H_0: \rho = 0$  frente a  $H_1: \rho \neq 0$*

$$\text{Sea } t = r \sqrt{\frac{n-2}{1-r^2}} \quad \begin{array}{l} \text{Se acepta } H_0 \text{ si } |t| < t_{n-2; \alpha/2} \\ \text{Se rechaza } H_0 \text{ si } |t| \geq t_{n-2; \alpha/2} \end{array}$$

DistribuciónTInversa( <Grados de libertad>, <Probabilidad>)

DistribuciónTInversa(n-2, 1-α/2)

# VARIABLES Y MODELOS DE DISTRIBUCIÓN

Se exige que el espesor de las hojas de aluminio para hacer latas esté entre 1 y 1.5 milésimas de pulgada. Sea  $X$  el espesor de una hoja de aluminio. Sea la función de densidad de probabilidad de  $X$  la siguiente:

$$f(x) = \begin{cases} c \cdot x & 0 < x < 2 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

- a. ¿Qué proporción de hojas cumplirá las exigencias?
- b. Halla la media del espesor de una hoja.
- c. Halla la varianza del espesor de una hoja.
- d. Halla la función de distribución del espesor.
- e. Una hoja especial tiene 0.8 milésimas de pulgada de espesor ¿Qué proporción de hojas tiene más espesor?

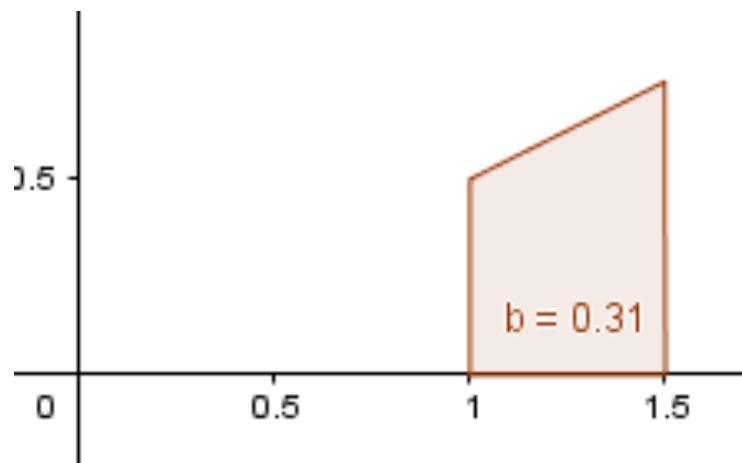
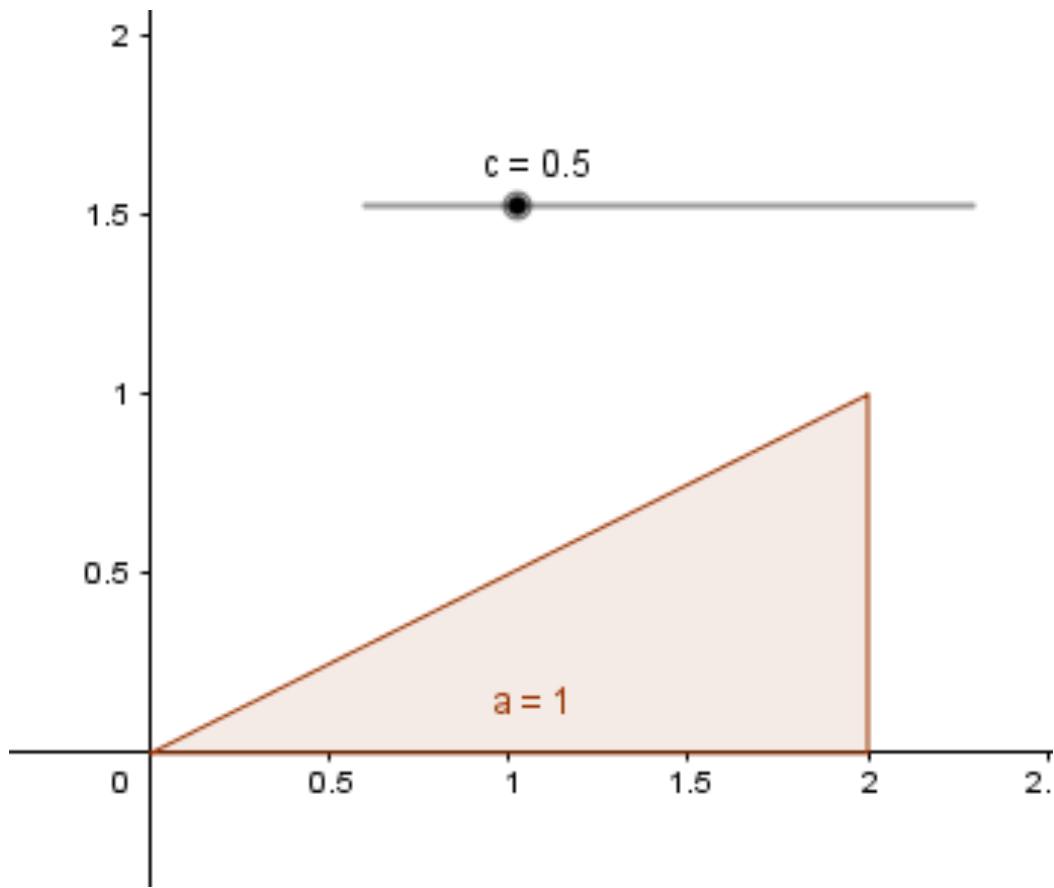
a) Creamos un deslizador para  $c$  y representamos gráficamente la función  $f(x)=c \cdot x$ . El deslizador se puede iniciar en 0, ya que una de las condiciones que debe cumplir una función de densidad es el ser positiva,  $f(x)>0$  y, en este caso,  $x \in [0,2]$

Se calcula el área en función del valor de  $c$  con el comando

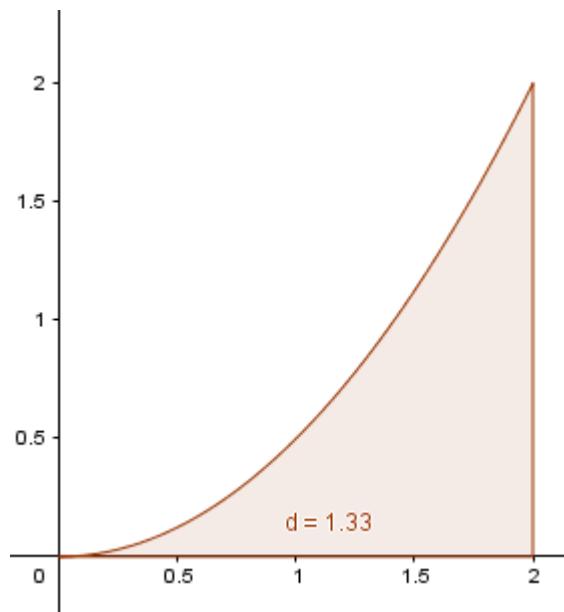
**Integral[ <Función>, <Valor Inicial de x>, <Valor Final de x> ]**

Después de obtener el valor de  $c$ , en este caso de 0.5, calculamos

$$\int_1^{1.5} cx dx$$

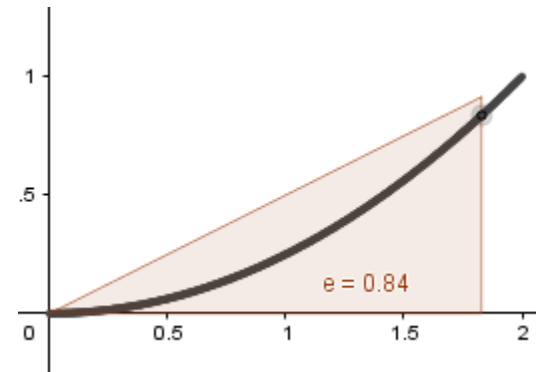


b) Para calcular la media usamos  
**Integral[x\*0.5\*x,0,2]**. Obtenemos

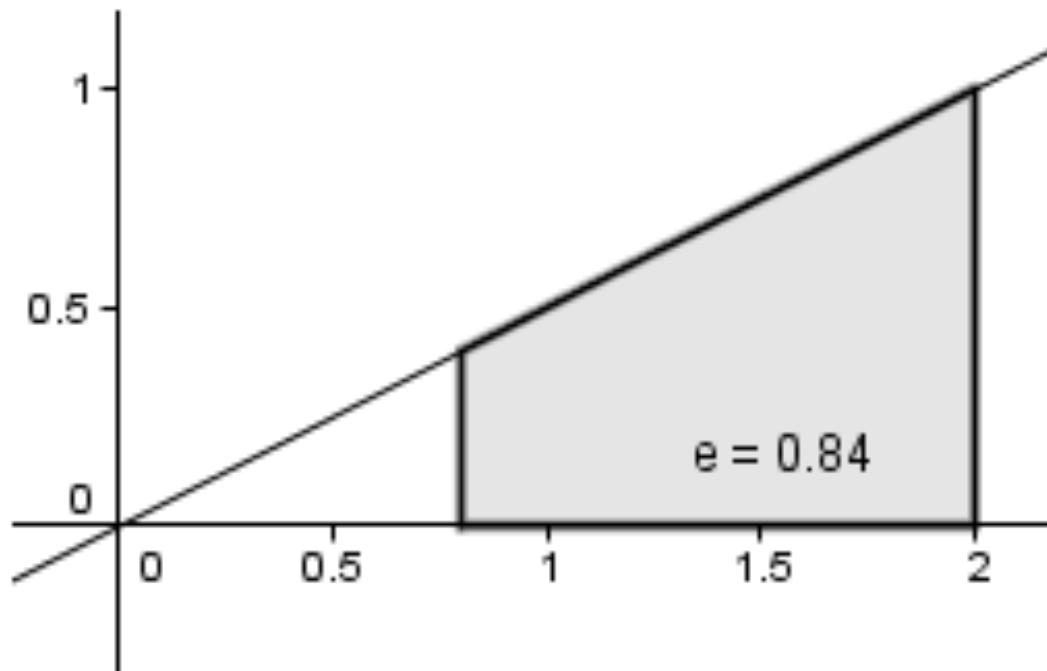


c) Para hallar la varianza, podemos usar  
**Integral[x^2 0.5 x, 0, 2] – (Integral[x 0.5 x, 0, 2])^2**  
d) Para hallar la función de distribución, podemos crear un nuevo deslizador, k, que tome valores en el intervalo  $[0,2]$ . A continuación, obtenemos el valor distribución=Integral[0.5 x, 0, k].

Para que se visualice la función de distribución creamos el punto A(k, distribución) y activamos el rastro, poniendo el deslizador k en animación automática.



e) Para responder a esta pregunta basta con calcular  
 $\text{Integral}[0.5*x, 0.8, 2]$



2. Las puntuaciones de una prueba estandarizada se distribuyen normalmente con media de 480 y desviación típica de 90.

¿Cuál es la proporción de puntuaciones mayores a 700?

¿Cuál es el 25º percentil de las puntuaciones?

Si la puntuación de alguien es de 600 ¿En qué percentil se encuentra?

¿Qué proporción de las puntuaciones se encuentra entre 420 y 520?

- a)  $P(\xi > 700) = 1 - \text{Normal}[\text{<media>}, \text{<desviación típica>}]$ ,  $\text{<valor>} = 1 - \text{Normal}[480, 90, 700] = 0.0073$
- b) **NormalInversa[ <Media>, <Desviación Estándar>, <Probabilidad> ]**  
 $\text{NormalInversa}[480, 90, 0.25] = 419.296.$
- c) **Normal[ <Media>, <Desviación Estándar>, x ]**  
 $\text{Normal}[480, 90, 600] = 0.909$ , lo que implica prácticamente un percentil 91
- d)  $\text{Normal}[480, 90, 520] - \text{Normal}[480, 90, 420] = 0.419$

# **APROXIMACIÓN DE LA BINOMIAL MEDIANTE LA NORMAL**

1. El 45% de los condensadores de una cierta partida presenta deficiencias

a) ¿Cuál es la probabilidad de que presenten deficiencias 1, 2, 3, 4,...,10 de los condensadores?

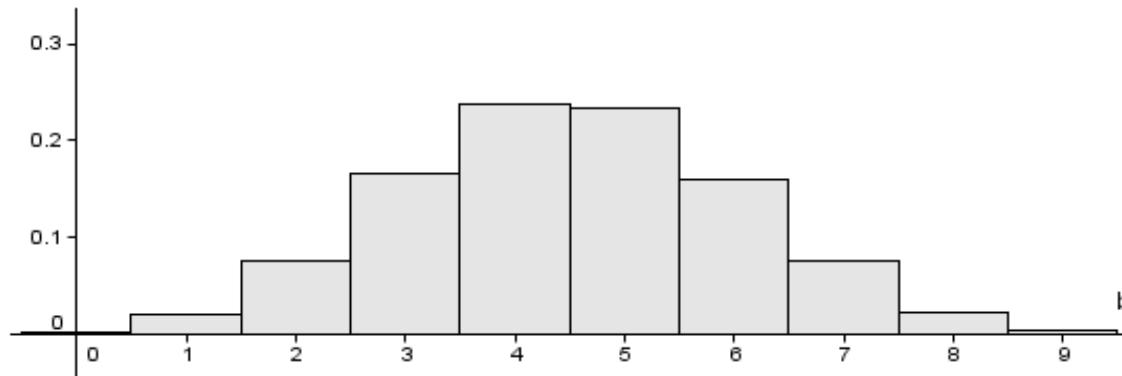
b) ¿Cuál es la probabilidad de que de diez condensadores presenten deficiencias un número menor o igual a la mitad.

c) ¿Puede la distribución aproximarse a la distribución normal?

