

# DADOS Y DATOS II

CÓMIC **DISCRETO** DE ESTADÍSTICA  
PARA UN APRENDIZAJE **CONTINUO**





# Dados y datos II

Cómic **discreto** de estadística  
para un aprendizaje **continuo**



**Govern de les Illes Balears**  
**Conselleria d'Economia, Hisenda i Innovació**  
Direcció General d'Economia

© Institut Balear d'Estadística (IBAE)  
C/ Sant Gaietà, 4, 1r  
07012 Palma (Mallorca)  
Tel. (34) 971 177 489  
Fax (34) 971 176 467  
<http://ibae.caib.es>  
e-mail: [ibae@caib.es](mailto:ibae@caib.es)

Edición: **Direcció General d'Economia**  
**Conselleria d'Economia, Hisenda i Innovació**  
**Govern de les Illes Balears**

Autor: **Javier Cubero**

Dirección del proyecto: **Maria Marquès Caldentey**

Dirección técnica: **Miquel Font Rosselló**

Gestión y producción: **inrevés SLL**   
Ilustraciones: **Alex Fito**  
Color y maquetación: **Samuel García Martorell**  
Coordinación y guión adaptado: **Pere Joan**

Colección: **Estadística al carrer. Volumen 2**  
Título: **Dados y datos II. Còmic discreto de estadística para un aprendizaje continuo**  
Nº IBAE: **II-MMV**  
Depósito legal: **PM 1.223-2005**  
ISBN: **84-934294-2-2**

Impresión: **Imprenta Son Espanyolet**  
Fecha de edición: **mayo 2005**

Me complace compartir contigo esta nueva publicación del IBAE que ahora tienes en tus manos: DAUS I DADES II.

Tratamos con ella de profundizar en el estudio de los conceptos estadísticos sin perder el atractivo y formato original con el que fue creada y su intención de acercar la estadística a la sociedad y, particularmente, a los estudiantes (ESO y BACHILLER) dentro de los planes educativos vigentes.

Tras la aparente simplicidad de su presentación en forma de cómic y la plasticidad e ingenio en que se resuelven conceptos con cierto grado de dificultad, subyace el más exquisito rigor científico.

Pero es que, además, el desarrollo de cada concepto se aborda desde la perspectiva de unos personajes que en este segundo volumen adquieren una personalidad muy definida. Cada concepto aparece en su momento y encuentra su encaje perfecto en el plan de la obra. Es su gran mérito didáctico.

La continuidad en el estilo y en los personajes, cada uno personificando actuaciones estadísticas, desde "Gráfica" hasta "55" -como representante del carácter de los datos, que hay que entenderlos, leerlos y tratarlos de forma que nos aporten conclusiones- pasando por "Binomio" -el rechazo o no de la hipótesis nula separados solamente por un valor crítico- entre otros.

DAUS I DADES II, como su sobretítulo indica, es una publicación de apoyo a los textos de clase o una creación de incógnitas a confirmar con los textos.

El orden, dentro de una forma, mantiene su discreción tratando una continuidad de conocimientos que persiguen el saber, al menos, interpretar los datos, resoluciones e inferencias estadísticas.

Seguimos pensando que este formato, elegido por el autor para transmitir conocimientos, es adecuado y alcanza la finalidad pretendida que se ha marcado el IBAE, desde el inicio de esta nueva etapa, para el acercamiento de la estadística a la sociedad.

Por ello, prologar esta nueva edición es motivo de satisfacción, no tan sólo en calidad de Directora General del área que engloba nuestro Instituto de Estadística, sino también personalmente en la vertiente de formación profesional que en la materia me alcanza.

Quiero expresar mi agradecimiento al autor por la densidad de los contenidos y al equipo de diseño por la plasticidad y frescura de su realización. Entre ambos han conseguido una obra singular en el panorama editorial, tanto de la estadística como del cómic.

