

**¿Cuántos peces hay en un lago?  
¿Cuántos taxis en una ciudad?**

***Ideas para actividades y proyectos utilizando la estadística***

---

**Pere Grima  
pere.grima@upc.edu**



**2º Encuentro Iberoamericano de Biometría**

---

## Contenido. Objetivo

**Peces** (estimación del tamaño de una población por el método de la captura, marcaje y recaptura)

**Taxis** (estimación del tamaño de una población con sus elementos numerados)

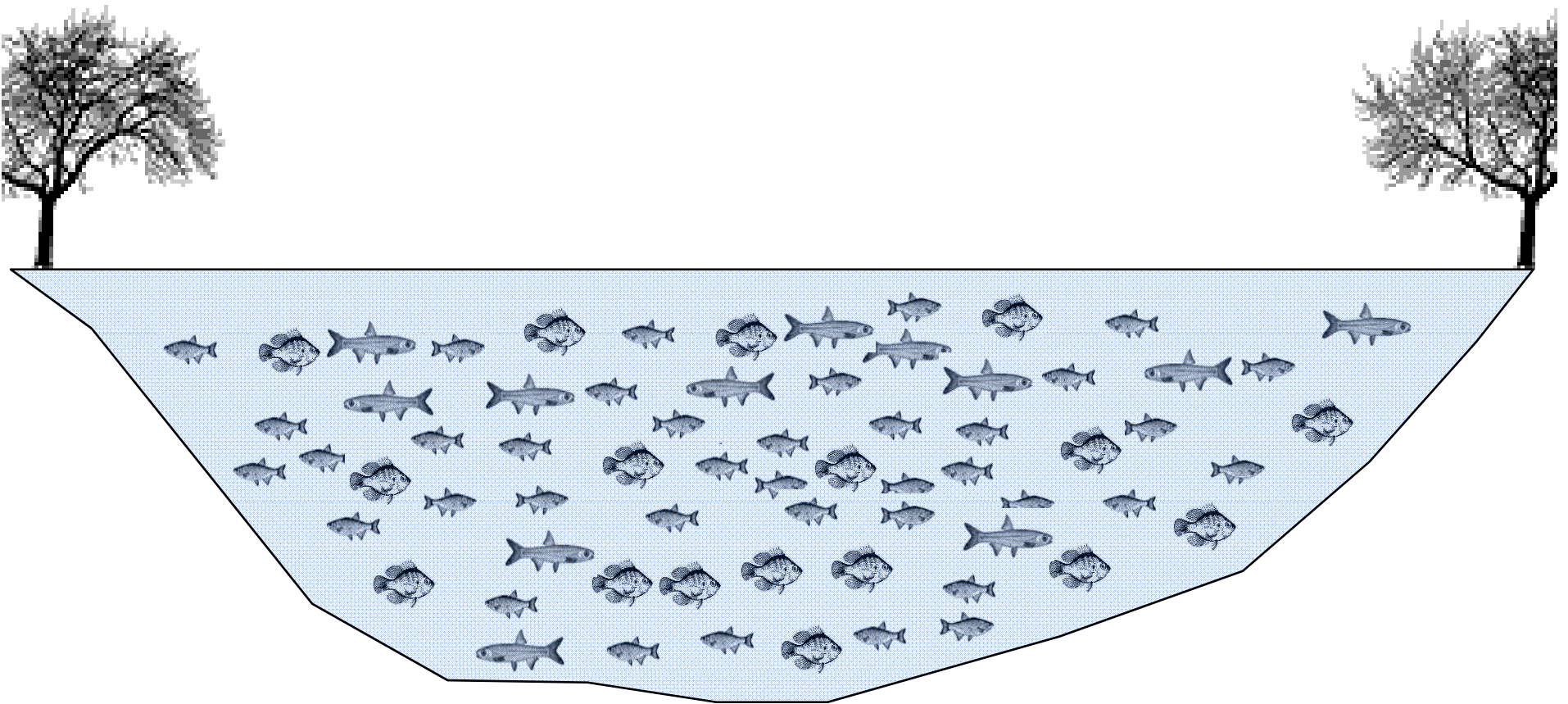
**“Estadística en acción”** (más ejemplos)

Objetivo: **Dar ideas sobre posibles actividades o trabajos relacionados con la estadística**