Calcula la probabilidad de que de diez condensadores examinados, la mitad presente dichas deficiencias utilizando la distribución normal.

**a) Distribución Binomial[ <Número de Ensayos>, <Probabilidad de Éxito>, <Acumulada Booleana> ]**

DistribuciónBinomial[10, 0.45, false]



b) DistribuciónBinomial[ <Número de Ensayos>, <Probabilidad de Éxito>, <Valor de Variable>, <Acumulada Booleana> ]

DistribuciónBinomial[ 10,0.45,5,true]=0.738

c)  $nq=5.5>5$ ,  $\mu=np=0.45\cdot 10=4.5$  no es mayor o igual que 5, pero no queda muy lejos, por lo tanto, aunque muy en el límite, lo podríamos admitir.

$$\sigma^2=npq=10\cdot 0.45\cdot 0.55=2.475; \quad \sigma=1.57$$

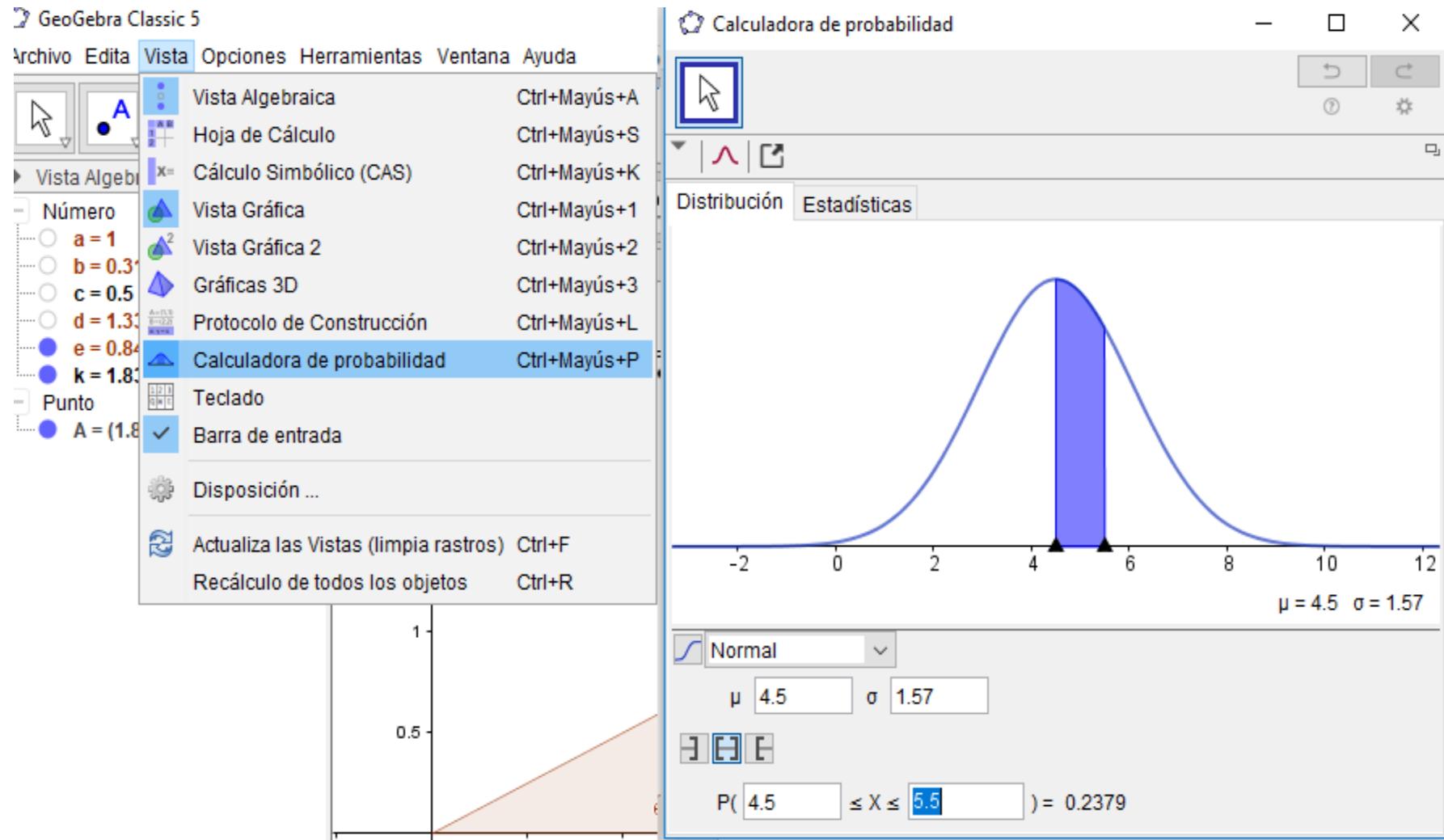
$$P(\chi = 5) = P\left(\frac{4.5 - 4.5}{1.57} < Z < \frac{5.5 - 4.5}{1.57}\right)$$

Normal[ <Media>, <Desviación Estándar>, x, <Booleana Acumulativa> ]

Normal[4.5, 1.57, 5.5,true]- Normal[4.5, 1.57, 4.5,true]=0.738-0.5=0.238

Si queremos comparar con el valor que se obtiene mediante el modelo binomial usamos, DistribuciónBinomial[ <Número de Ensayos>, <Probabilidad de Éxito>, <Valor de Variable>, <Acumulada Booleana> ] = DistribuciónBinomial[10, 0.45, 5,false ]= 0.234

## Otra forma de hacer lo mismo con **Calculadora de probabilidades**:



Vamos a hacer un estudio más exhaustivo de la aproximación del modelo binomial por el modelo normal, analizando modelos binomiales con diferentes medias y desviaciones típicas y comparándolo con el modelo normal

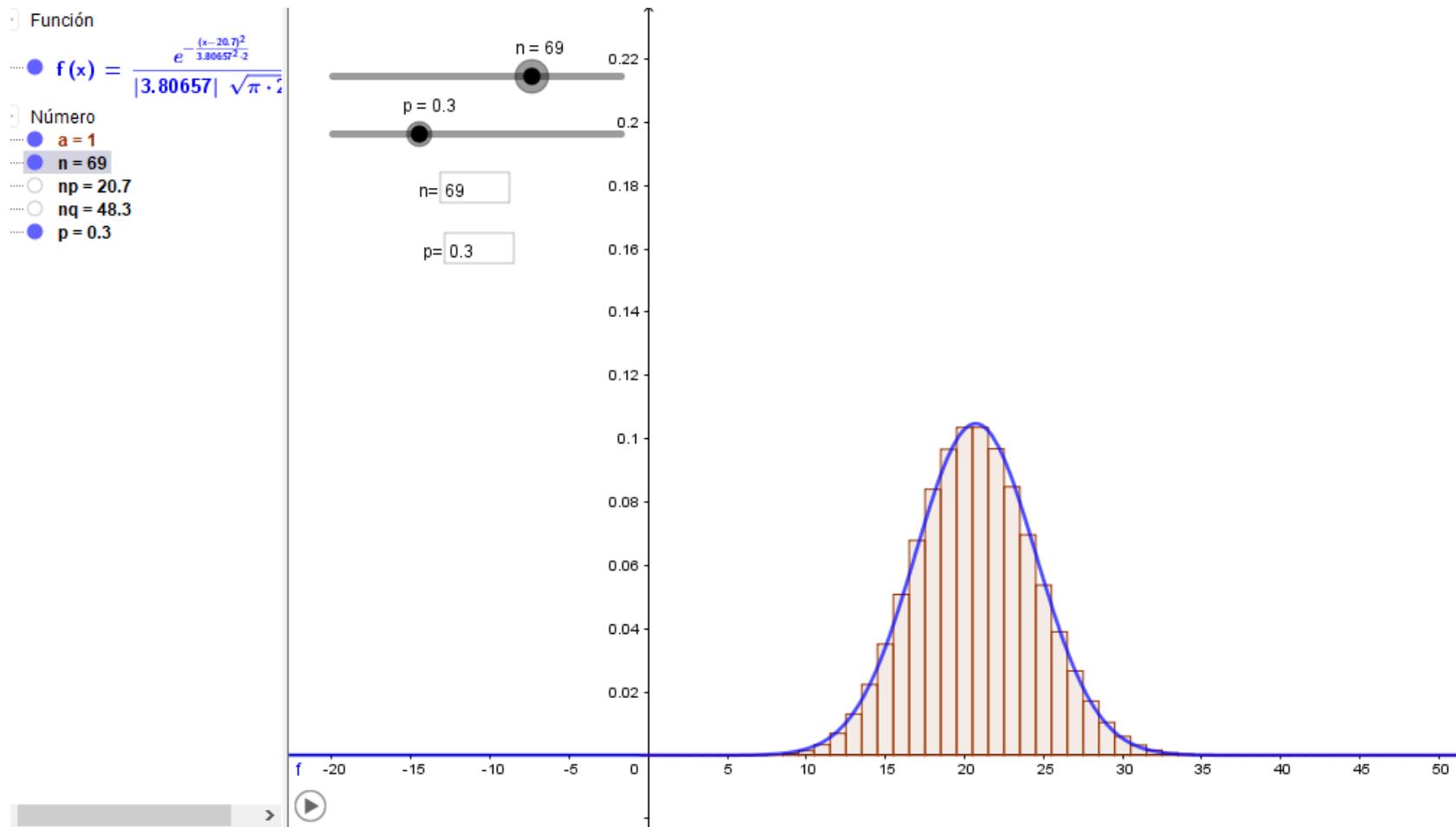
Vamos a crear dos **deslizadores**, uno que tomará los valores **para n** y otro **para p**.

El deslizador para n abarcará un rango que vaya, por ejemplo, desde 1 hasta 200, con pasos de 1 en 1. El deslizador para p irá desde 0 a 1, con pasos de 0.1.

Aunque especifiquemos así los pasos, es difícil que los saltos no sean mayores y que haya valores que nos saltemos. Para solventar este problema, podemos crear **casillas de entrada**. Les damos los nombres **n=** y **p=** y los **vinculamos al deslizador correspondiente**. Podemos **modificar la longitud de la casilla de entrada en estilo**.

# Sobre Distribución Binomial[n, p]

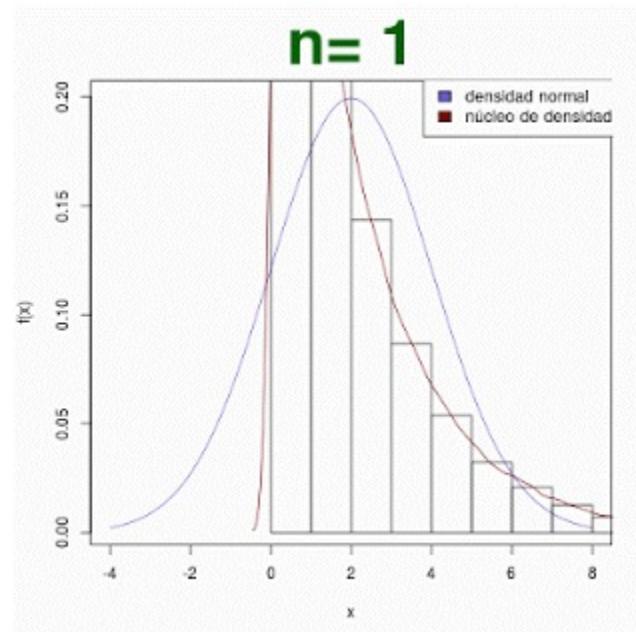
Ajustamos una curva normal de media  $\mu$  y desviación típica  $\sigma$ , con **Normal[n · p,  $\sqrt{n \cdot p(1 - p)}$ , x, false]**.



# DISTRIBUCIÓN DE LA MEDIA MUESTRAL TEOREMA CENTRAL DEL LÍMITE

# Teorema central del límite

Si  $\bar{X}$  es la media de una muestra aleatoria de tamaño  $n$  que se toma de una población con media  $\mu$  y varianza  $\sigma^2$ , entonces cuando  $n \rightarrow \infty$ , la variable  $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$  sigue una distribución normal tipicada.



## MODELO DE DISTRIBUCIÓN UNIFORME

En la casilla A1 ponemos **UniformeAleatoria[0, 1]**.

Extendemos hacia abajo para generar una muestra, tantas casillas como tamaño queramos para la muestra (tomemos  $n=40$ ).

A continuación, nos movemos hacia la derecha para generar más muestras del mismo tamaño.

En la casilla A41 escribimos **Media[A1:A40]**, se obtiene la media y nos movemos hacia la derecha para obtener las medias de todas las muestras. Creamos una lista con esas medias (llamémosla lista1).