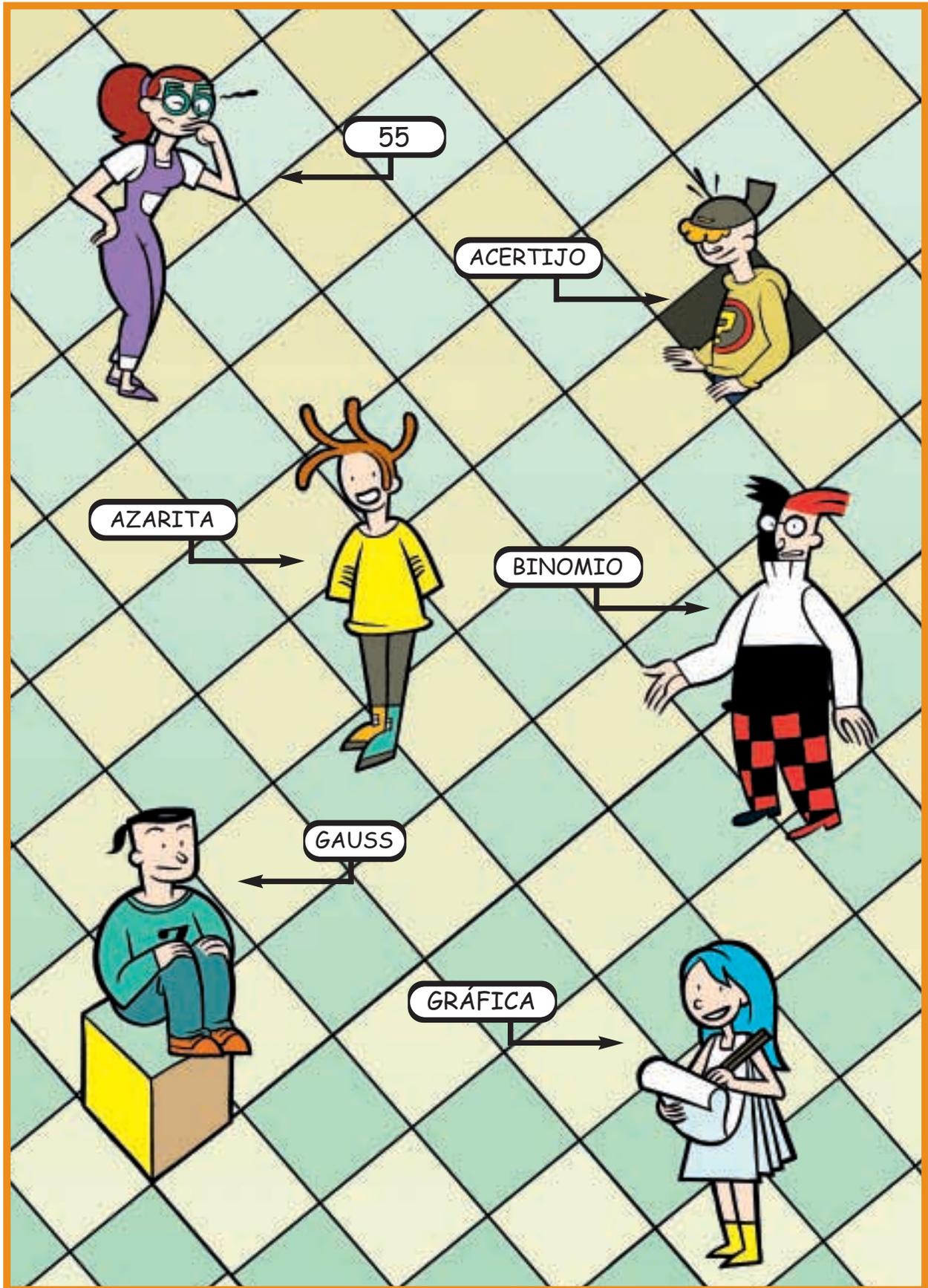
Así también a todos los que han colaborado de alguna forma en esta publicación.

María Marquès Caldentey  
Directora General de Economía



Capítulo 1 - FRANCIS GALTON	pág. 8
Capítulo 2 - KARL PEARSON	pág. 22
Capítulo 3 - RONALD AYLMER FISHER GEORGE SNEDECOR	pág. 35
Capítulo 4 - GERTRUDE MARY COX	pág. 50
Capítulo 5 - ANDREI NIKOLAEVICH KOLMOGOROV	pág. 73
Capítulo 6 - JOHN WILDER TUKEY	pág. 86





# CAPÍTULO 1



FRANCIS GALTON, BIRMINGHAM (1822-1911)



QUIEREN SABER QUÉ OPINAMOS, BASADOS EN NUESTRA PEQUEÑA EXPERIENCIA ESTADÍSTICA.

Institutos	Chicas	Aprobados	Porcentaje
100	573	57	30%
3 Institutos	1000	471	47'10%

TENDRÍAMOS QUE VER SI ESA DIFERENCIA EN PORCENTAJE ES SIGNIFICATIVA.

¿QUÉ?

QUIERE DECIR SI PUEDE SUPONERSE QUE ES CAUSAL O CASUAL, O SEA, POR ALGUNA CAUSA O POR ALEATORIEDAD.

ES COMPLICADA LA CUESTIÓN.