## El lago

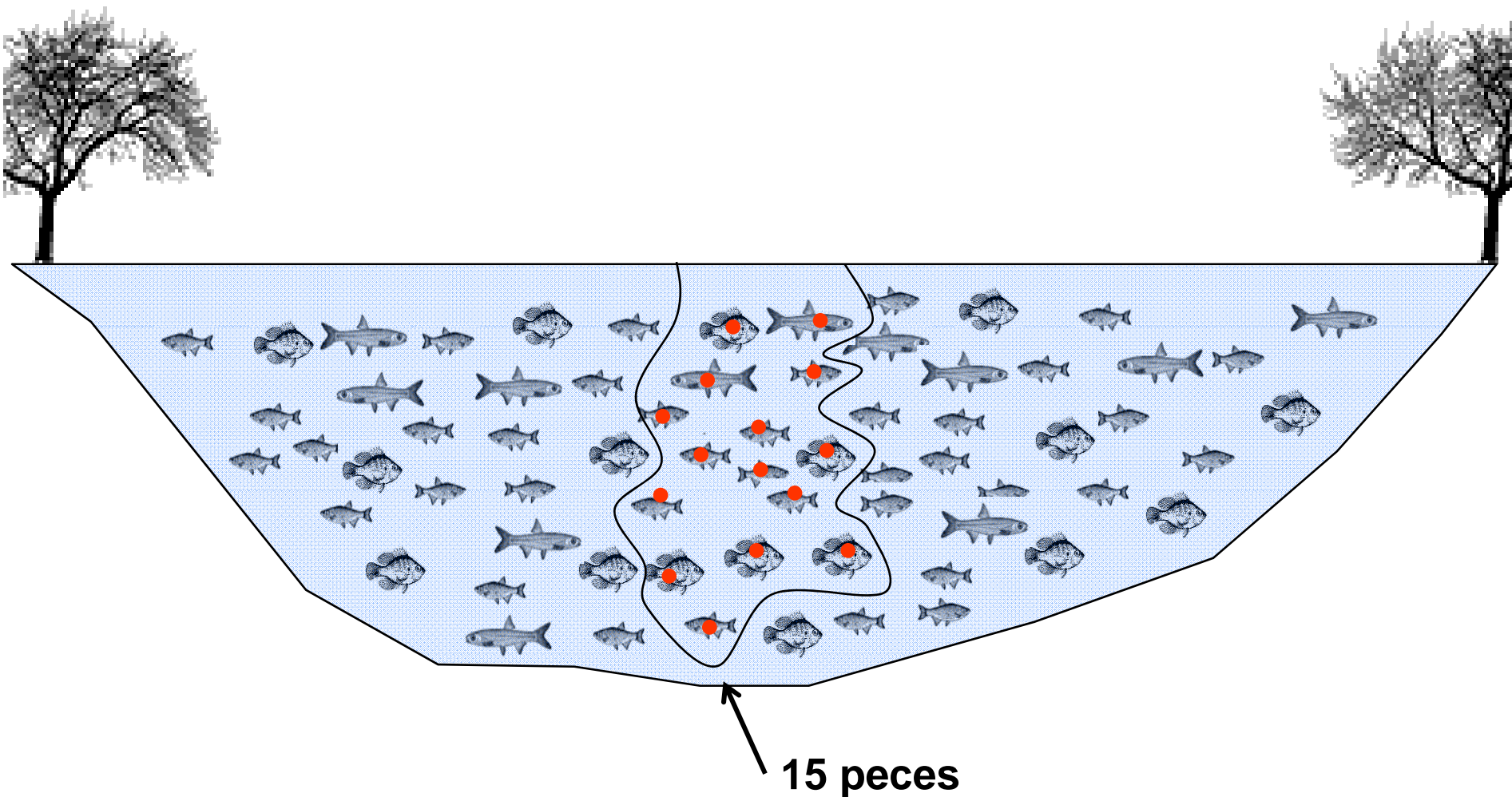


# Si pudiéramos ver lo que hay dentro...



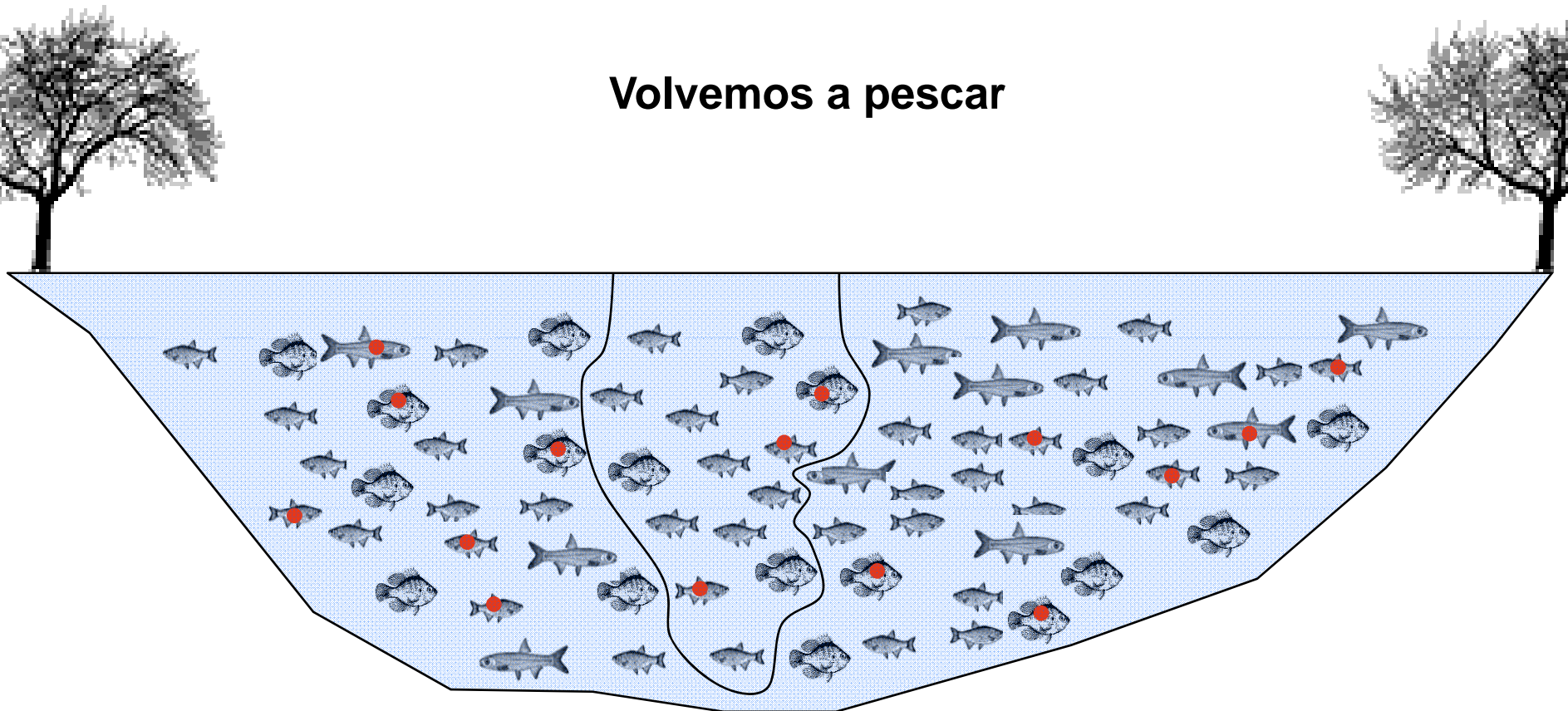


# Pescamos, marcamos y los soltamos



# Los peces marcados se dispersan...

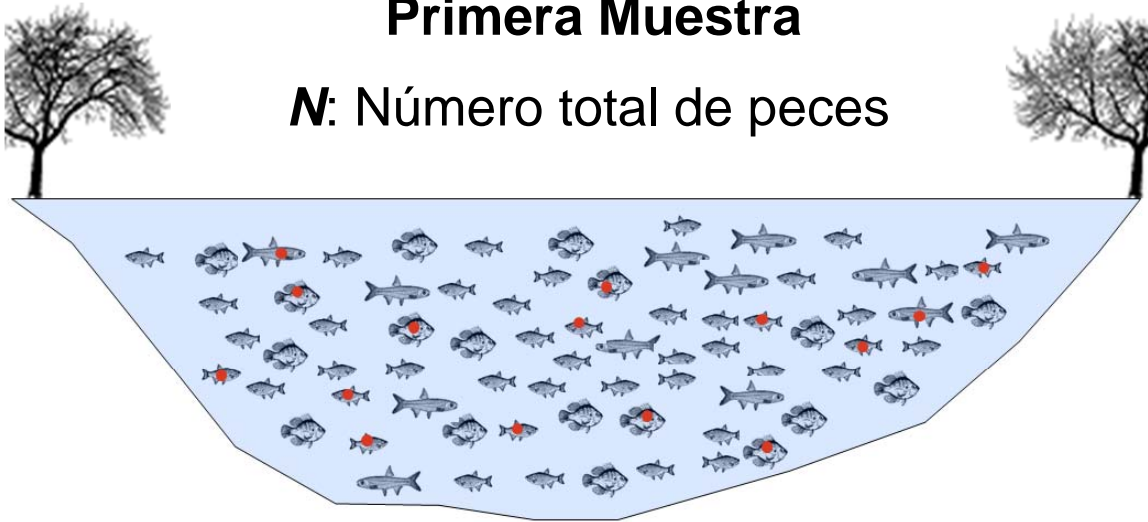
Volvemos a pescar



# Hipótesis. Cálculos

## Primera Muestra

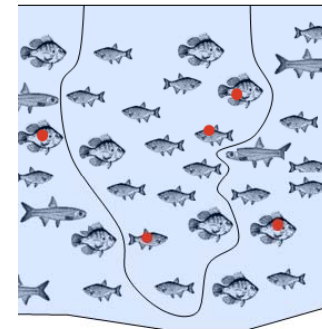
**N:** Número total de peces



**M:** Marcados

## Segunda Muestra

**C:** Capturados



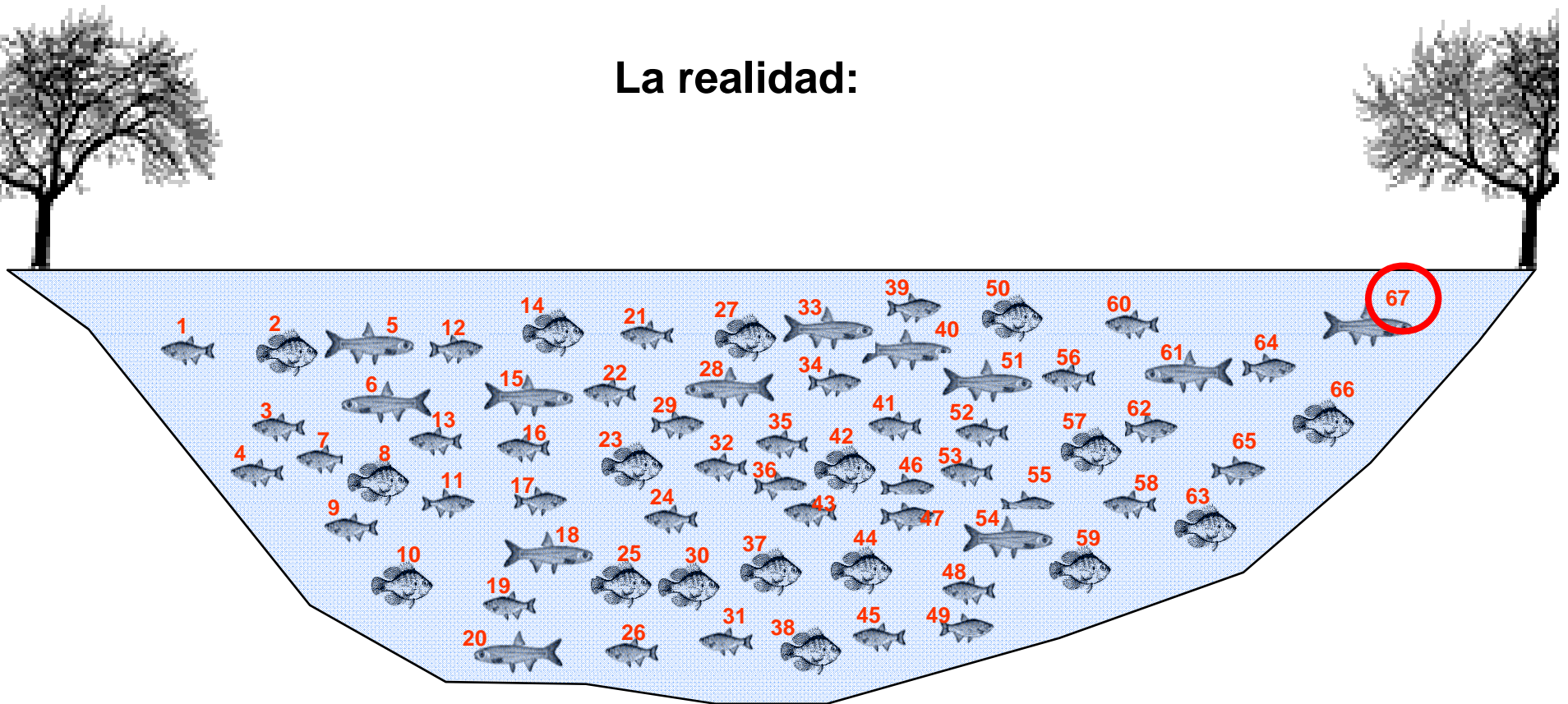
**R:** Aparecen marcados

$$\frac{M}{N} = \frac{R}{C} \quad \longrightarrow \quad \hat{N} = \frac{M \cdot C}{R}$$

$$\hat{N} = \frac{15 \cdot 15}{3} = 75$$

# Si pudiéramos ver lo que hay dentro...

La realidad:



Hay 67 peces. Error del 12% sobre el valor real



## Calidad de la estimación

¿Qué significa que  $N$  está “alrededor de” 75?