En la casilla de Entrada ponemos **Histograma[Clases[lista1,10],lista1,false]**

Se ha usado la instrucción:

**Histograma[<Lista de límites de clase>,<Lista de datos brutos>,Usar densidad o no (true/false)>,<Factor de escala de densidad (opcional)>]**

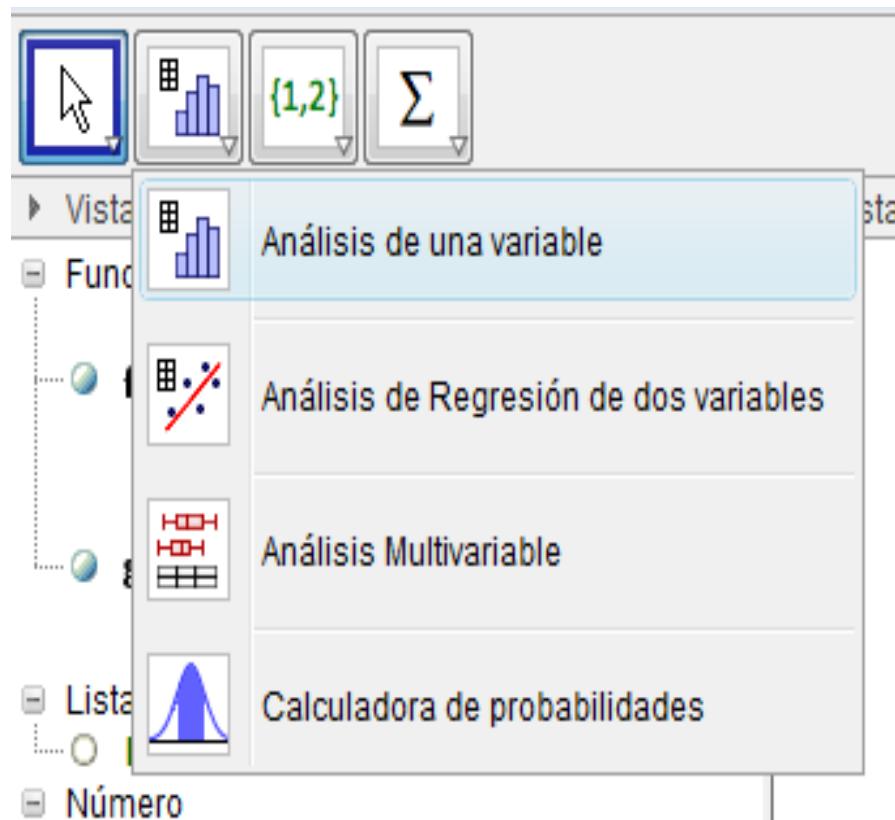
Para **<Lista de límites de clase>** se ha usado la instrucción **Clases[lista1,10]** para especificar el número de intervalos.

Para que la imagen se vea bien, pondremos en Vista gráfica como valores mínimo y máximo de la x, -0.1 y 1.1

A continuación, representaremos la curva normal con una media 0.5 y una desviación típica  $1/\sqrt{12*40}$

**Normal[0.5, 1/sqrt(12\*40),x,false]**

Si queremos hacer un estudio de la variable correspondiente a las medias muestrales, podemos seleccionar la columna y hacer clic sobre **Análisis de una variable**.



Si marcamos, histograma normalizado, podemos superponer la curva normal correspondiente de media la media de las medias muestrales y desviación típica la de las medias muestrales. También obtendremos un resumen de las medidas de centralización y dispersión de la variable medias muestrales.

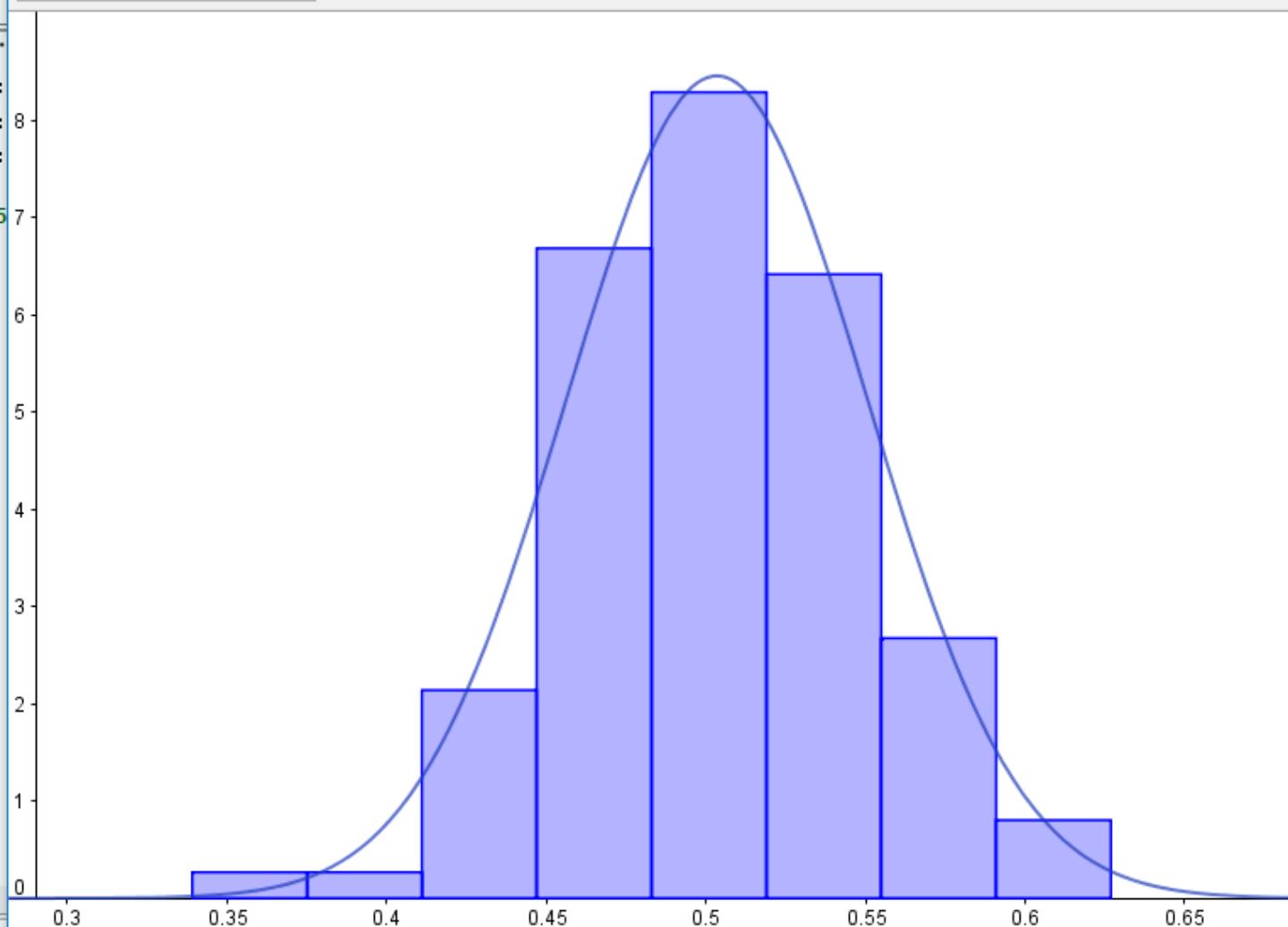
# Análisis de datos - distribución de la media muestral (uniforme).ggb

X



Σx 123  
456

Histograma



Histograma Gráfica

#### Clases

 Establecer las clases manualmente

#### Reglas de clase:

  $\leq x <$   $< x \leq$ 

#### Tipo de frecuencia

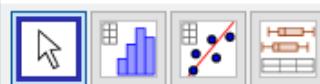
 Acumulada Cuenta Relativo Normalizado

#### Muestra

 Histograma Tabla de frecuencias Polígono de frecuencias Curva Normal

# Análisis de datos - distribución de la media muestral (uniforme).ggb

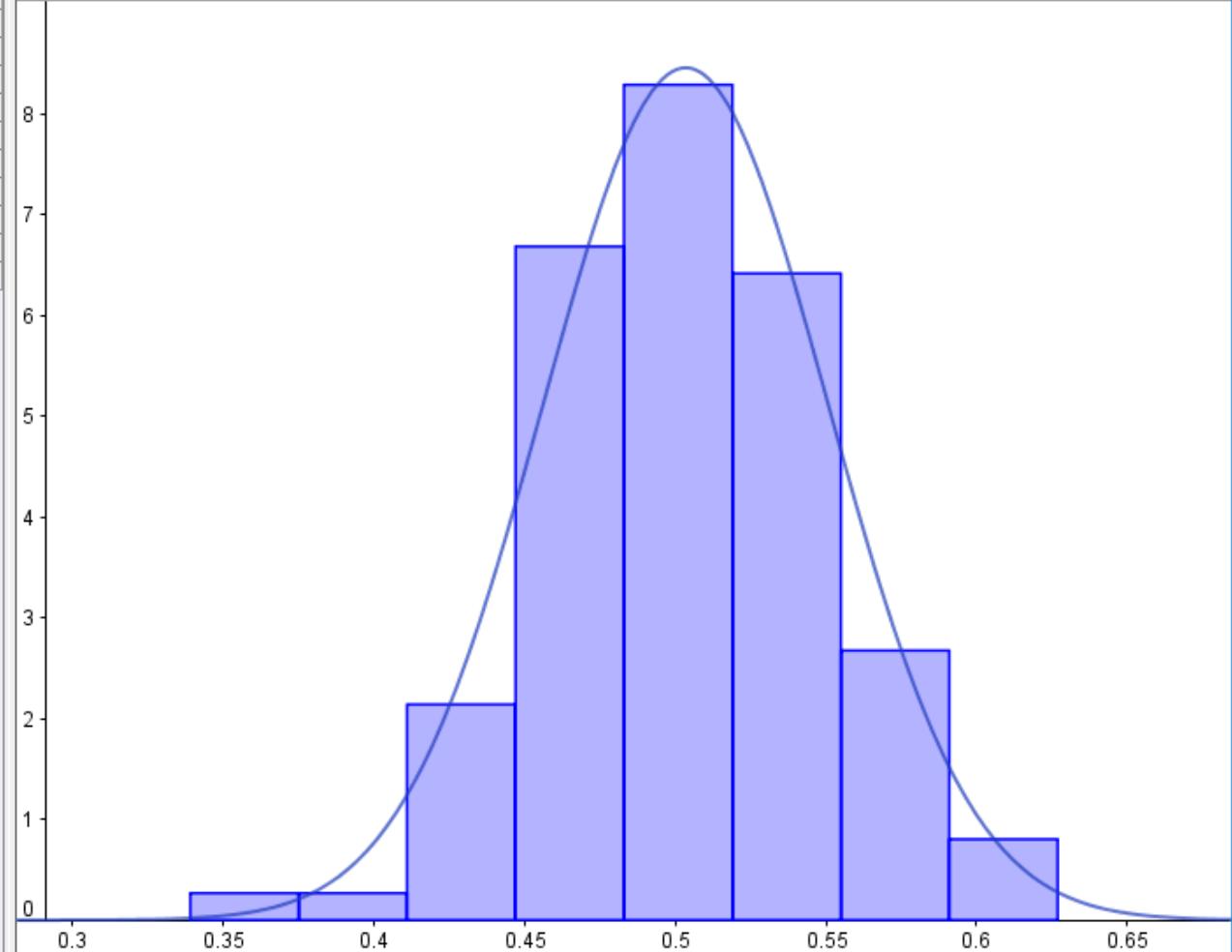
X



a  
cción de conjunto de celdas

n	104
Media	0.5035
$\sigma$	0.0472
s	0.0474
$\Sigma x$	52.3601
$\Sigma x^2$	26.5927
Mín	0.3392
Q1	0.4739
Mediana	0.5057
Q3	0.5326
Máx	0.6269

Histograma



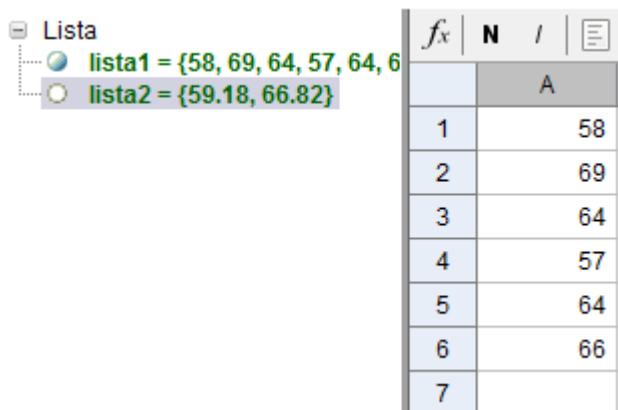
## A PRACTICAR

Analizad la media muestral para muestras procedentes de un modelo de Poisson

# **INTERVALOS DE CONFIANZA Y CONTRASTE DE HIPÓTESIS**

1. Un atleta efectúa seis lanzamientos, obteniendo distancias de 58, 69, 64, 57, 64 y 66 metros. Halla un intervalo de confianza para la media del 90%. Contrastar la hipótesis de que la media poblacional es mayor o igual que 66.

**IntervaloMediaT[ <Lista de datos (muestra)>, <Nivel> ]**



The screenshot shows the GeoGebra interface. On the left, there is a 'Lista' (List) panel with two entries: 'lista1 = {58, 69, 64, 57, 64, 66}' and 'lista2 = {59.18, 66.82}'. To the right is a table with columns labeled  $f_x$ , N, and I. The table has 7 rows, indexed from 1 to 7. Row 1 contains 'A'. Rows 2 through 7 contain the values 58, 69, 64, 57, 64, and 66 respectively.

$f_x$	N	I
A		
1	58	
2	69	
3	64	
4	57	
5	64	
6	66	
7		

En lista 2, geogebra nos devuelve los extremos del intervalo

Para realizar el contraste de hipótesis, consideramos como hipótesis alternativa  $H_1: \mu < 66$ , frente a la hipótesis nula  $H_0: \mu \geq 66$

**TestMediaT[ <Lista de datos (muestra)>, <Media (hipótesis)>, <Cola> ]**

TestMediaT[ <lista1, 66, “<” ].