EN SERIO HABRÍA QUE ESTUDIAR VARIOS CONCEPTOS, PERO CREO QUE DEBERÍAMOS PROFUNDIZAR EN LOS DATOS EXAMINÁNDOLOS CON UNA AGRUPACIÓN MENOR.

DIVIDÁMONOS POR PAREJAS Y CONSULTEMOS LOS DATOS EN CADA UNO DE LOS INSTITUTOS.

VALE.



**PARA AHORRAR PAPEL Y TINTA**

$1 \times 2 \times 3 = 3!$   
 $1 \times 2 \times 3 \times 4 \times 5 = 5!$   
 .....  
 $1 \times 2 \times 3 \times 4 \times 5 \times \dots \times (n-2) \times (n-1) \times n = n!$

BUEN TRABAJO.

INSTITUTOS			Examinados	Aprobados	Porcentaje
Pearson	Chicos		410	285	69'50%
	Chicas		152	114	75'00%
Wilcoxon	Chicos		98	18	18'36%
	Chicas		352	71	20'17%
Kolmogorov	Chicos		492	270	54'88%
	Chicas		496	286	57'66%

AQUÍ TENEMOS LOS DATOS DESAGREGADOS.

¿QUÉ?

¡EJEM!  
¡EJEM!

O SEA, QUE MIRANDO SÓLO EL TOTAL DE TODOS LOS INSTITUTOS, EL PORCENTAJE DE CHICOS APROBADOS CON RESPECTO A LOS EXAMINADOS ES MAYOR QUE SI MIRAMOS EL DE LAS CHICAS.

SACANDO CONCLUSIONES A LA LIGERA, ES MÁS PROBABLE APROBAR SIENDO CHICO QUE SI SE ES CHICA.

PERO EN CAMBIO EN CADA INSTITUTO POR SEPARADO, Y OCURRE EN LOS TRES, RESULTA TODO LO CONTRARIO.

PUES ES DIFÍCIL ATREVERSE A DECIR CON ESTOS DATOS QUE LA MANIFESTACIÓN ESTÉ JUSTIFICADA.

ESTO ES ALGO RARO, TODOS LOS NÚMEROS CUADRAN, LOS HEMOS COMPROBADO Y VALIDADO, PARECE UNA PARADOJA, ESTOS INSTITUTOS DEBEN DE SER MUY DIFERENTES O...

LO QUE EN ESTE CASO ES SEGURO ES QUE LA MEDIA DE LOS PORCENTAJES DE LAS MUESTRAS, LOS INSTITUTOS, NO COINCIDE CON EL PORCENTAJE DE LA POBLACIÓN ESTUDIADA, LOS TRES INSTITUTOS AGRUPADOS.





UNA SERIE DE PADRES DAN UNA SUBSISTENCIA A SUS HIJOS DE 6€ SEMANALES... ESO ES UNA VARIABLE DETERMINISTA (LA QUE NOSOTROS CONOCIÁMOS).



MEJOR AÚN, OTROS PADRES ABONAN A SUS HIJOS UNA GRATIFICACIÓN SEMANAL DE 2€ POR HORA DE ESTUDIO, MULTIPLICADA POR LA CALIFICACIÓN GLOBAL, DIVIDIDA POR 10, O SEA:

$$\text{Gratificación} = \frac{2 \times \text{Horas de estudio} \times \text{calificación}}{10}$$



ES DETERMINISTA, PORQUE PODEMOS SABER QUÉ CANTIDAD VAN A RECIBIR, SOLAMENTE REALIZANDO LAS OPERACIONES,

PERO HAY OTRO PADRE QUE ADOPTA LA MISMA FÓRMULA, AÑADIENDO UN POCO DE SUSPENSE, ES DECIR:



$$\text{Gratificación} = \frac{2 \times \text{Horas de estudio} \times \text{calificación}}{10}$$



MÁS 5€ SI LANZANDO UNA MONEDA SALE CARA, O MENOS 6€ SI SALE CRUZ.

LA GRATIFICACIÓN QUE ANTES ERA UNA VARIABLE DETERMINISTA, AHORA SE HA CONVERTIDO EN UNA VARIABLE ALEATORIA.