¿Tenemos un buen estimador?

Cuál es la calidad de la estimación si:

$$N = 1000$$

$$M = 100$$

$$C = 100$$

# Simulación con Excel

Podemos  
simular  
con Excel



Simulacion Excel [Modo de compatibilidad] - Micro

Inicio Insertar Diseño de página Fórmulas Datos Revisar Vista Programador Complementos

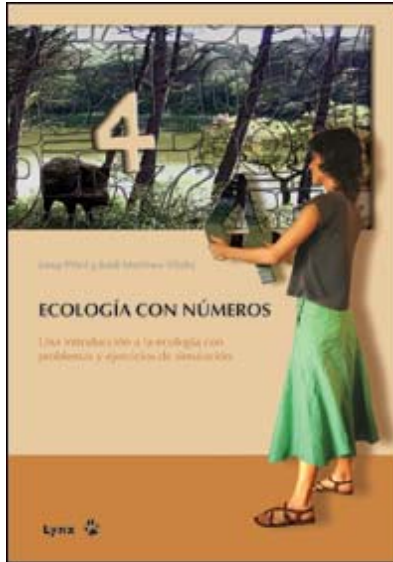
Calibri 11 A A

Pegar Fuente Alineación Ajustar texto General Combinar y centrar Número Forma condicio

D2 =SUMA(C2:C101)

	A	B	C	D	E
1	Población con la primera muestra marcada	Números aleatorios (columna auxiliar)	Muestra repescada	Peces que se encuentran marcados en la repesca	Estimación del tamaño de la población
2	0	0,747579	0	11	909,0909091
3	0	0,923519	0		
4	0	0,831943	0		
5	0	0,045243	0		
6	0	0,872980	0		
7	0	0,690584	0		
8	0	0,541898	0		
9	0	0,311406	0		
10	0	0,962678	0		
11	0	0,560683	0		
12	1	0,642250	1		
13	1	0,405491	1		
14	0	0,312869	0		
15	0	0,419439	0		
16	0	0,506876	0		
17	1	0,033154	1		
18	0	0,342271	0		
19	0	0,939727	0		
20	0	0,330959	0		
21	0	0,027463	0		

# Applet



3.1. Marcaje y recaptura - Mozilla Firefox

Archivo Editar Ver Historial Marcadores Herramientas Ayuda

[http://www.ecologiaconnumeros.uab.es/Libre/AppletsLibre/3-MuestreoPoblaciones/Applet\\_3\\_1/MarcajeRecaptura.html](http://www.ecologiaconnumeros.uab.es/Libre/AppletsLibre/3-MuestreoPoblaciones/Applet_3_1/MarcajeRecaptura.html)

MARCAJE Y RECAPTURA (2)

Marcar Recapturar Calcular Nuevo N real= 1000 Ver N M= 300 C= 100 R= N est=

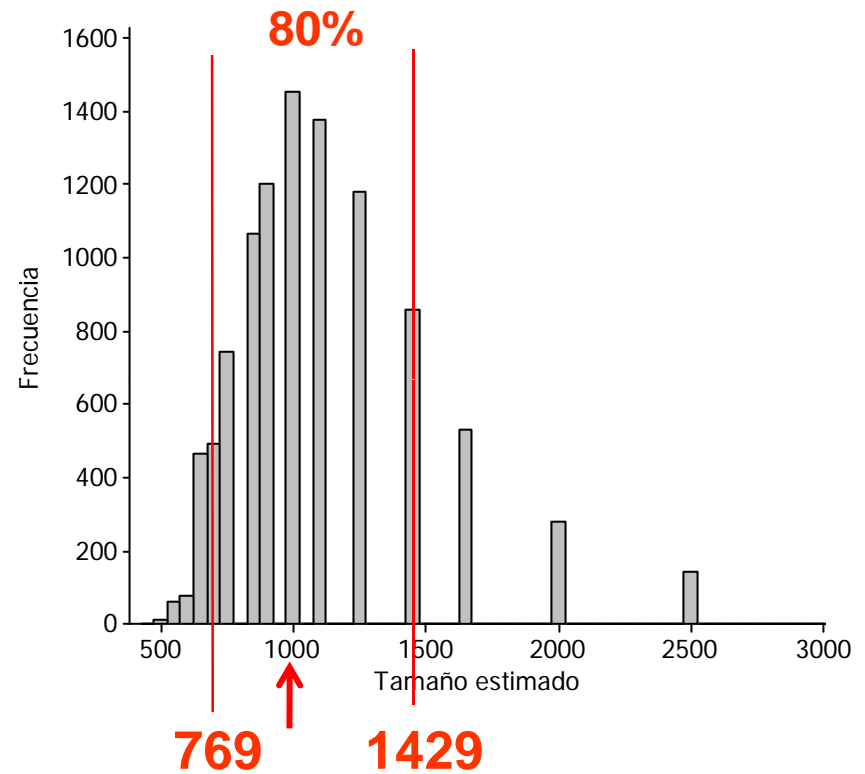
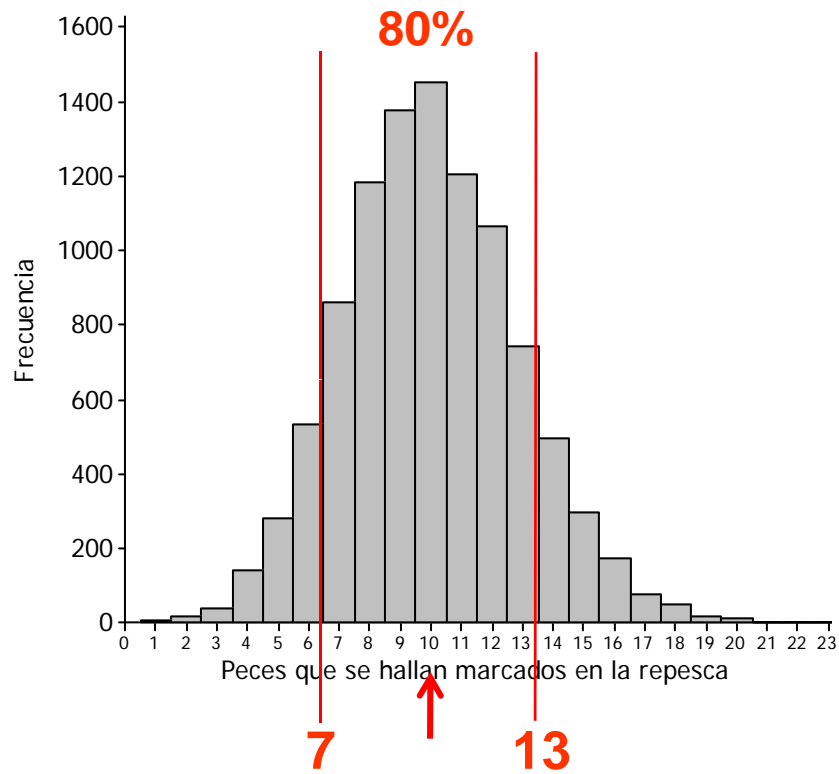
674 985 1446