La respuesta es **lista2 = {0.09, -1.58}**.

El **primer valor** corresponde al **p-valor**

El **segundo valor**, corresponde al **estadístico de contraste**.

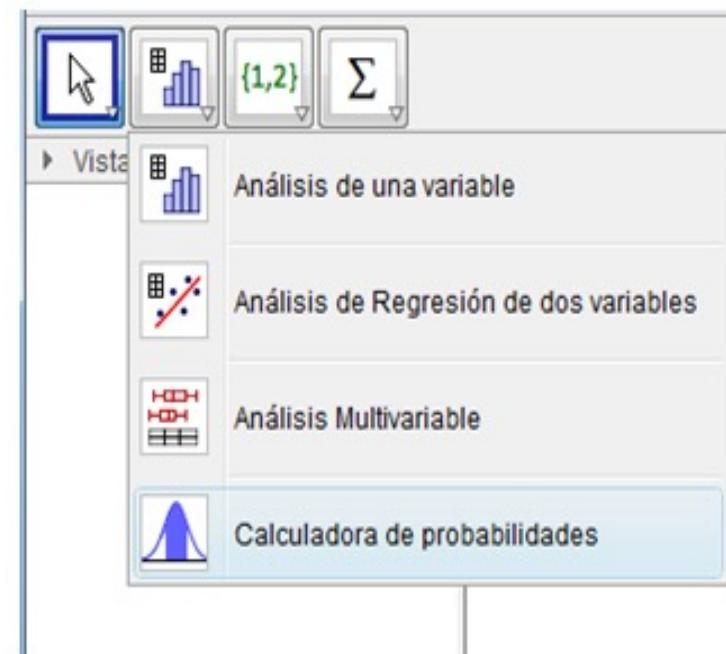
2. En una muestra de tamaño  $n=16$ , se mide una media 6 y una desviación típica  $s=12$ . ¿Es el valor de la media significativamente mayor que 0?

**En este caso no disponemos de los datos en bruto,  
podemos usar**

**TestMediaT[ <Media (muestra)>, <Desviación estándar (muestra)>, <Tamaño (muestra)>, <Media (hipótesis)>, <Cola> ]**

**Hay otra forma:**

**Vamos a calculadora  
de probabilidades,  
dentro del menú que  
ofrece la hoja de  
cálculo**



# O bien, dentro del menú de la vista gráfica

The screenshot shows the GeoGebra interface with the 'Vista Gráfica' menu open. The menu includes options like Texto, Imagen, Lápiz, Dibujo a mano alzada, Relación, Calculadora de probabilidades, and Analizador de funciones. To the right, a dialog box for a 'Test T de una media' is displayed, showing input fields for null hypothesis ( $\mu = 0$ ), alternative hypothesis (radio button selected for  $>$ ), sample statistics (Media: 6, s: 12, N: 16), and results (Media: 6, s: 12, ES: 3, N: 16, df: 15, t: 2, P: 0.032).

Distribución Estadísticas

Test T de una media

Hipótesis nula  $\mu = 0$

Hipótesis alternativa  <  >  ≠

Muestra

Media 6

s 12

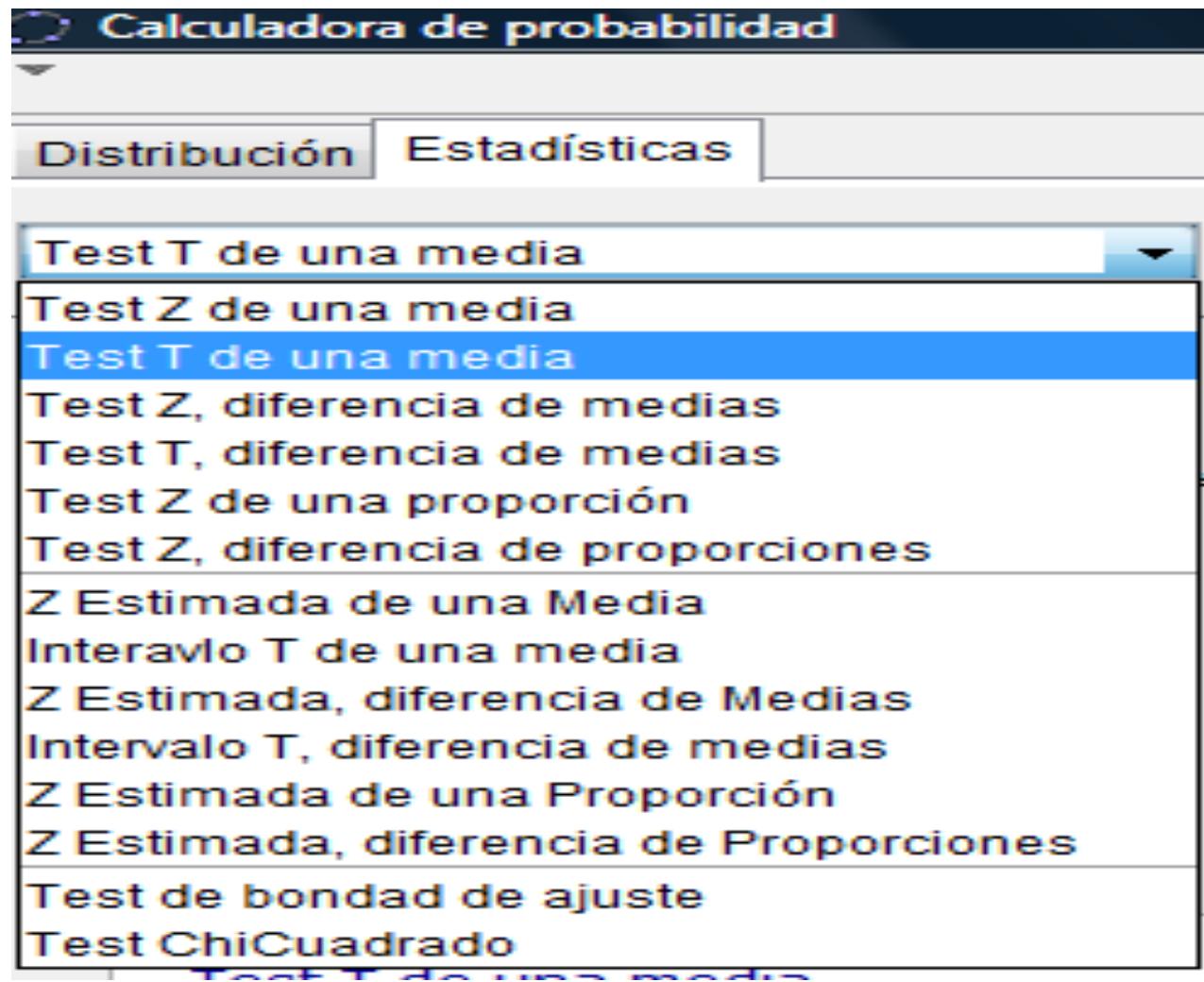
N 16

Resultado

Test T de una media

Media	6
s	12
ES	3
N	16
df	15
t	2
P	0.032

# Diferentes tests que se pueden utilizar



Si queremos obtener un intervalo T de confianza para la media...

Calculadora de probabilidad

Distribución Estadísticas

Interavlo T de una media

s 12  
N 16

Resultado

Interavlo T de una media

Media	6
s	12
ES	3
N	16
df	15
Límite por la derecha	-0.3943
Límite por la izquierda	12.3943
Intervalo	6 ± 6.3943

## A PRACTICAR

3. Se quiere averiguar si ha habido una reducción significativa en el porcentaje de votantes a un determinado partido político, en el último año. Para ello se eligieron al azar 100 personas y se les preguntó si votarían al partido en cuestión, obteniéndose un porcentaje de respuestas afirmativas del 39%. Si el porcentaje de votantes a favor del partido considerado era del 42% hace un año, cuando se preguntó a 150 personas, contrasta a nivel de significación  $\alpha=0.05$  si la reducción habida ha sido significativa.

# ANOVA

Para trabajar el análisis de la varianza, podemos teclear en la línea de entrada **ANOVA[ <Lista>, <Lista>, ... ]**

1. La siguiente tabla da el contenido de nitrógeno, en miligramos, de plantas de trébol rojo inoculadas con cultivos de *Rhizobium trifolii* más cinco cepas diferentes de *Rhizobium meliloti*, además de un compuesto de trébol rojo que se utilizó como control.

3DOk1	19.4	32.6	27	32.1	33
3DOk5	17.7	24.8	27.9	25.2	24.3
3DOk4	17	19.4	9.1	11.9	15.8
3DOk7	20.7	21	20.5	18.8	18.6
3DOk13	14.3	14.4	11.8	11.6	14.2
Compuesto	17.3	19.4	19.1	16.9	20.8

Contrasta a un nivel de significación  $\alpha=0.01$ , si todas las plantas de trébol rojo pueden considerarse homogéneas desde el punto de vista de la cantidad de nitrógeno que contienen.

Introducimos los datos utilizando la hoja de cálculo y creamos **varias listas, tantas como tratamientos diferentes haya.**

Utilizamos la instrucción ANOVA[ <Lista>, <Lista>, ... ]

**En una nueva lista aparecen dos valores**

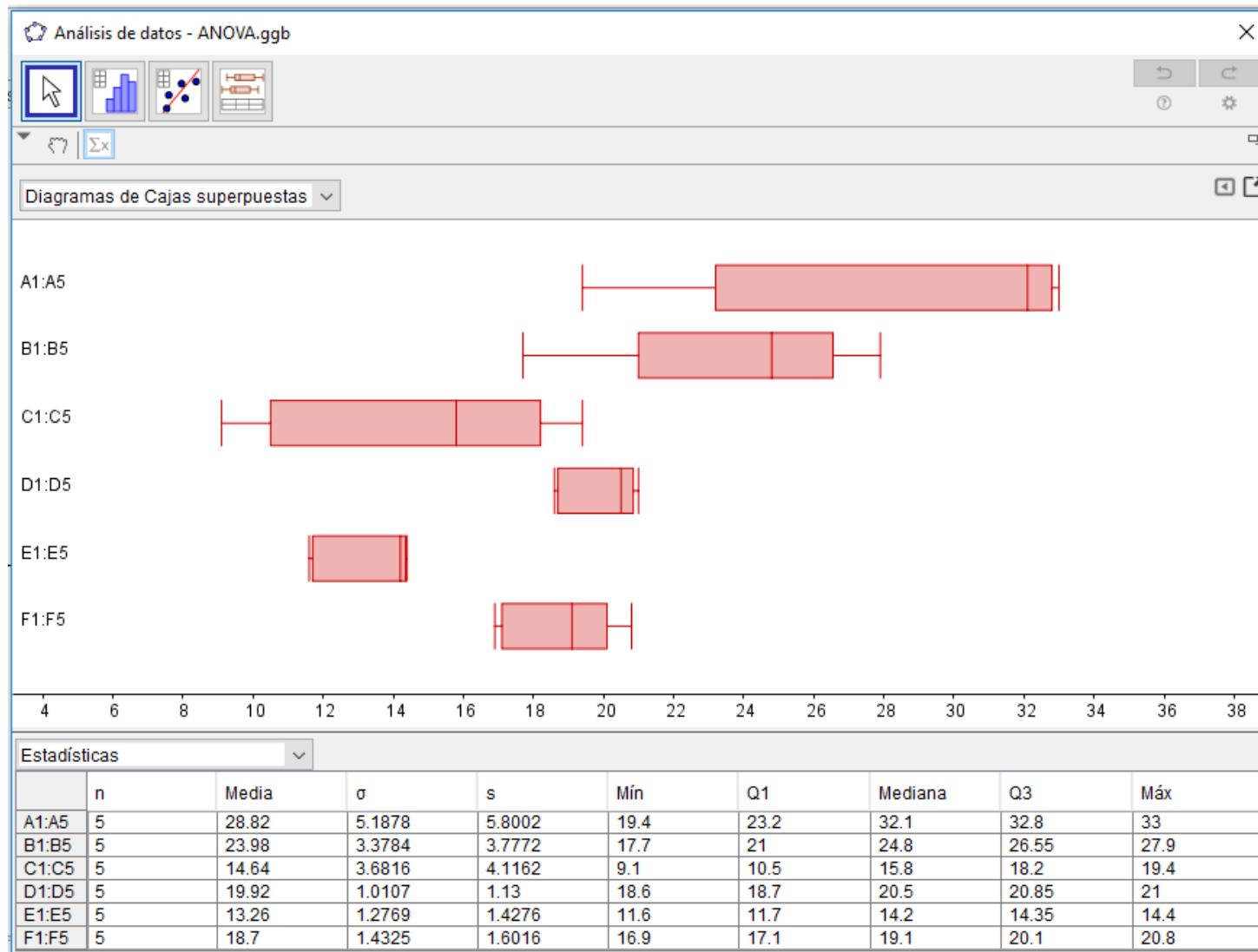
**El primero** corresponde al **p-valor**