Ejemplo: Horas de estudio; calificación obtenida 5	
SALE "CARA"	SALE "CRUZ"
Gratificación:	Gratificación:
<b>+ 5 €</b>	<b>- 6 €</b>
Probabilidad de conseguirla: 1/2	Probabilidad de conseguirla: 1/2





**PARA AHORRAR PAPEL Y TINTA**

$$1+2+3+4+5+6+7 = \sum_{n=1}^7 n$$

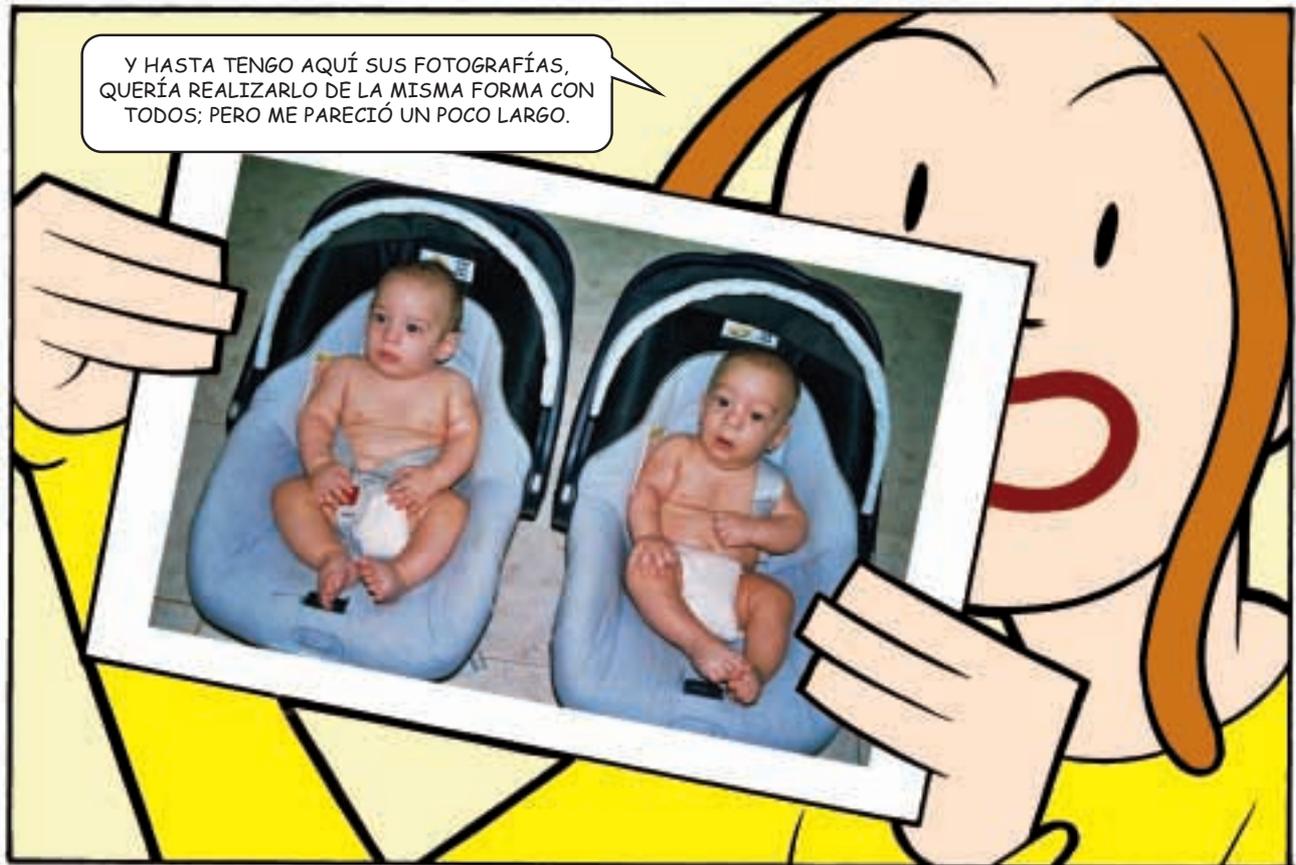
$$1+2+3+\dots+m+\dots+= \sum_{n=1}^{\infty} n$$