Auto k = 100 int = 95 N medio = 985 N inf = 722 N sup = 1320





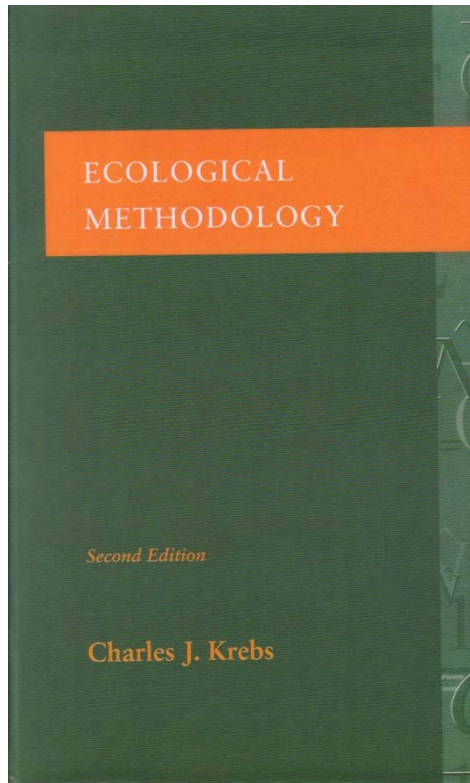
# Haciendo gráficos



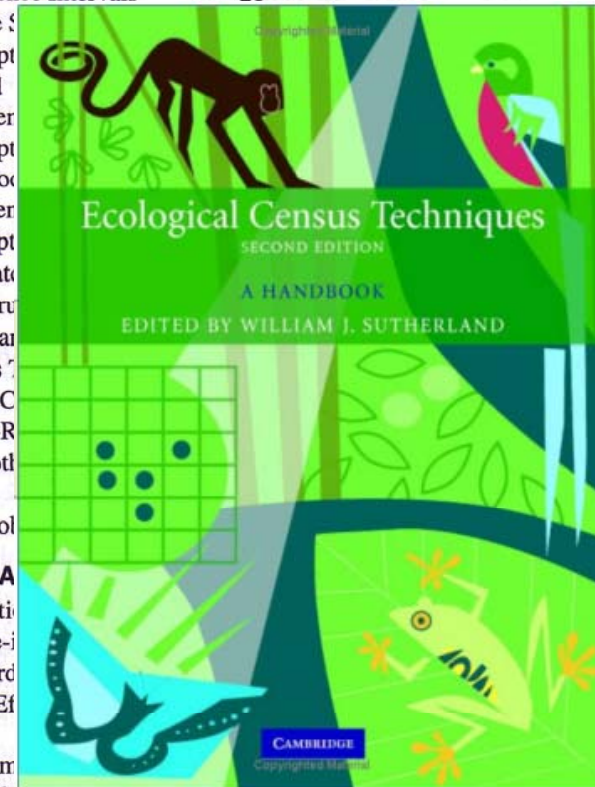
$$\hat{N} = \frac{M \cdot C}{R} \quad \text{Es un estimador sesgado}$$

# ¿Se puede mejorar?

¿Se puede calcular u  
 ¿Cómo varía la calid  
 ¿Y si se hacen varias  
 ¿Y si...?



<b>Chapter 1</b>	<b>Ecological Data</b>		<b>1</b>
1.1	Designing Field Studies	3	
1.2	Scales of Measurement	6	
1.3	Statistical Inference	10	
1.4	Data Records	12	
	Questions and Problems	14	
<b>Part One</b>	<b>Estimating Abundance in Animal and Plant Populations</b>		<b>17</b>
<b>Chapter 2</b>	<b>Estimating Abundance: Mark-Recapture Techniques</b>		<b>19</b>
2.1	Petersen Method	20	
	2.1.1 Confidence Intervals	21	
	2.1.2 Sample Size		
	2.1.3 Assumptions		
2.2	Schnabel Method		
	2.2.1 Confidence Intervals		
	2.2.2 Assumptions		
2.3	Jolly-Seber Method		
	2.3.1 Confidence Intervals		
	2.3.2 Assumptions		
2.4	Tests of Equal Catch		
	2.4.1 Zero-Truncated Poisson		
	2.4.2 Chapman's		
	2.4.3 Leslie's		
	2.4.4 Leslie's		
2.5	Planning a Mark-Recapture Study		
2.6	What to Do If Not All Marked Animals are Recaptured		
2.7	Summary		
	Questions and Problems		
<b>Chapter 3</b>	<b>Estimating Abundance of Exploited Populations</b>		<b>70</b>
3.1	Exploited Populations		
	3.1.1 Change-in-Ratio		
	3.1.2 Eberhardt's		
	3.1.3 Catch-Effort		
3.2	Resight Methods		
3.3	Computer Programs		
3.4	Enumeration Methods		



## Con bolas....





## ¿Cuántos garbanzos hay en 1 kg?



- Contarlos
- Pesar 100 y hacer la proporción
- Pesca i repesca





# ¿Cuántos taxis hay en Barcelona?

---



# ¡Fácil!

<http://www.taxibarcelona.cat/tabid/308/Default.aspx>

QUI SOM DIRECTORI AGENDA EL VOSTRE CORREU SMS CONTACTE TAXI SOSTENIBLE

Treballèm per la qualitat del servei

► TAXISTA ► USUARI ► PER A SER TAXISTA

Inici : Sala de Premsa : El món del taxi : Dades sectorials TORNEU ►

## Dades del sector de taxi de l'AMB

Nre. de llicències	10.481 Nocturnes: 285 Doble torn: 137
Nre. de credencials de taxistes	16.878 (11.235 actives)
Carreres diàries	més de 225.000
Vehicles autoritzats per a taxi	13 marques / 48 models
Antiguitat	5 anys: 67,5% Entre 5 y 7 anys: 19,6% Entre 7 y 10 anys: 9,5% + 10 anys: 3,4%
Mitjana d'edat del parc de vehicles de taxi a l'AMB	4 anys



## Sin repesca....



**¡Los taxis están numerados!**

## Sólo con la intuición...

**Muestra de valores de una población numerada:**

**16 28 45 48 68 72 81**

**¿Cuántos elementos tiene la población?**



# Posibilidades

Si tenemos toda la población numerada se verifica que:

$$\left. \begin{array}{l} \text{Tamaño de la población: } N \\ \text{Media: } \bar{X} \\ \text{Mediana: } \tilde{X} \end{array} \right\} \begin{array}{l} N = 2\bar{X} - 1 \\ N = 2\tilde{X} - 1 \end{array}$$

**Ejemplo:**

1, 2, 3, 4, 5, 6, 7, 8, 9, 10



**Media = Mediana = 5,5**

## Posibilidades

En nuestra muestra:

16 28 45 48 68 72 81

Mediana: 48

Media: 51,14

Estimaciones:  $\hat{N}_1 : 2\bar{X} - 1 = 2 \cdot 51,14 - 1 \cong 101$

$\hat{N}_2 : 2\tilde{X} - 1 = 2 \cdot 48 - 1 \cong 95$

¿Inconveniente de este método?

## El inconveniente

Puede dar una estimación evidentemente falsa...

Ejemplo: 3, 4, 6, 15

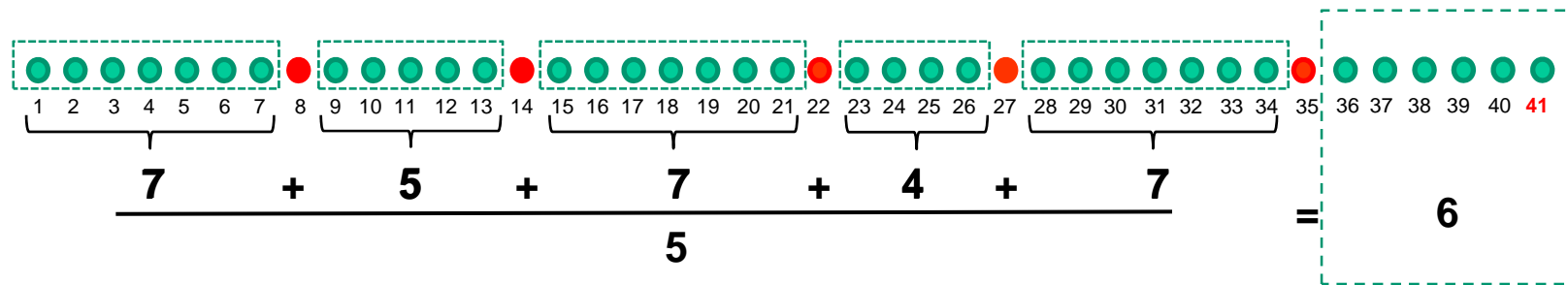
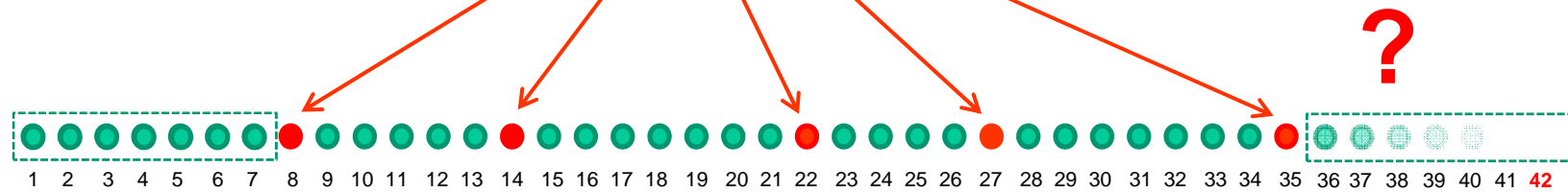
Media: 7       $\hat{N}_1 : 2\bar{X} - 1 = 2 \cdot 7 - 1 = 13$

Mediana: 5       $\hat{N}_2 : 2\tilde{X} - 1 = 2 \cdot 5 - 1 \cong 9$

# Solucionando el inconveniente

Supongamos que los valores son:

8, 14, 22, 27, 35



¡Este es un estimador excelente! (UMVUE)



## ¡ La fórmula !

**El promedio de las diferencias es:**

$$\frac{\{[X_{(1)} - 1] + [X_{(2)} - X_{(1)} - 1] + [X_{(3)} - X_{(2)} - 1] + \dots + [X_{(n)} - X_{(n-1)} - 1]\}}{n} = \left[ \frac{X_{(n)}}{n} \right] - 1$$