**El segundo valor** corresponde al **estadístico de contraste** (cociente de cuadrados medios).

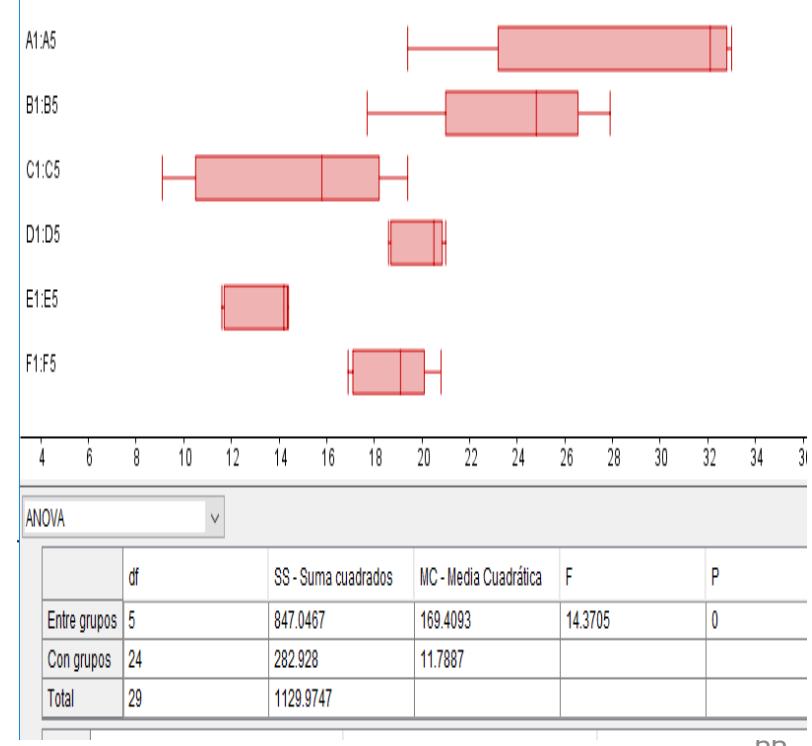
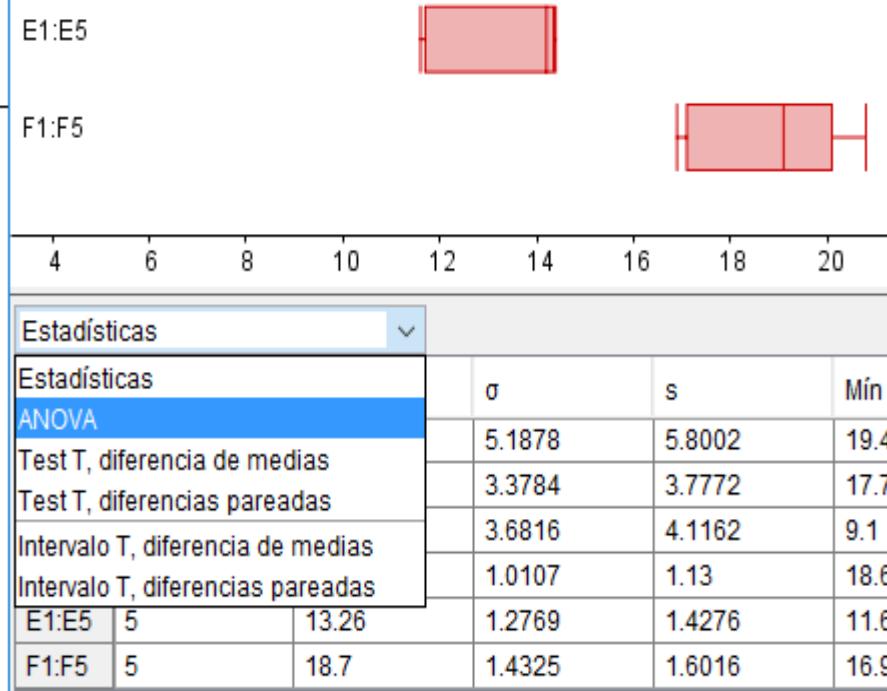
The screenshot shows the TI-Nspire CX CAS software interface. On the left, the 'Vista Algebraica' window displays a list of seven lists named lista1 through lista7, each containing numerical values. On the right, the 'Hoja de Cálculo' window shows a 7x7 grid of data. The columns are labeled A through F, and the rows are numbered 1 through 7. The data entries are as follows:

	A	B	C	D	E	F
1	19.4	17.7	17	20.7	14.3	17.3
2	32.6	24.8	19.4	21	14.4	19.4
3	27	27.9	9.1	20.5	11.8	19.1
4	32.1	25.2	11.9	18.8	11.6	16.9
5	33	24.3	15.8	18.6	14.2	20.8
6						
7						

Si queremos realizar un análisis individual de las variables, podemos utilizar “Análisis multivariable”



También podemos realizar el análisis de la varianza desde el mismo sitio, e incluso contrastes de hipótesis e intervalos de confianza para la diferencia de medias



## A PRACTICAR

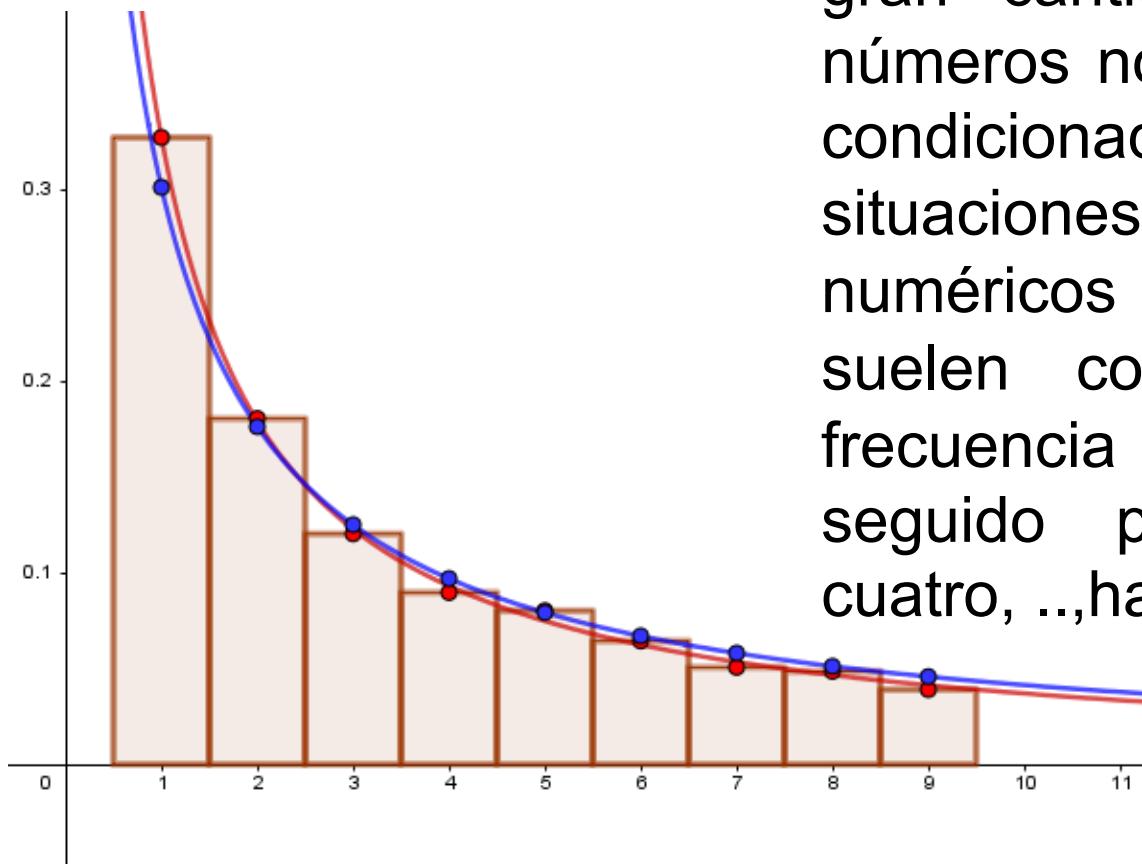
2. Se quiere averiguar si tres tipos de gasolina presentan diferencias significativas en cuanto a sus efectos contaminantes. Para ello se seleccionaron al azar doce vehículos en los que se aplicaron aleatoriamente los tres tipos de gasolina, obteniéndose los siguientes datos respecto a reducción de óxido de nitrógeno:

Gasolina I	23	26	25	25
Gasolina II	28	29	27	25
Gasolina III	22	25	26	27

Con estos datos, ¿pueden inferirse diferencias significativas entre los tres tipos de gasolina, con un nivel de significación  $\alpha=0.05$ ?

# **EL TEST DE LA CHI-CUADRADO**

# La ley de Benford

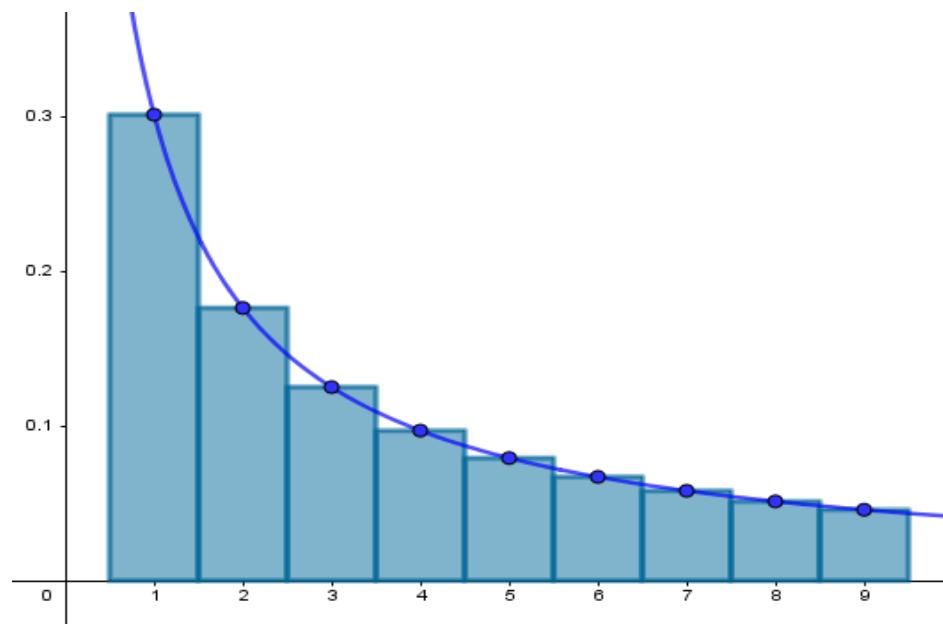


La ley de Benford, también conocida como ley del primer dígito, nos dice que, en una gran cantidad de series de números no aleatorios, ni muy condicionados que se dan en situaciones reales, los valores numéricos que se obtienen suelen comenzar con más frecuencia por el dígito uno, seguido por el dos, tres, cuatro, ..., hasta el nueve.

## Formulación matemática de la ley de Benford

En una distribución de Benford, la probabilidad de que el primer dígito de la serie considerada sea  $n$  es  $P(n) = \log\left(1 + \frac{1}{n}\right)$      $n=1, 2, 3, 4, 5, 6, 7, 8, 9$

Primer dígito	Probabilidad de aparición
1	0.301029995663982
2	0.176091259055681
3	0.124938736608300
4	0.0969100130080566
5	0.0791812460476250
6	0.0669467896306134
7	0.0579919469776869
8	0.0511525224473814
9	0.0457574905606752



# Ley de Benford en el censo

	A	B	C	D	E	F	
1	Dígitos	Frecuencias absolutas	Frecuencias relativas	Porcentajes Benford	Frecuencias esperadas	Chi cuadrado	
2	1	310	0.327	0.30103	285.37644	2.12463	
3	2	171	0.18038	0.17609	166.93451	0.09901	
4	3	114	0.12025	0.12494	118.44192	0.16659	
5	4	85	0.08966	0.09691	91.87069	0.51384	
6	5	76	0.08017	0.07918	75.06382	0.01168	
7	6	61	0.06435	0.06695	63.46556	0.09578	
8	7	48	0.05063	0.05799	54.97637	0.88528	
9	8	46	0.04852	0.05115	48.49259	0.12812	
10	9	37	0.03903	0.04576	43.3781	0.9378	
11		948				4.96273	
12							

$$\sum_{i=1}^9 \frac{(o_i - e_i)^2}{e_i}$$

ChiCuadradoInversa( <Grados de libertad>, <Probabilidad> )  
ChiCuadradoInversa(8, 0.95) =  $\chi^2_{9-1;0.05} = 15.50731$