## Capítulo 1

3300	2750	3150	3250	3750	3880	3700	3245	3270	3840
3740	3060	3450	3650	2870	3990	3475	3250	2910	2960
3260	3935	2500	3000	3410	3600	2340	3740	3060	2450
2750	3260	3800	3250	3800	3000	3200	2950	3870	2400
3150	3012	2925	3490	3140	4000	2900	2700	1860	3400
3400	3280	3250	3650	3720	2550	3200	3000	2050	3580
3200	3220	3050	3000	3440	3960	3580	2690	3900	2950
2940	3300	3975	3370	3200	3150	3560	3390	3360	4080
2850	3080	3850	3330	2860	3000	1900	3330	3550	3630
3000	2800	3880	2314	3550	3180	3090	3650	3160	3400
2850	3050	2900	3200	1056	3200	2800	3340	3090	3200
3040	2585	2360	4319	3750	3250	2650	3750	3060	3100
3250	3580	3000	3200	3040	3440	3650	2950	2600	3600
3250	3680	3200	4200	2550	3000	3000	2750	2040	3650
3750	2980	3700	2700	3150	3200	3420	3550	3250	3470
3250	3105	3105	3370	3300	3750	3460	2900	3140	3970
3350	3000	3480	4320	3320	2700	2440	3800	3250	3050
2500	3050	1900	3100	3480	2850	3270	3350	2850	4190
3160	3895	3030	3610	3500	3400	3400	2750	3890	3900
2750	3250	3080	3250	3100	3600	2830	3520	3040	3550
3600	2840	2910	3090	3750	2800	3110	3000	3650	2300
2700	3750	3300	3475	3120	3550	3980	3500	4500	3050
3980	3100	3300	2800	3550	3080	2740	3200	3750	3325
3170	3200	3120	3710	3300	3300	4070	3170	3850	1224
2425	3155	3240	3550	2900	2500	3300	2940	3250	3650
3220	2290	3600	2850	3600	3950	3620	3600	4350	3400
4250	3270	2880	2950	2860	3700	3590	2910	3350	2630
3840	3200	2625	2450	3520	3320	2600	3200	2690	4350
3250	3330	3575	3350	2880	3250	2650	3730	4050	4180
3700	3640	3400	3400	2400	2700	3480	3210	3000	3460
4000	3200	3300	3180	3680	3700	2700	3840	2460	3850
3270	2990	3400	3240	3130	3220	3905	2850	3330	2860
2750	2725	3300	3940	3050	3700	2900	3070	4000	4000
3400	3575	3300	3000	2970	3100	3150	3040	3150	3550
2900	2670	3340	3600	2400	4200	2650	3480	2860	3250
3580	2590	2725	3770	3550	3290	2575	2950	500	3695
3150	3315	2500	3220	3200	3750	3567	2600	1605	3530
2800	3620	3550	2875	3220	2800	3105	3050	1408	3430
2665	2680	3200	3420	3230	2950	3400	3400	3490	2665
3250	3310	3660	3195	3250	3960	2060	2040	3265	3500
2620	2295	3550	3440	2900	4000	2780	3440	1990	2530
2550	3315	3600	3240	4050	3700	3649	2850	2695	3670
3500	3680	3050	3340	3640	3080	2860	3450	3290	3000
3100	2510	3370	2550	3530	3200	4470	3910	2820	3090
3950	3330	3000	3160	3900	3100	2825	3380	2690	2200
2620	3620	3200	3000	3500	3250	3470	3490	3880	3350
2590	3950	2630	2930	2675	3870	3150	2980	2600	2910
3120	3100	3000	2250	3200	2840	4365	3660	3100	3520
2500	3750	2900	3180	2540	2670	3250	3160	3800	3300
3560	3400	3750	3870	2950	2900	3000	3750	2380	3600