$X_{(1)}$  : Primer valor, ordenados de menor a mayor

$X_{(n)}$  : Último valor, ordenados de menor a mayor

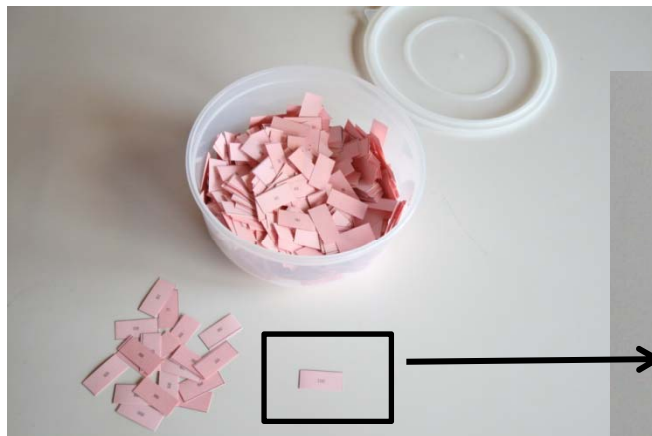
Por tanto:

$$\hat{N} = X_{(n)} + \frac{X_{(n)}}{n} - 1$$

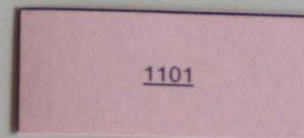
## Comprobemos que funciona



20 veces



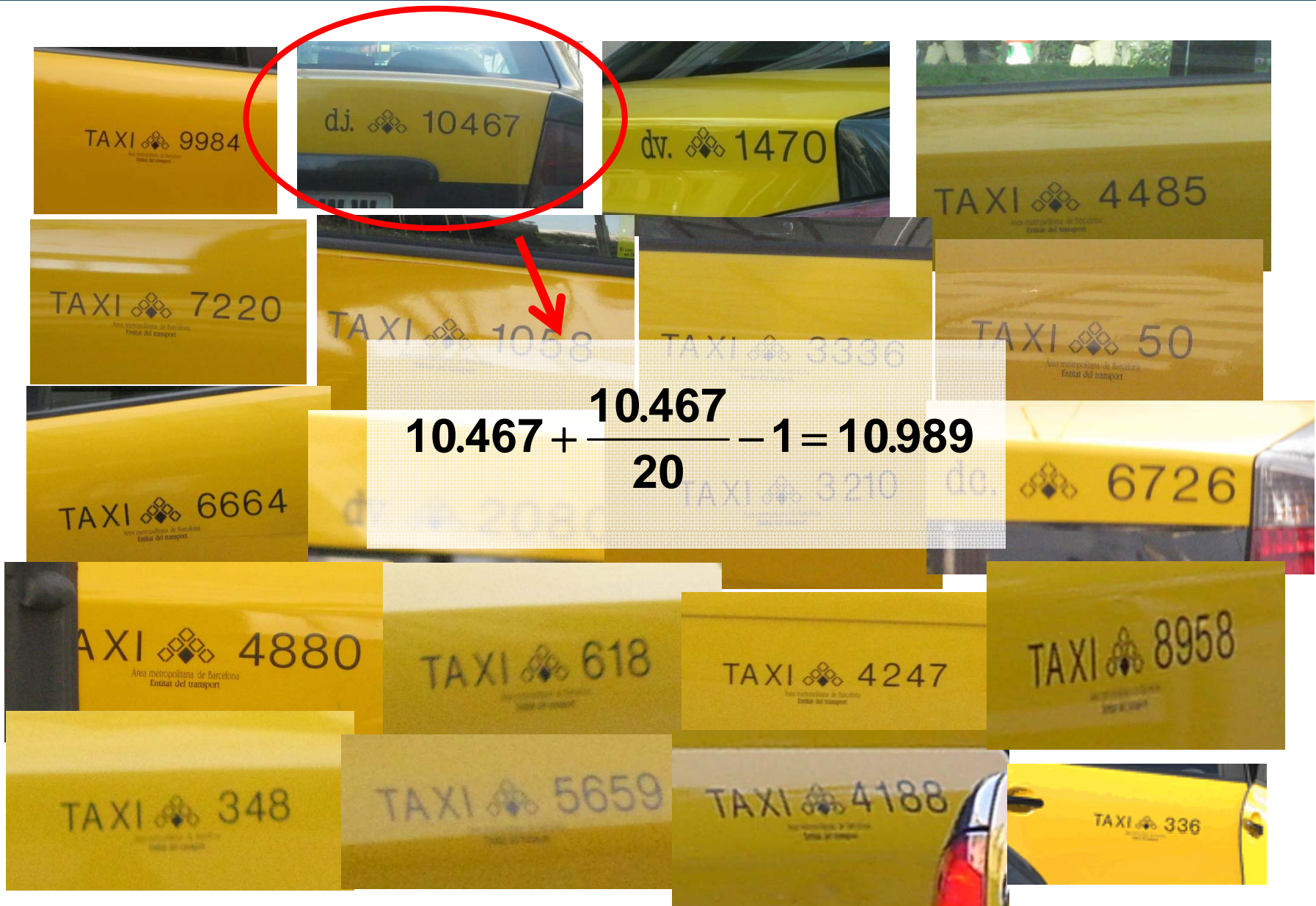
El mayor valor :



Mejor estimación:

$$1101 + \frac{1101}{20} - 1 = 1155$$

# Caso práctico





## Otras situaciones similares

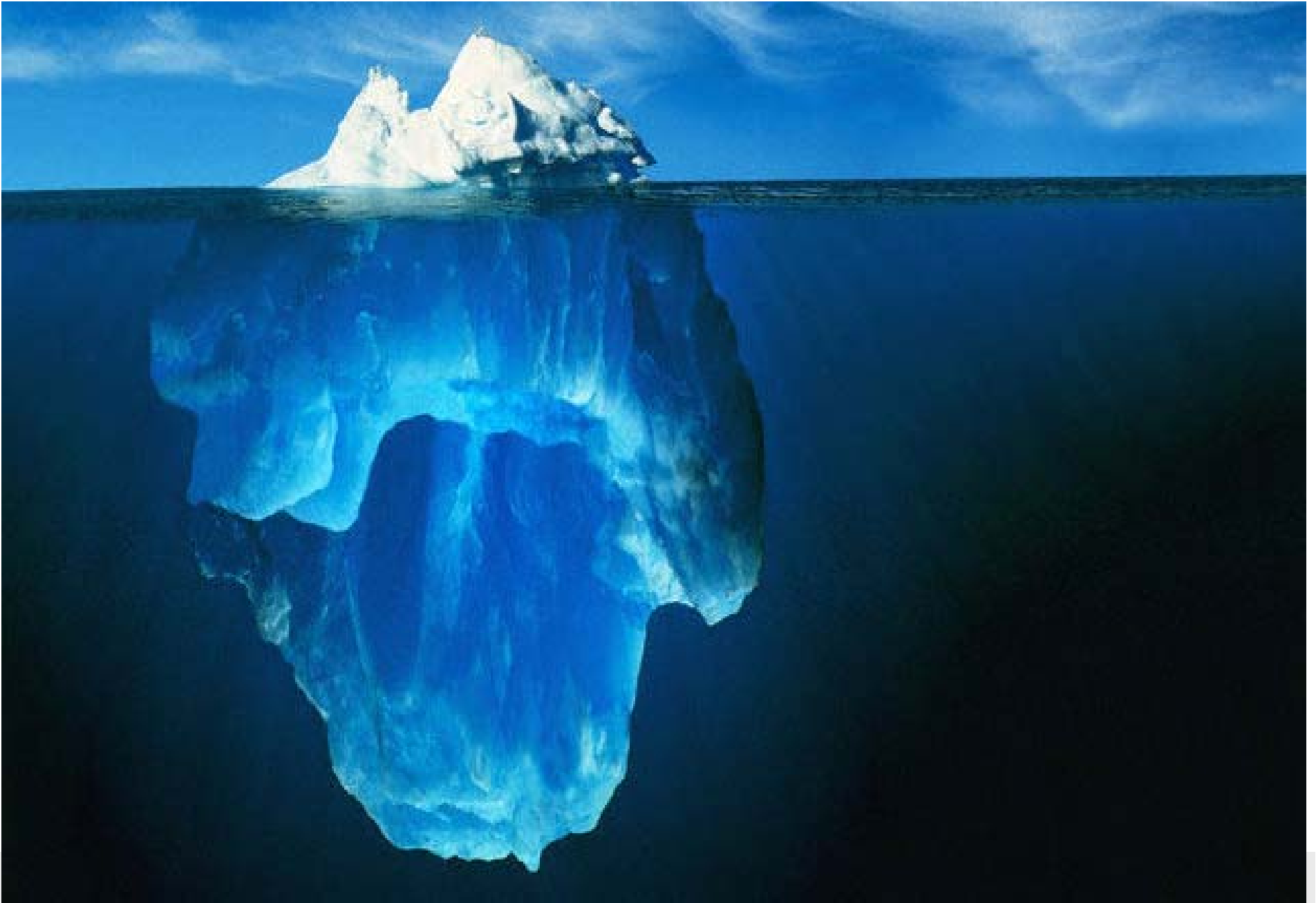


We close with an anecdote of Colonel Trevor Dupuy (1991). “In the Middle East a few years ago I was given permission by

Israeli military authorities to go through the entire Merkava Tank production line. At one time I asked how many Merkavas had been produced, and I was told that this information was classified. I found it amusing, because there was a serial number on each tank chassis.”