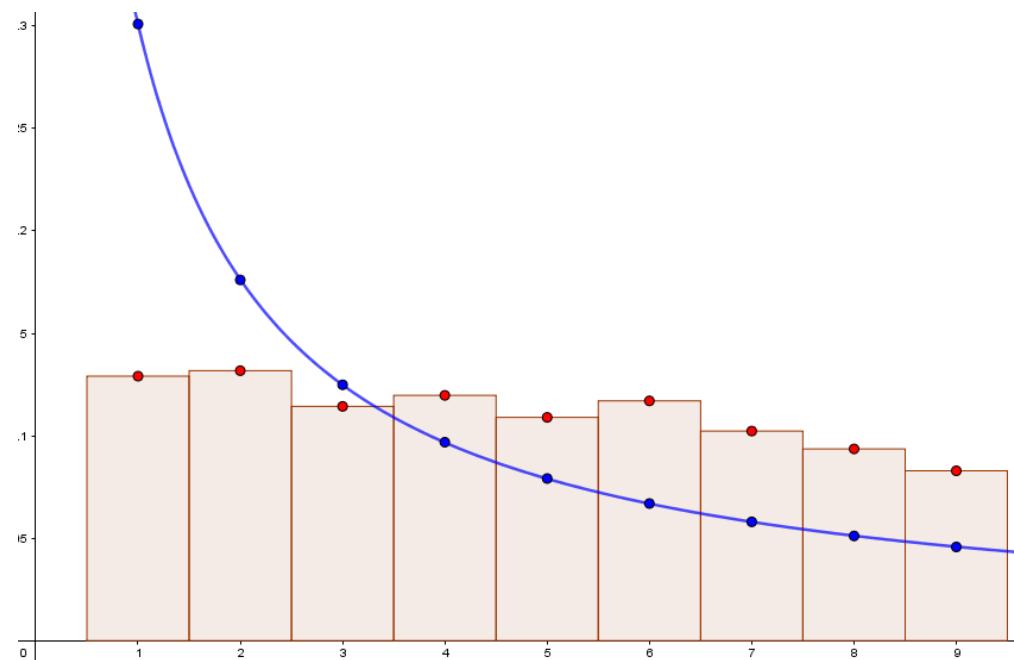
# Ley de Benford en la lotería

	A	B	C	D	E	F	
1	Dígitos	Frecuencias absolutas	Frecuencias relativas	Porcentajes Benford	Frecuencias esperadas	Chi cuadrado	
2	1	310	0.327	0.30103	285.37644	2.12463	
3	2	171	0.18038	0.17609	166.93451	0.09901	
4	3	114	0.12025	0.12494	118.44192	0.16659	
5	4	85	0.08966	0.09691	91.87069	0.51384	
6	5	76	0.08017	0.07918	75.06382	0.01168	
7	6	61	0.06435	0.06695	63.46556	0.09578	
8	7	48	0.05063	0.05799	54.97637	0.88528	
9	8	46	0.04852	0.05115	48.49259	0.12812	
10	9	37	0.03903	0.04576	43.3781	0.9378	
11		948				4.96273	
12							

$$\sum_{i=1}^9 \frac{(o_i - e_i)^2}{e_i}$$

ChiCuadradoInversa( <Grados de libertad>, <Probabilidad> )  
ChiCuadradoInversa(8, 0.95) =  $\chi^2_{9-1;0.05}$  = 15.50731

	A	B	C	D	E	F	I
1	Dígitos	frecuencias	frecuencias relativas	Porcentajes Benford	Esperados	chicuadrado	
2	1	193	0.1291	0.30103	450.03984	146.80807	
3	2	197	0.13177	0.17609	263.25643	16.67543	
4	3	171	0.11438	0.12494	186.78341	1.33372	
5	4	179	0.11973	0.09691	144.88047	8.03519	
6	5	163	0.10903	0.07918	118.37596	16.82187	
7	6	175	0.11706	0.06695	100.08545	56.07398	
8	7	153	0.10234	0.05799	86.69796	50.70431	
9	8	140	0.09365	0.05115	76.47302	52.77256	
10	9	124	0.08294	0.04576	68.40745	45.17829	
11		1495				394.40342	
12							



# EPIDEMIOLOGÍA

*Se realiza una investigación sobre una nueva vacuna contra la gripe. Se elige una muestra aleatoria de 900 individuos y se clasifica a cada uno de ellos según haya contraído la gripe o no durante el último año y según haya sido o no vacunado. Se obtiene la siguiente información:*

		Contraída la gripe	
Vacunado		Sí	No
Sí	150	200	
	300	250	

¿Existe relación entre haber recibido o no la vacuna y contraer la gripe?

# Introducimos los recuentos en forma de tabla y se crea una matriz

The screenshot shows a software interface with a sidebar on the left and a main workspace on the right.

**Sidebar (Left):**

- Lista
  - matriz1 =  $\begin{pmatrix} 150 & 200 \\ 300 & 250 \end{pmatrix}$
- Texto
  - A2 = "Sí"
  - A3 = "NO"
  - B1 = "Sí"

**Main Workspace (Right):**

A table with columns labeled A, B, and C, and rows labeled 1, 2, and 3. The table contains the following data:

	A	B	C
1		SI	NO
2	Sí	150	200
3	NO	300	250

## TestChicuadrado[matriz]

...  
• **Resultado** = {0.00063, 11.68831}  
...  
• matriz1 =  $\begin{pmatrix} 150 & 200 \\ 300 & 250 \end{pmatrix}$

El primer valor es el p-valor y el segundo el estadístico de contraste

# **EL MODELO SIR EN EPIDEMIOLOGÍA**

- Las matemáticas aportan modelos matemáticos para describir y explicar la evolución de las epidemias y para predecir el número de afectados por la misma.
- Uno de los modelos más utilizados en el terreno de la epidemiología es el modelo SIR (Población de Susceptibles-Población de Infectados-Población de Recuperados).
- Con él se explica y predice la evolución de agentes infecciosos que pueden ocasionar daños en poblaciones, sobre todo en la población humana.
- A través de este modelo se establece la relación entre las tres poblaciones mencionadas utilizando, sobre todo, tres elementos fundamentales, como son incidencia, prevalencia y número reproductivo básico.

# **CONCEPTOS BÁSICOS EN EPIDEMIOLOGÍA**

**Epidemiología:** Es la ciencia que estudia cómo se distribuyen las enfermedades y otros eventos de salud en las poblaciones humanas, sus patrones de aparición y los factores ambientales y sociales que los determinan.

Una **Pandemia** es una epidemia que alcanza proporciones geográficas muy extensas.

**Número reproductivo básico:** Es el principal parámetro utilizado en Epidemiología. Se le representa como  $R_0$  y viene definido como el número promedio de infecciones causadas por un individuo infectado cuando está en medio de la población de susceptibles. Con este número se intenta establecer la capacidad reproductiva de la enfermedad, es decir, la velocidad a la que una enfermedad puede propagarse en una población

¿Cómo interpretamos este número reproductivo básico?

Si  $R_0 > 1$  los individuos infectados al inicio de la enfermedad infectarán a otros individuos susceptibles y se producirá un brote epidémico.

Si  $R_0 = 1$  cada individuo se reemplaza a sí mismo y no habrá brote epidémico.

Si  $R_0 < 1$  los individuos infectados serán transmisores poco eficientes y la enfermedad terminará por desaparecer

**Tamaño de la epidemia:** Es el número total de individuos infectados. Es un valor muy importante, ya que está relacionado con los costos de la epidemia.

**Intervalo de serie:** Es el tiempo transcurrido entre la aparición de la enfermedad en un caso y el siguiente.

**Incidencia:** Es la proporción de casos nuevos de la enfermedad infecciosa, que se presentan durante un periodo de tiempo determinado. La incidencia nos da la probabilidad de que una persona en esa población resulte contagiada (riesgo medio de contraer la enfermedad). La incidencia es más útil cuando se trata de enfermedades de corta duración como la varicela

$$\frac{Nº\ de\ casos\ nuevos}{Nº\ de\ susceptibles}$$

**Prevalencia:** Es la proporción de casos existentes en una población en un momento dado o en un período de tiempo determinado. Digamos que es la probabilidad de contraer una enfermedad. Por lo general se representa en forma de fracción, porcentaje o número de casos por cada 10000 o 100000 personas. La prevalencia es más útil cuando se habla de enfermedades duraderas como el AIDS

$$\frac{Nº\ de\ infectados}{Nº\ de\ susceptibles}$$

**Factor de Riesgo:** Es una característica del individuo que determina que tenga mayor probabilidad de contraer una enfermedad, de sufrirla con mayor virulencia o de morir por causa de ella.

## El modelo SIR en epidemiología

El modelo SIR es más apropiado para modelizar enfermedades infecciosas que se desarrollan a lo largo del tiempo y en las que superar la enfermedad implica inmunidad.

Poblaciones consideradas: **Individuos susceptibles (S)**, **Individuos infectados (I)** e **individuos recuperados o fallecidos (R)**.

En el modelo SIR se relacionan las variaciones de las tres poblaciones mencionadas a través de la tasa de infección, que es la probabilidad de que una persona enferme al entrar en contacto con un infectado y el período infeccioso promedio que es el tiempo que un individuo tarda en recuperarse.

$$\begin{cases} S'(t) = -\frac{a}{N} S(t)I(t) \\ I'(t) = \frac{a}{N} S(t)I(t) - bI(t) \\ R'(t) = bI(t) \end{cases}$$

$\frac{a}{N} \rightarrow$  Tasa de transmisión  
Expresa la velocidad con la que se pasa de S a I  
 $b \rightarrow$  tasa de recuperación e indica las personas que pasan de I a R

## Nuestro trabajo

- Se simula con Geogebra el modelo SIR de manera discreta, de modo que se calcula el número de susceptibles, infectados, recuperados y fallecidos dando saltos diarios.
- Aunque lo lógico sería obtener valores enteros positivos, esto no es posible, dado que se trabaja con probabilidades y tasas de prevalencia e incidencia.
- Hemos simulado diversas situaciones obtenidas modificando el número medio de contactos por persona (este valor está relacionado con el cumplimiento de las normas referentes a vida social, profesional o familiar), la probabilidad de contagio y la tasa de supervivencia.
- La modificación de estos parámetros se ha realizado usando deslizadores.
- Hemos hecho una simulación simple de la evolución de la pandemia en la provincia de Albacete.
- Hemos tomado el número de habitantes (388167), el número de infectados (4063) y de fallecidos (577) hasta el 26 de julio.
- Se ha considerado que el número de días durante los cuales se podía transmitir la enfermedad era de 15 y el estudio se ha extendido durante 200 días (de primeros de marzo a septiembre).

- Los valores obtenidos corresponden a la población de susceptibles, infectados, recuperados, fallecidos, contagios nuevos diarios, número de individuos recuperados y fallecidos a diario.
- Con los datos iniciales se podría calcular la tasa de prevalencia, tasa de incidencia, tasa de mortalidad, población total de contagios, población total de recuperados, población total de fallecidos, la población máxima infectada simultáneamente, el máximo de contagios diarios, el máximo de recuperaciones diarias y el máximo de fallecimientos al día.
- Se podría obtener una aproximación muy rudimentaria al número reproductivo básico usando cocientes entre contagiados y recuperados ( $\frac{\beta}{\gamma} = \frac{\text{tasa de transmisión}}{\text{tasa de recuperación}}$ ).

Actividad 1: Considerando que el número de habitantes de la ciudad de Albacete es de aproximadamente 388167 y que, hasta el día 27 de julio, la información oficial que se tiene es de 4063 infectados y 577 fallecidos por coronavirus en la provincia, simula, con la ayuda de GeoGebra, la evolución diaria de la enfermedad en cuanto a personas susceptibles, infectadas, recuperadas y fallecidas, así como el número diario de contagios nuevos, personas recuperadas y fallecidas.

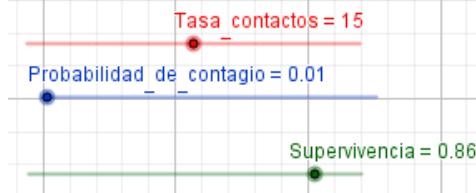
Consideraremos que la duración media de la enfermedad es de 15 días y que es durante este período de tiempo en el que se puede contagiar la enfermedad.

Para poder estudiar, sin repetir los cálculos y procedimientos, la evolución de la enfermedad en muy diversas situaciones usaremos deslizadores.

Construye tres deslizadores. El primero de ellos te permitirá modificar la tasa diaria de interacción, entendiéndose ésta como el número medio de contactos diarios por persona. Estableceremos un rango de valores de 0 a 30.

El segundo deslizador tendrá como misión el permitir modificar la probabilidad de contagio. ¿Cuál debería ser el rango de valores para esta tasa?

Con el tercer deslizador modificaremos la tasa de mortalidad o, si se prefiere, su complementario, es decir, la tasa de recuperación. Al igual que para el segundo deslizador, ¿cuál sería el rango de valores para esta tasa?



Comienza el estudio usando los valores oficiales para el cálculo de la probabilidad de contagio y la tasa de supervivencia y después modifica dichos valores para ver qué ocurre.

¿Según los resultados obtenidos, crees que es importante tener en cuenta las medidas de distanciamiento social limitando el número medio de contactos diarios?

Calcula una aproximación al número reproductivo básico

¿Cuál es el número máximo de infectados? ¿Cuántos de los infectados se han recuperado? ¿Cuántos han fallecido? ¿Qué cantidad de habitantes no se han visto afectados por la enfermedad?

¿Cuál ha sido la máxima cantidad de gente infectada a la vez? ¿Y el máximo número de contagios, recuperados o fallecidos diarios?

Representa gráficamente la evolución de la pandemia teniendo en cuenta los valores obtenidos.

- El número de susceptibles se calcula restando cada día los nuevos contagios al número de susceptibles.
- Para recalcular diariamente el total de infectados sumamos a los del día anterior los nuevos contagios. Lo mismo para los recuperados y fallecidos.
- El incremento de los contagios se obtiene de manera proporcional al número de susceptibles, número de infectados, a la tasa diaria de interacción y a la probabilidad de contagiarse.
- El número de nuevas recuperaciones es directamente proporcional al número de infectados y a la tasa de recuperación e inversamente proporcional a la duración media de la enfermedad. Algo similar ocurre para el número de fallecidos, sustituyendo tasa de recuperación por tasa de mortalidad.
- Para evitar problemas con GeoGebra, se recomienda introducir las fórmulas en las celdas en el siguiente orden: G3 / B3 / H3 / I3 / C3 / D3 / E3. Después, es conveniente seleccionar toda la fila y arrastrar hacia abajo, para generar los valores de la manera correcta.