3250	3260	3360	3200	3280	3220	3160	3820	2315	3250
4320	2800	3160	3300	2950	3300	3650	3040	2870	3500
3100	2750	3050	3610	3200	3650	2740	3460	3150	3630
2700	3380	3800	3700	3300	2640	4020	3980	3750	3200
3150	2200	2550	3000	3400	3680	3750	3520	3000	3600
3700	3645	2800	3135	3200	3365	3200	4190	4250	2780
3080	2750	3340	3200	3250	4200	3240	2490	2960	2550
3550	3610	3970	3750	3050	4350	3715	3180	3370	3750
3600	3570	1700	4050	3500	3000	3250	2200	3700	2950
3420	2800	3256	3900	2920	3600	2500	3520	4200	3740
2880	2470	3100	3900	3090	3650	3200	3200	3180	3170
3700	2920	3370	3990	3604	3750	3400	3750	4050	4120
3600	3200	3760	3100	3270	3250	3100	3470	2650	4200
2940	3450	4100	2300	2690	3100	2945	3790	3875	3500
3430	3225	3130	3300	2950	3750	2800	3740	3800	2800
2700	3260	3500	3400	3250	3720	3100	3060	2900	2650
2700	3800	4820	3050	3260	3310	3415	3760	2970	2745
2800	3350	3870	3210	3250	4200	3280	2810	3400	3700
2870	3550	3350	2850	3200	2750	3100	3400	1470	3350
3050	3370	2820	3000	3220	3950	2800	3280	2625	3225
3110	2630	3320	2850	3750	3350	3225	3030	2515	2100
3420	3200	2770	3080	2680	3430	3560	3250	2000	2770
3720	3050	3850	3050	3250	3495	2250	3460	2540	3200
2700	4080	2910	4150	2300	4350	3600	3350	2650	3090
3550	2180	3370	2780	3700	3390	3605	3070	3565	3050
3250	3650	2850	3140	3350	2350	2980	2850	3690	2700
3180	3230	3300	3400	3280	3195	2625	4080	3200	3950
3150	3750	3060	3400	2960	3250	2900	3250	3750	3830
2750	3550	2590	3050	3140	3250	2890	3570	3350	3250
3300	3600	3600	3500	2900	3710	2350	3300	1600	2950
3640	2820	3350	3550	3470	3890	2860	3360	3435	3350
2310	3530	2100	2800	2930	3080	3065	2910	3350	3200
3800	3350	3200	2900	3320	2310	2985	2770	2760	3800
3650	3280	3100	3400	2670	4060	3200	3180	4000	3750
2880	3100	2800	2970	3190	3750	2300	3210	2620	3300
2420	3500	3200	3150	3190	3240	3515	1810	3950	3710
2950	3200	3450	3680	3410	3200	3460	3920	3750	3180
3800	3020	4300	3600	3530	2660	2950	3650	3350	3420
3405	3310	3300	2950	2300	3300	3400	3060	3600	2470
3000	2370	3600	3400	3850	3480	2390	3700	3350	3370
3410	2550	3170	2550	4300	3820	3090	3700	3100	2780
3290	2780	3380	2450	3390	3860	2950	3650	648	3690
4500	3950	2700	3500	3900	2800	3350	2830	3650	4700
3610	2980	3076	3000	2950	3750	2900	3600	3000	3390
2910	3215	1800	3300	3895	3440	3250	3220	3380	3400
3170	3900	3450	4175	3810	2950	3080	4000	3500	3150
3330	3950	3300	2600	2750	3300	3215	3870	3260	2675
3250	3530	3200	4200	2835	3500	2750	3045	3690	2830
3220	3600	2930	3600	2720	3400	2620	3600	3850	3455
3500	3810	2690	2650	4120	3410	3770	2820	3550	3100





MEDIA, VARIANZA Y DESVIACIÓN ESTÁNDAR

Media aritmética:  $\mu = \frac{\sum x_i}{n} = 3234,218$

Varianza:  $\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} = 250697,7025$

Desviación típica o estándar:  $\sigma = \sqrt{250697,7025} = 500,6972$



PUEDES PROPONER TAMBIÉN UN ACERTIJO, PARA RESOLVERLO POCO A POCO, ANTES DE FINAL DE CURSO.



### ACERTIJO:

- EXISTEN 5 CASAS CON FACHADAS EN DIFERENTES COLORES.
- EN CADA UNA DE LAS CASAS VIVE UNA PERSONA CON UNA DIFERENTE NACIONALIDAD.
- LOS 5 DUEÑOS BEBEN UNA DETERMINADA BEBIDA, TIENEN UNA DETERMINADA PROFESIÓN Y UNA DETERMINADA MASCOTA.
- NINGÚN DUEÑO TIENE LA MISMA MASCOTA, NI TIENE LA MISMA PROFESIÓN, NI BEBE LA MISMA BEBIDA.

LA PREGUNTA ES: **¿QUIÉN TIENE EL PEZ?**



### CLAVES:

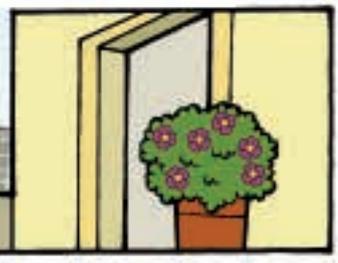
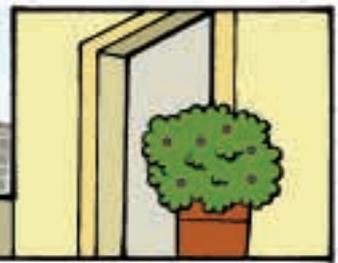
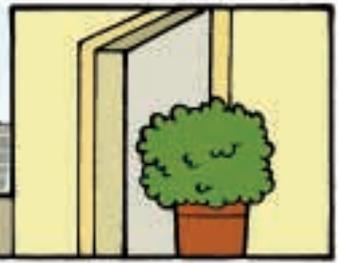
1. EL BRITÁNICO VIVE EN LA CASA ROJA.
2. EL SUECO TIENE COMO MASCOTA UN PERRO.
3. EL DANÉS TOMA TÉ.
4. LA CASA VERDE ESTÁ A LA IZQUIERDA DE LA CASA BLANCA.
5. EL DUEÑO DE LA CASA VERDE TOMA CAFÉ.
6. LA PERSONA QUE ES FÍSICO TIENE UN PÁJARO.
7. EL DUEÑO DE LA CASA AMARILLA ES BIÓLOGO.
8. EL QUE VIVE EN LA CASA DEL CENTRO TOMA LECHE.
9. EL NORUEGO VIVE EN LA PRIMERA CASA.
10. LA PERSONA QUE ES QUÍMICO VIVE JUNTO A LA QUE TIENE UN GATO.
11. LA PERSONA QUE TIENE UN CABALLO VIVE JUNTO AL QUE ES BIÓLOGO.
12. EL QUE ES INFORMÁTICO BEBE ZUMO DE POMELO.
13. EL ALEMÁN ES MATEMÁTICO.
14. EL NORUEGO VIVE JUNTO A LA CASA AZUL.
15. EL QUÍMICO TIENE UN VECINO QUE TOMA AGUA.