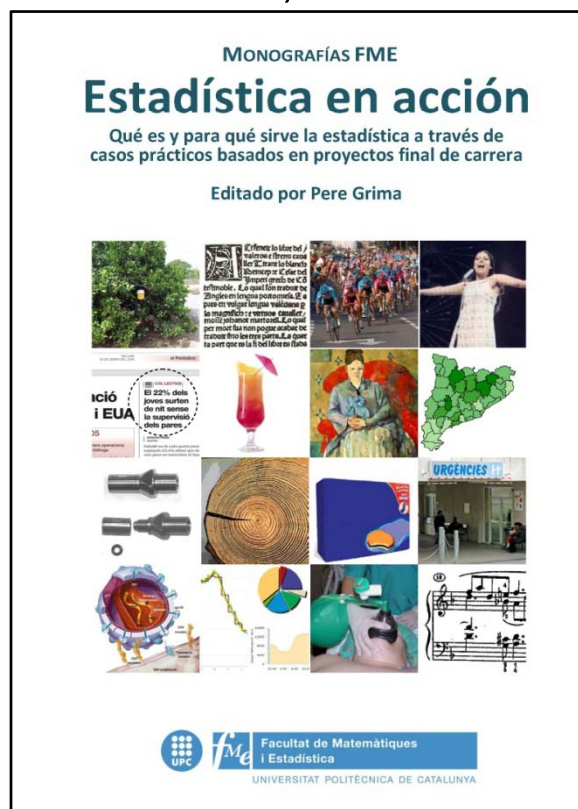
*Roger W. Johnson: “Estimating the Size of a Population. Teaching Statistics <http://www.rsscse.org.uk/ts/gtb/contents.html>*

## Iceberg y estadística



# Hay mucho más de lo que se ve

Porcentajes  
Sondeos electorales  
Medias





# Análisis estadístico del estilo literario. Discusión sobre la autoría del *Tirant lo Blanc*

Proyecto realizado por: **Susanna Cabos Ruiz**  
Dirigido por: **Josep Ginebra Molins**

**Cuantificación del estilo literario:** longitud de las frases, frecuencia de uso de determinadas palabras (palabras herramienta), selección dntre posibles alternativas,...

**Autor  
A**

**Autor  
B**

**Autor  
disputado  
A o B**

# Estilometría. Software

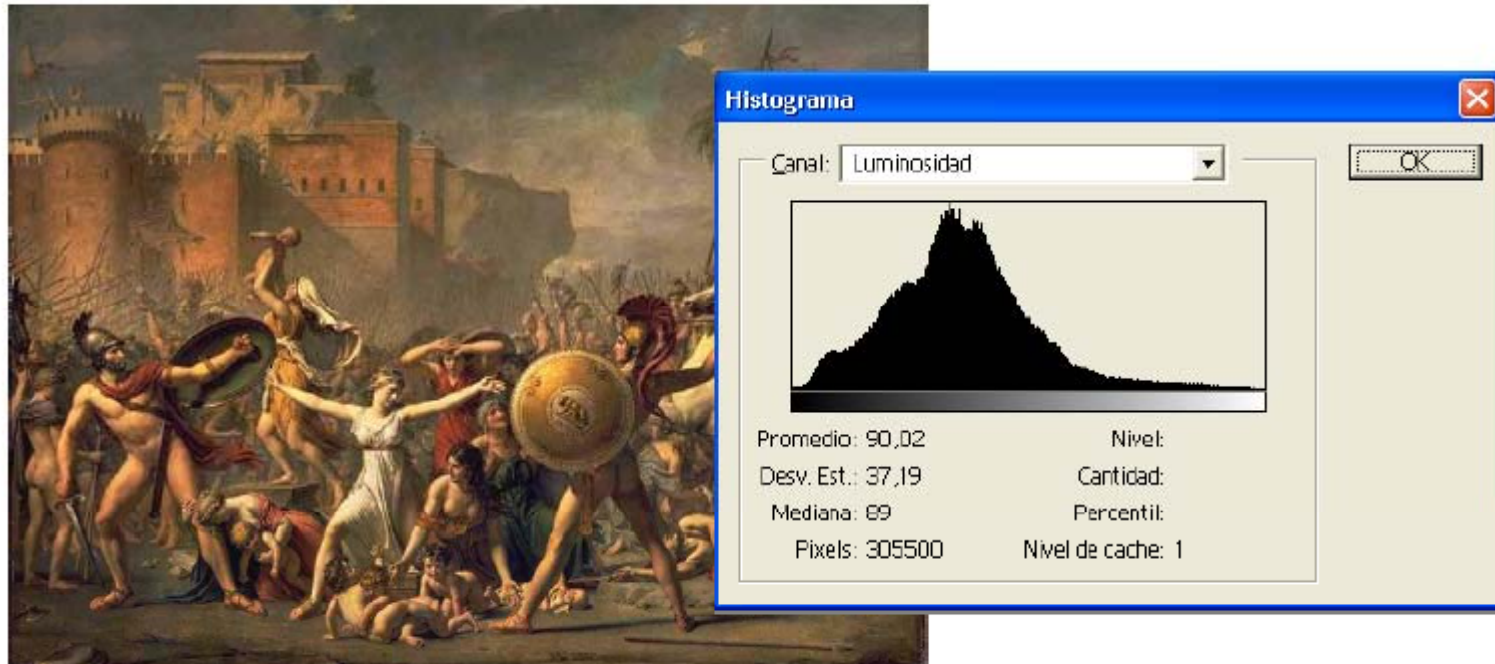


Ejercicios



# Aplicación de técnicas estadísticas al estudio del arte pictórico de los siglos XV al XIX

Proyecto realizado por: **Miquel Romero Obón**  
Dirigido por: **Lidia Montero Mercadé**



# Cuantificación de una imagen

**Histogram**

Display channel: Blue

Value(s):  
% in Range:  
% Below:  
% Above:  
Mean: 147  
Median: 163

Red  Green  Blue  Greyscale  Use all layers  
 Hue  Saturation  Lightness  Selection only

**Image Information**

Item	Value
Orientation	Landscape (Top Left)
X resolution	72.0 dpi
Y resolution	72.0 dpi
Resolution unit	Inches (in)
Pixel height	2532
Pixel width	3888
Component configur...	YCbCr
Color space	sRGB
Exposure program	Manual
Scene capture type	Standard
Exposure mode	Manual exposure
Exposure bias	F/1.0 (0.00)
Exposure time	1/30 second
F number	5.0
Lens aperture	F/5.0 (4.64)
Shutter speed	1/30 sec. (4.91)
Focal length	37.0 mm

**Paint Shop Pro X**

The number of unique colors used in this image is 148164.

Aceptar

**Image Information**

Orientation: Landscape (Top Left)