Valores numéricos de las variables											
	A	B	C	D	E	F	G	H	I	J	K
1	Tiempo	Susceptibles	Infectados	Recuperados	Fallecidos		Contagios diarios	Recuperaciones diarias	Fallecimientos diarios		
2	0	388167	19	0	0						
3	1	388164...	20.5831...	1.08933...	0.17733...		2.849860505	1.089333333	0.177333333		Duración Media de la enfermedad
4	2	388161...	22.2982...	2.26943...	0.36944...		3.0873067009	1.1801031134	0.1921098092		Tasa diaria de interacción
5	3	388157...	24.1562...	3.54787...	0.57756...		3.3445314584	1.2784351567	0.2081173511		Probabilidad de contagio
6	4	388154...	26.1690...	4.93283...	0.80301...		3.6231813863	1.3849592832	0.225458488		Tasa de recuperación
7	5	388150...	28.3494...	6.43318...	1.04726...		3.9250400233	1.5003577305	0.2442442817		Mortalidad
8	6	388145...	30.7115...	8.05855...	1.31185...		4.2520391981	1.6253695098	0.2645950365		
9	7	388141...	33.2703...	9.81935...	1.59849...		4.6062713277	1.7607951231	0.2866410666		Población total contagiados
10	8	388136...	36.0423...	11.7268...	1.90902...		4.9900027295	1.907501671	0.3105235278		Población total recuperados
11	9	388130...	39.0452...	13.7932...	2.24541...		5.4056880313	2.0664283828	0.3363953181		Población total fallecidos
12	10	388125...	42.2981...	16.0318...	2.60984...		5.8559857675	2.2385926044	0.3644220519		Población no afectada
13	11	388118...	45.8220...	18.4569...	3.00462...		6.3437752578	2.4250962815	0.3947831156		
14	12	388111...	49.6394...	21.0841...	3.43229...		6.8721748734	2.6271329775	0.4276728103		Máximo de población infectada simultánea
15	13	388104...	53.7747...	23.9301...	3.89559...		7.4445618023	2.8459954717	0.4633015884		
16	14	388096...	58.2543...	27.0131...	4.39749...		8.0645934337	3.0830839836	0.5018973927		Máximo de contagios diarios
17	15	388087...	63.1069...	30.3530...	4.94120...		8.7362304934	3.3399150749	0.5437071052		Máximo de recuperaciones diarias
18	16	388078...	68.3635...	33.9712...	5.53020...		9.4637620699	3.6181312849	0.5889981161		Máximo de fallecimientos al día
19	17	388067...	74.0578...	37.8907...	6.16826...		10.2518326813	3.9195115579	0.638060021		

- Para representar gráficamente la evolución se utilizarán dos gráficos. El primero de ellos para representar a susceptibles, infectados, recuperados y fallecidos. Se crearán listas con los datos obtenidos para los cuatro parámetros y para el tiempo y se representará cada una de ellas como función del tiempo. Se puede usar la instrucción **Poligonal((listatiempo,listaparametro))**

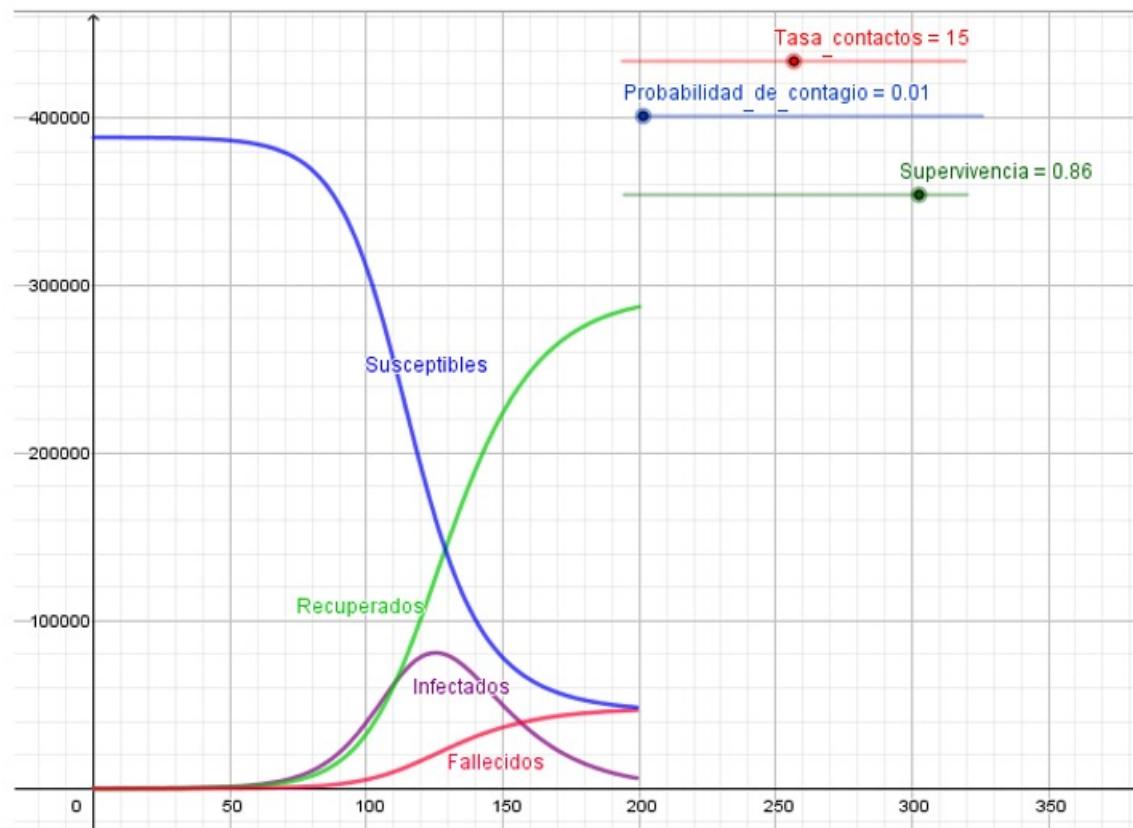


Figura 3: Representación gráfica de la evolución diaria de la enfermedad infecciosa (tasa media de contactos por persona =15)

- Del mismo modo se representará la evolución correspondiente al incremento diario de contagiados, recuperados y fallecidos

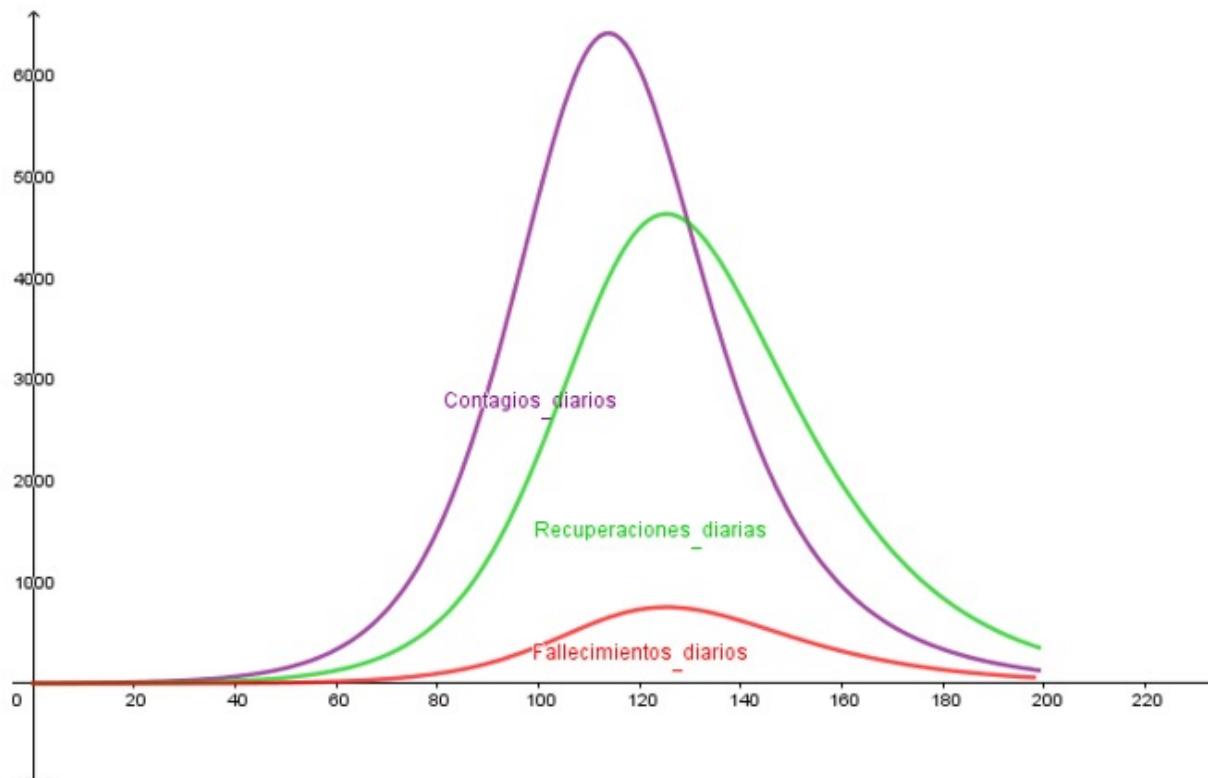


Figura 4: Representación gráfica de los casos nuevos (tasa media de contactos por persona =15)

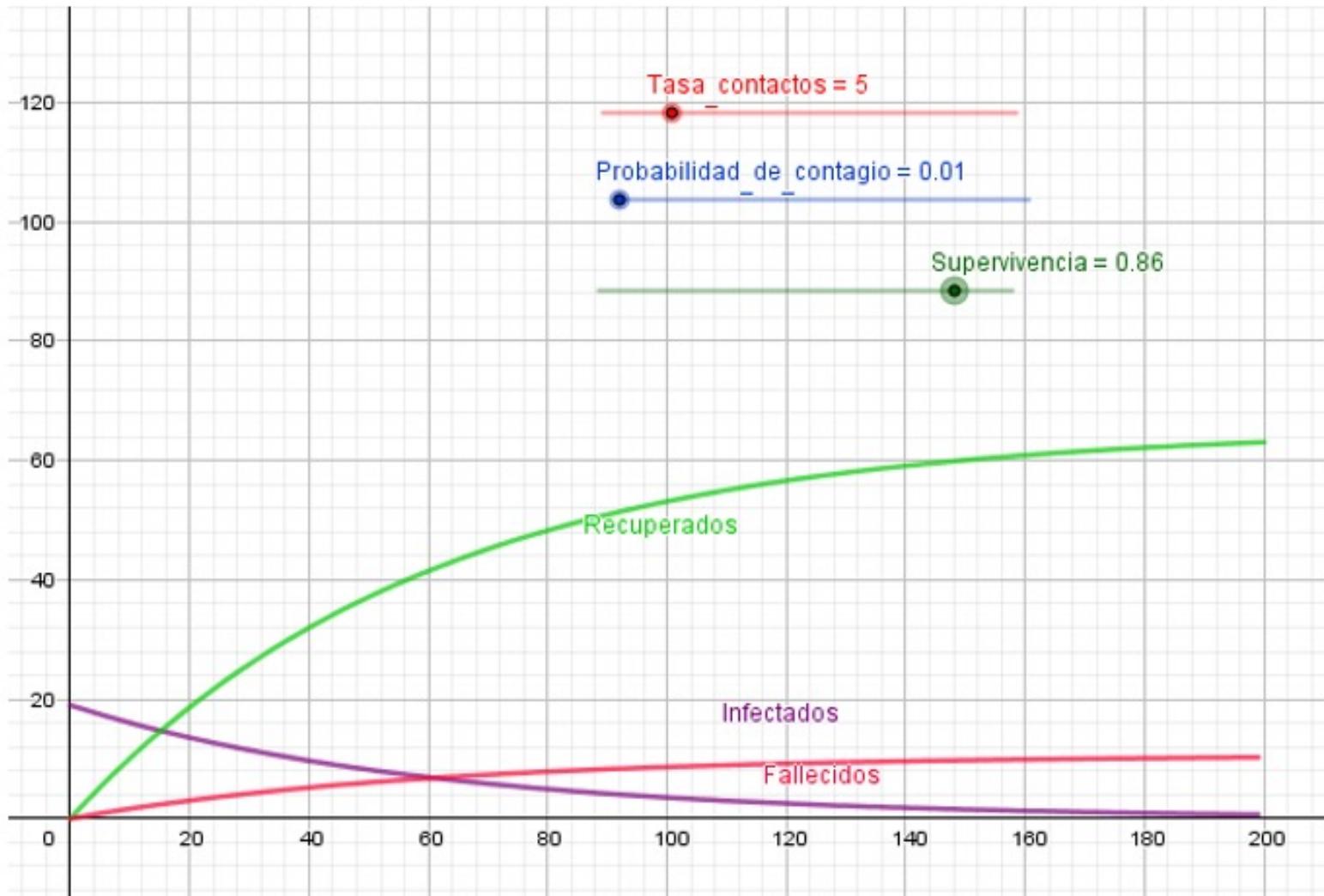


Figura 5: Representación gráfica de la evolución diaria de la enfermedad infecciosa (tasa media de contactos por persona =5)

# Investigación Operativa

# Investigación operativa con GeoGebra

- En general, GeoGebra no incluye herramientas específicas para resolver problemas de optimización e investigación operativa.
- Sin embargo, sí que podemos hacer uso de GeoGebra para representar gráficamente estos problemas.
- Muy útil para visualizar cómo resolver este tipo de problemas.