DICEN QUE **EINSTEIN** ESCRIBIÓ UN ACERTIJO SIMILAR EN EL SIGLO PASADO Y DIJO QUE EL 90% DE LA POBLACIÓN MUNDIAL NO LO PODRÍA RESOLVER. NO ES DIFÍCIL, SÓLO DEBES PONER MUCHA ATENCIÓN, CONCENTRACIÓN Y SER PACIENTE.

## CAPÍTULO 2



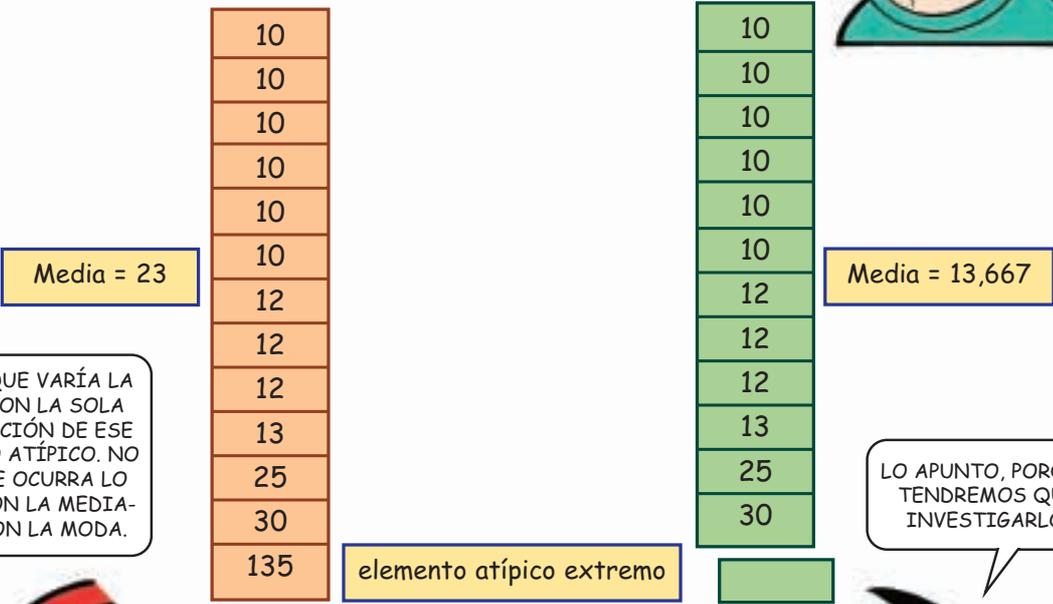
KARL PEARSON, LONDRES (1857-1936)





ES VERDAD QUE LA MEDIA, BUENO LA MEDIA ARITMÉTICA, A LA QUE LLAMAMOS SIMPLEMENTE MEDIA, SE VE MUY AFECTADA POR LOS VALORES

MIRAD ESTE EJEMPLO QUE HE ENCONTRADO:



PUES SÍ QUE VARÍA LA MEDIA, CON LA SOLA DESAPARICIÓN DE ESE ELEMENTO ATÍPICO. NO CREO QUE OCURRA LO MISMO CON LA MEDIANA, NI CON LA MODA.



LO APUNTO, PORQUE TENDREMOS QUE INVESTIGARLO.



Suma	299	164
nº elementos	13	12
Cociente	23	13,667



ADEMÁS, SI LA DISTRIBUCIÓN VINIERA DADA EN CLASES DE LA SIGUIENTE FORMA:

Clases:	Frecuencias:	Marcas de clase:
De 0 a 20	5	10
Más de 20 a 40	12	30
Más de 40 a 70	7	55
Más de 70	4	¿?