Adjust date taken

Editable

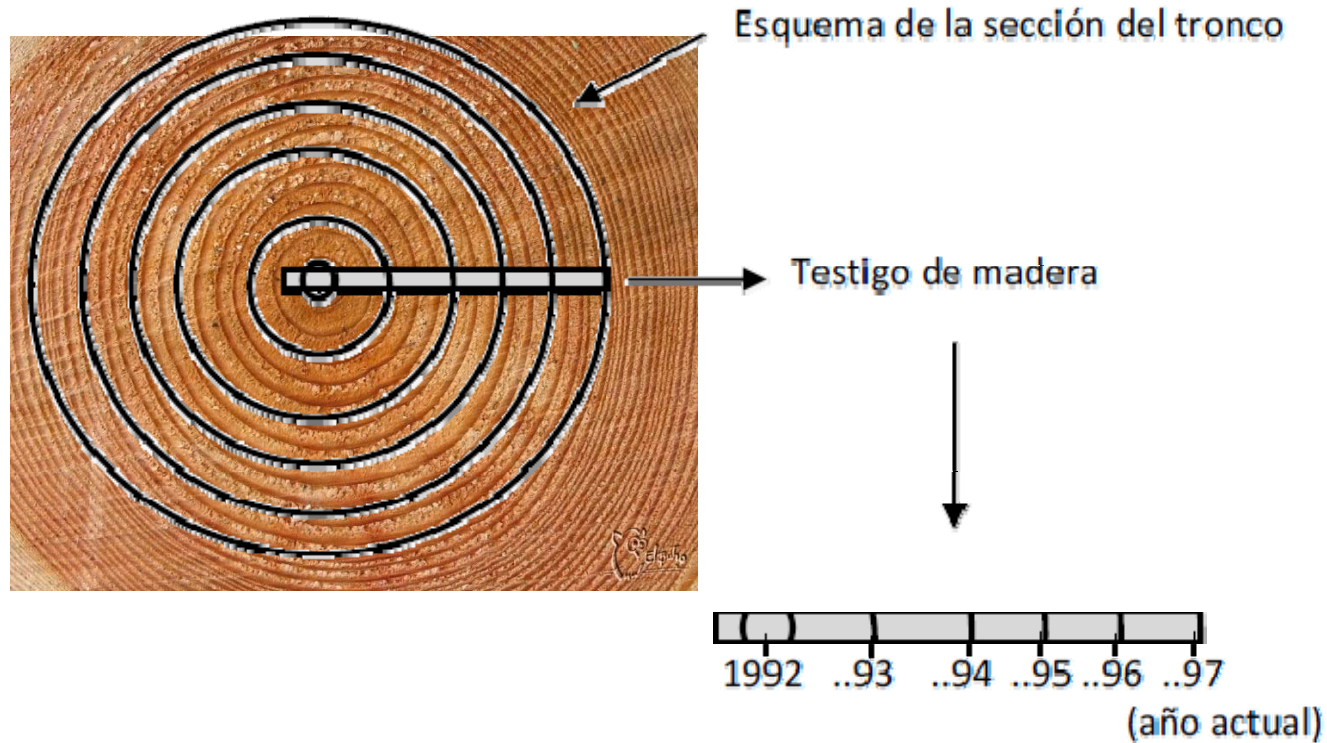
Aceptar Cancelar Ayuda

Image: 3888 x 2592 x RGB - 8 bits/channel



# Influencia de las condiciones climáticas en el crecimiento del pino silvestre en Cataluña

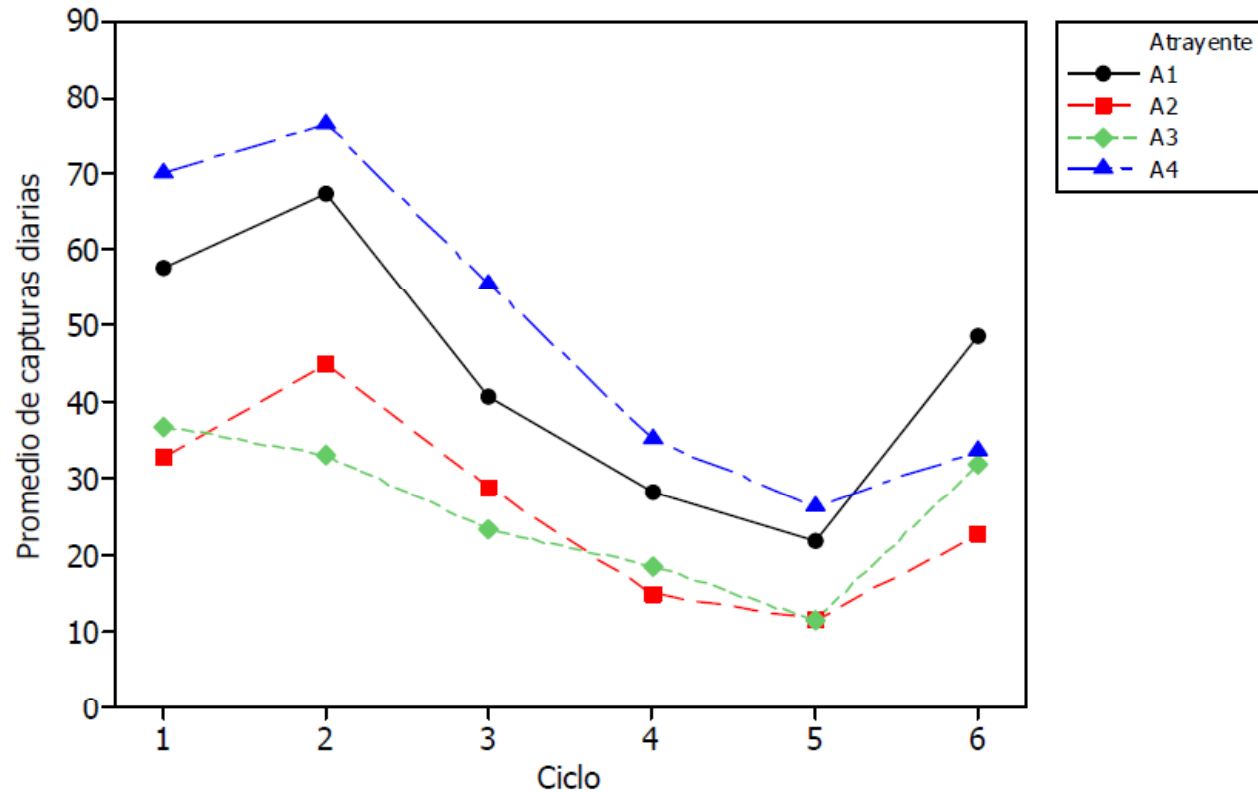
Proyecto realizado por: **Natàlia Adell Calvet**  
Dirigido por: **Llorenç Badiella Busquets**  
Tutelado per: **Josep Anton Sánchez Espigares**





# Buscando productos más limpios y eficaces para luchar contra las plagas de los frutales

Proyecto realizado por: **Lourdes Roderó de Lamo**  
Dirigido por: **Josep Ginebra i Molins**