# Programación lineal

- Un problema de programación lineal viene definido por:
  - Variables de decisión (continuas y no negativas).
  - Función objetivo (a maximizar o minimizar).
  - Restricciones.
- Tanto la función objetivo como las restricciones son lineales en las variables de decisión.

# Refinería de petróleo<sup>1</sup>

Una empresa refina petróleo de Arabia Saudí y Venezuela, produce gasolina, fuel de avión y lubricantes. Para producir un barril de esos derivados, necesita utilizar la cantidad de barriles que se indica en la tabla siguiente. Ahí también aparece la cantidad máxima de barriles de Arabia Saudí y Venezuela disponibles por día, así como la cantidad mínima requerida de barriles de gasolina, fuel de avión y lubricantes.

<sup>1</sup> El ejemplo ha sido tomado de los apuntes de la asignatura "Investigación Operativa" de José Niño Mora (UC3M), que están disponibles en <https://ocw.uc3m.es/course/view.php?id=171> .

# Refinería de petróleo

(barriles/barril)	A. Saudí	Venezuela	requerido/día
Gasolina	0,3	0,4	$\geq 2000$
Fuel de avión	0,4	0,2	$\geq 1500$
Lubricantes	0,2	0,3	$\geq 500$
Disponible/día	$\leq 9000$	$\leq 6000$	
coste (€/barrel)	20	15	

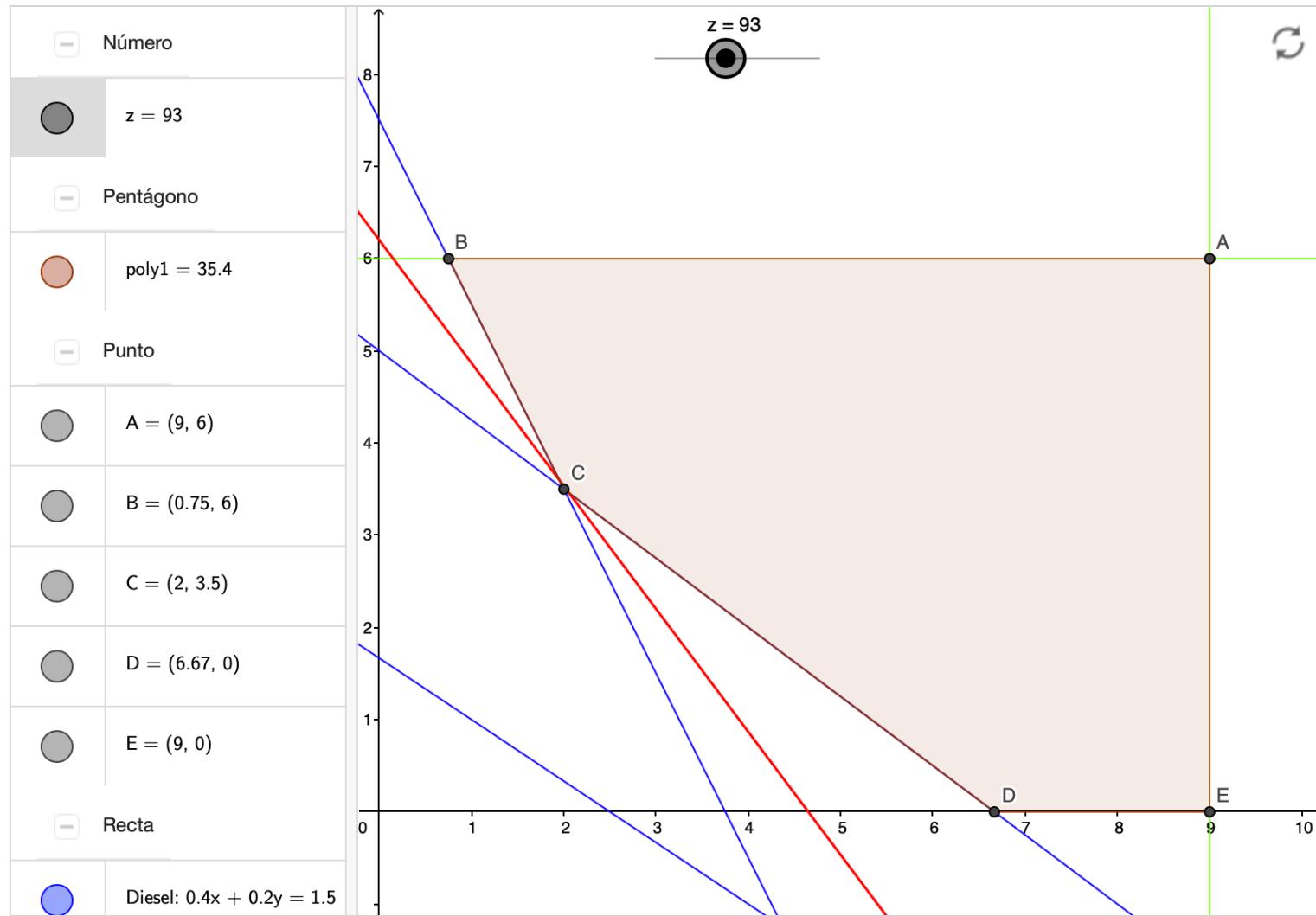
¿Cómo podemos minimizar el coste de la producción y a la vez cumplir las restricciones diarias de disponibilidad y producción?

# Refinería de petróleo

GeoGebra nos permite:

- Dibujar la región factible (como un polígono).
- Dibujar la función objetivo para un valor, y asociarlo a un deslizador para ver cómo varía.

# Refinería de petróleo



# **Geogebra y R: El paquete GeogebraR**

# Integración de Geogebra y R

- R es el software estadístico más utilizado en el mundo académico (<http://www.r-project.org>)
- Gran capacidad de análisis y visualización de datos
- Se podría utilizar para incrementar las rutinas estadísticas de Geogebra
- Los desarrolladores de Geogebra ya han planteado la posibilidad de integrar R en Geogebra:  
<https://dev.geogebra.org/trac/wiki/GeoGebraStatistics>

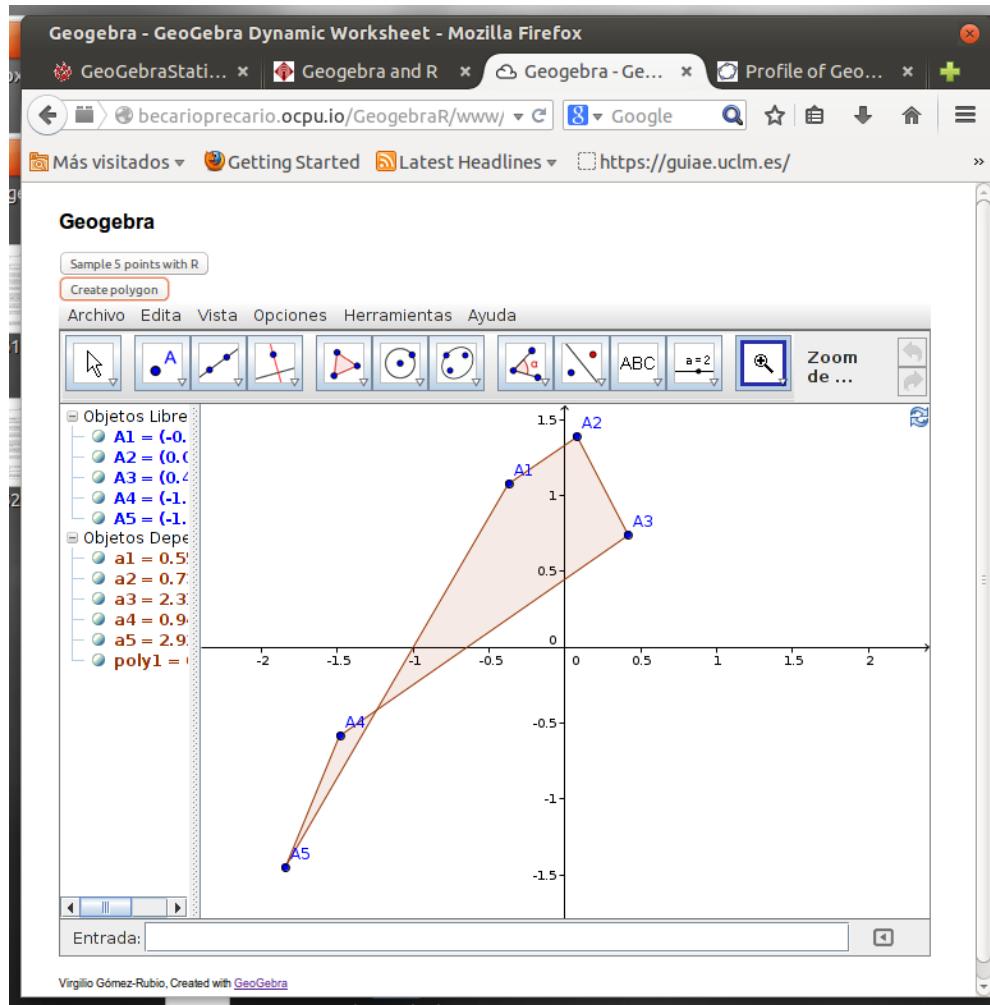
# GeogebraR

- Hemos desarrollado un paquete de R llamado GeogebraR, que integra Geogebra y R en una web
- Basado en el paquete de R opencpu, que permite que R se comporte como un servidor web
- opencpu incluye una API en Javascript para llamar a R
- Además, es posible hacer disponibles los ejemplos en una web  
<http://becarioprecario.ocpu.io/GeogebraR/www/>
- También se pueden ejecutar localmente instalando el paquete GeogebraR

# Estructura

- Código en Javascript para llamar a R
- Código en R que es visible desde Javascript (no se permite el acceso directo a los paquetes de R)
- Posibilidad de enviar y recibir datos entre Geogebra y R
- Opciones para el desarrollo de gráficos con R y mostrarlos en la web
- Geogebra también proporciona una API para manipular objetos desde Javascript

# Ejemplo: Polígono aleatorio



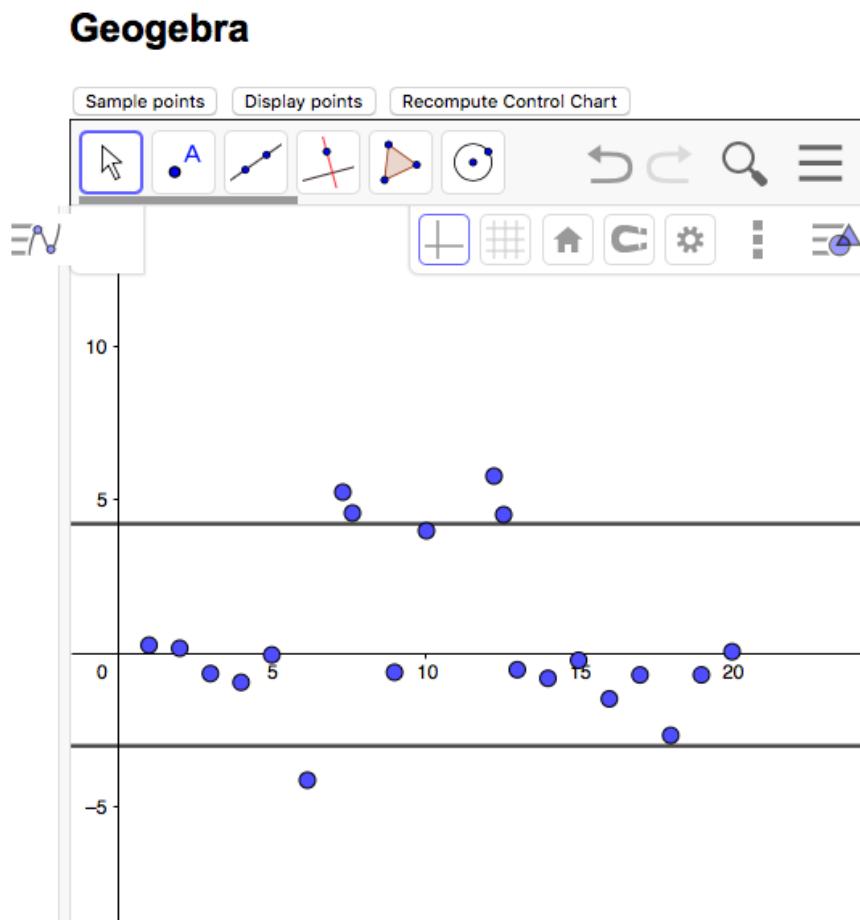
# Ejemplos

- ❑ Disponibles en el paquete GeogebraR  
<http://becarioprecario.ocpu.io/GeogebraR/www/>
- ❑ Teorema Central del Límite
- ❑ Control de Calidad
- ❑ Mapas
- ❑ Procesos puntuales (casos y controles)
- ❑ Ajuste de curvas no lineales

# Ejemplo: Control de Calidad

- ❑ Estudio de la sensibilidad de las bandas de control de calidad para el estudio de la media del proceso
- ❑ **Ejemplo interactivo**
- ❑ ¿Cómo cambian las bandas al mover uno de los valores hacia adentro? ¿Y hacia afuera?
- ❑ ¿Qué ocurriría si tuviésemos un valor muy atípico?
- ❑ ¿Puede ocurrir que todos los valores estén fuera de las bandas?

# Ejemplo: Control de Calidad



Virgilio Gómez-Rubio, Created with [GeoGebra](#)

# MUCHAS GRACIAS