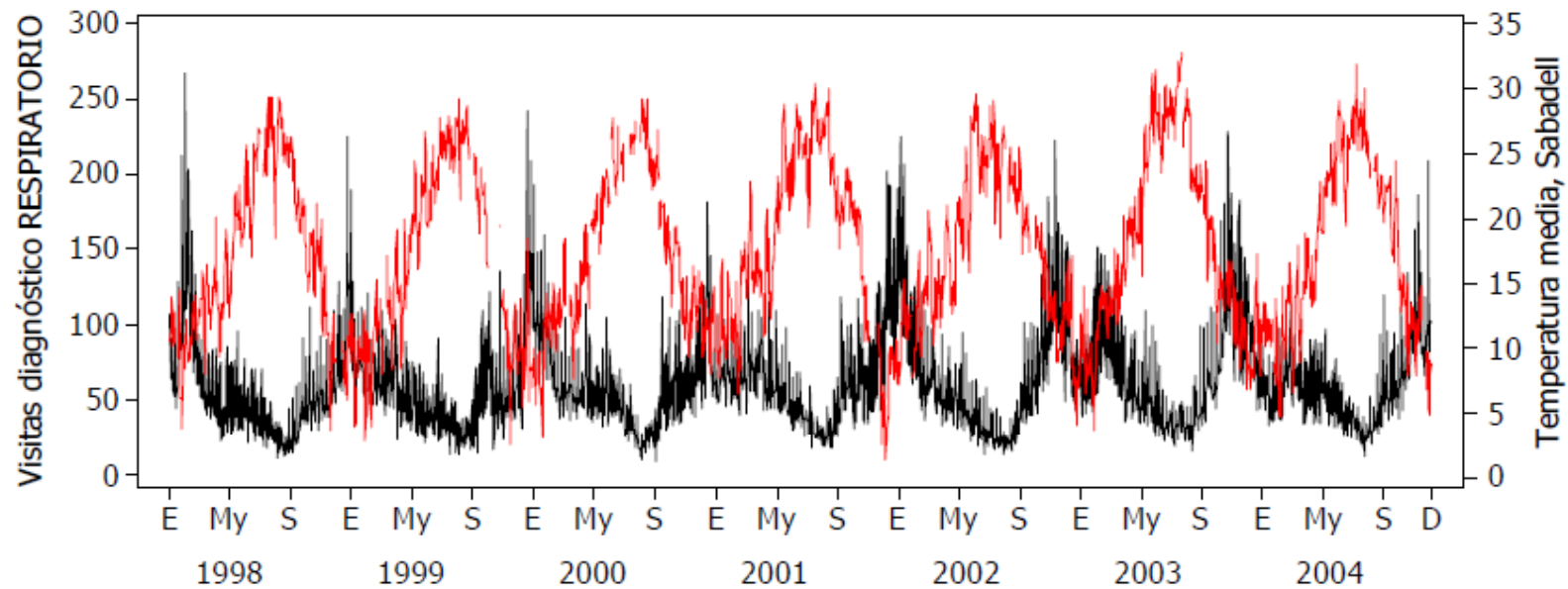




# Comportamiento de la demanda de urgencias hospitalarias y factores meteorológicos asociados

Proyecto realizado por: **Jordi Real Gatius**

Dirigido por: **Josep Anton Sánchez Espigares y Aureli Tobías Garcés**

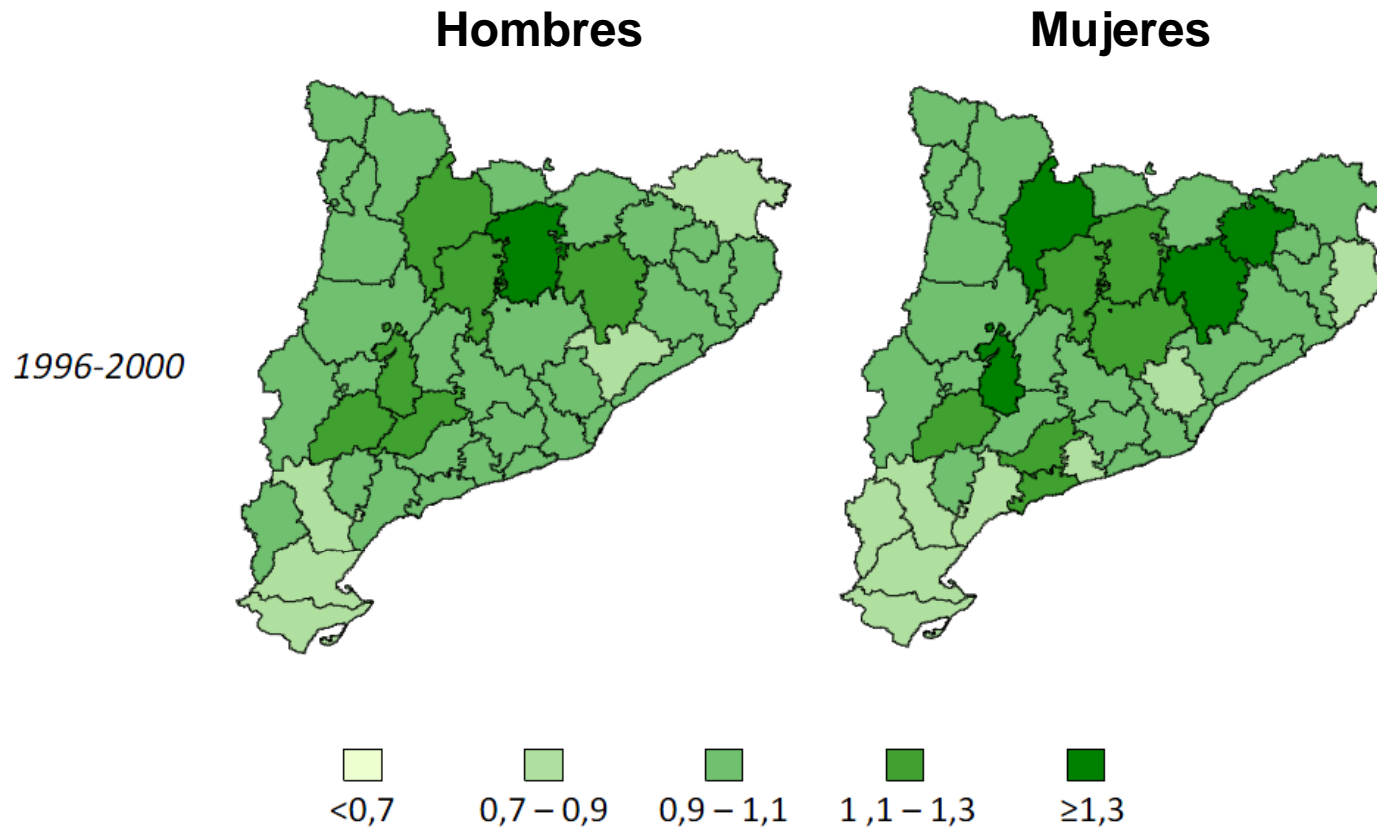




# Análisis de la mortalidad por tumores malignos de mama y estómago en Cataluña

Proyecto realizado por: **Xavier Puig Oriol**

Dirigido por: **Josep Ginebra i Molins**

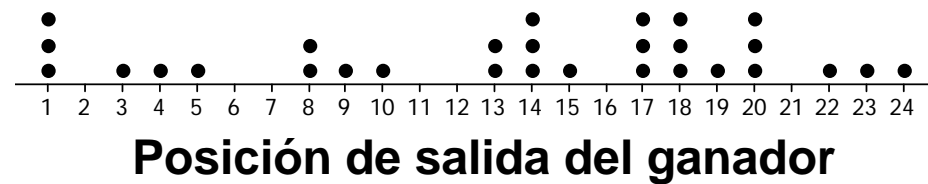
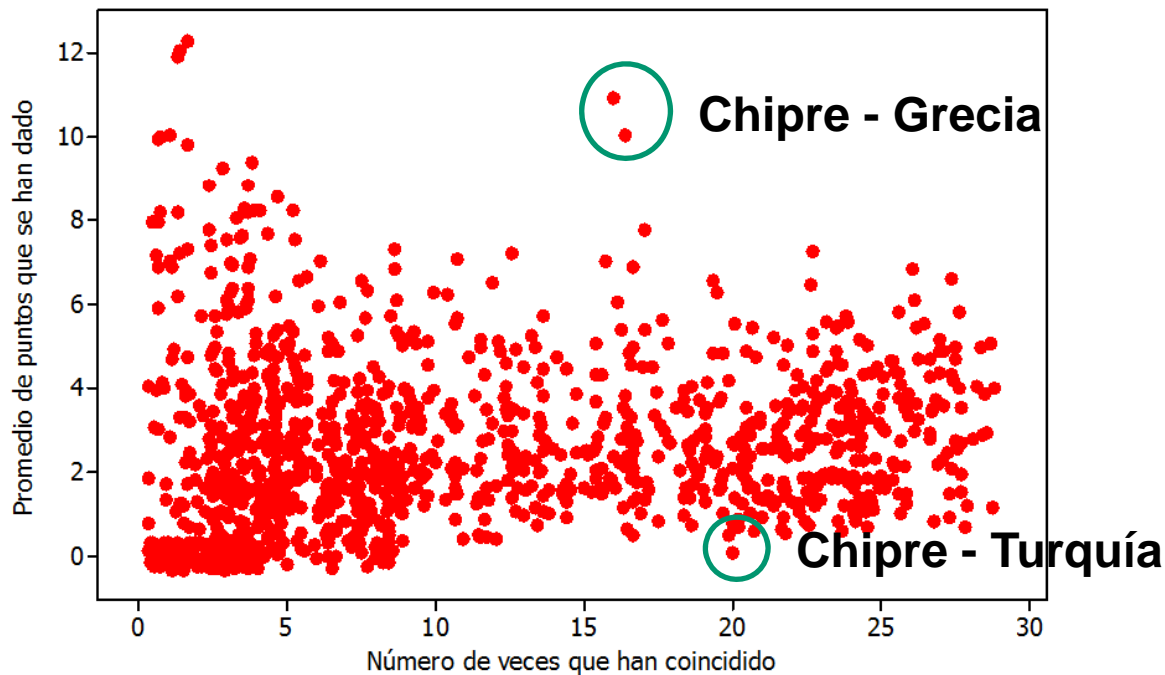




# Análisis de patrones y tendencias en las votaciones del festival de Eurovisión

Proyecto realizado por: **Laura Marí Tomàs**

Dirigido por: **Lluís Marco Almagro**



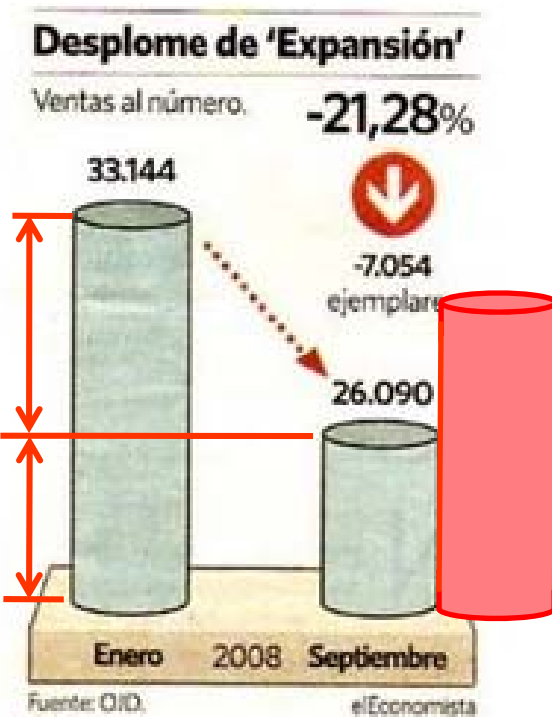
# La estadística en la prensa. Estudio crítico

Proyecto realizado por: Sara Fontdecaba i Rigat  
 Maria Montón i Domingo

Dirigido por: Pere Grima Cintas

27.10.08

## Consejos vendo y para mí no tengo



## 'Expansión' se contrae

Si, siempre es cierto aquello de renovarse o morir, en tiempos de crisis el dicho cobra aún más sentido. El tradicional periódico económico *Expansión*, anclado en un estilo abigarrado, plano y falto de rigor, resiste mal la incertidumbre. Según los datos de OJD,

Fuente: [www.malaprensa.com](http://www.malaprensa.com)

## 'Expansión' se contrae

## **Bonus track**

(más información sobre peces, taxis, bolas, palas para sacarlas... )

**<http://www-eio.upc.es/~grima/>**