¿QUÉ MARCA DE CLASE ASIGNAMOS A LA ÚLTIMA...? ¿CÓMO HALLAMOS LA MEDIA? ¿NO SERÍA MEJOR, EN ESTE CASO LA MEDIANA?...



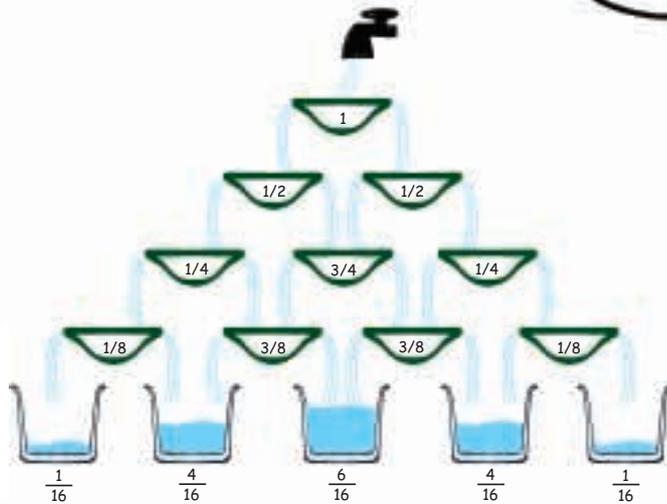
ESTAMOS DESCUBRIENDO QUE A PESAR DE LA IMPORTANCIA Y LA FACILIDAD DE CÁLCULO DE LA MEDIA ARITMÉTICA, DEBEREMOS TENER EN CUENTA OTRO TIPO DE MEDIDAS CENTRALES, SEGÚN QUÉ CASOS Y, SI ES POSIBLE, EN TODOS PARA UNA MAYOR Y MEJOR DESCRIPCIÓN DE LA DISTRIBUCIÓN.



SERÍA, PUES, INTERESANTE REPARAR OTRAS MEDIAS, ... POR EJEMPLO LA MEDIA GEOMÉTRICA.



PRECISAMENTE EL OTRO DÍA, REVISANDO NUESTRAS PRIMERAS EXPERIENCIAS, ME ENCONTRÉ CON EL GRÁFICO DE LAS TIRADAS DE LAS MONEDAS.



Y PENSÉ QUE SI EN UNA TIRADA HAY UNA PROBABILIDAD DE SALIR CARA DE  $\frac{1}{2}$  Y EN CINCO TIRADAS LA PROBABILIDAD DE QUE TODAS SEAN CARAS ES DE  $\frac{1}{32}$  PODRÍA HALLAR LA PROBABILIDAD EN TRES TIRADAS, MEDIANTE UNA MEDIA.

PROBEMOS LA MEDIA ARITMÉTICA:

$$\bar{p} = \frac{\frac{1}{2} + \frac{1}{32}}{2} = \frac{16 + 1}{32} = \frac{17}{64}$$

MAL, POR ESTE CAMINO NO LLEGAMOS.

PUES PROBEMOS LA MEDIA GEOMÉTRICA:

¡EUREKA! CUANDO HABLEMOS DE DISTRIBUCIÓN VEREMOS ALGO SOBRE ESTA CUESTIÓN...

Media geométrica =  $\sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \dots \cdot x_i^{n_i}}$       $n = \sum_{j=1}^i n_j$

$$M_g = \sqrt{\frac{1}{2} \cdot \frac{1}{32}} = \sqrt{\frac{1}{64}} = \frac{1}{8}$$

PERO SÍ DEBEMOS ACLARAR VARIAS COSAS ANTES DE SEGUIR.

UNA PUEDE SER QUE LAS MEDIAS, SEAN LAS QUE SEAN, SON SÓLO ESTADÍSTICOS QUE CALCULAMOS EN LA BÚSQUDA DE UNA REPRESENTACIÓN MÁS SIMPLIFICADA DE LA TABLA DE DATOS INICIAL.

TÚ TE HAS DADO CUENTA DE ESA CUESTIÓN, CON EL ÚLTIMO EJEMPLO QUE PUSO GAUSS, EN DONDE LAS MEDIAS ERAN 23 Y 13,667.

SÍ, QUE ERAN VALORES NO COINCIDENTES CON NINGUNO DE LOS DATOS, O SEA NI SIQUIERA PERTENECÍAN COMO TALES A LA DISTRIBUCIÓN.

Y EN EL EJEMPLO DE LA MEDIA GEOMÉTRICA, LA PERFECCIÓN DEL RESULTADO "UN OCTAVO" ES DEBIDA A LA ELECCIÓN DE LOS DATOS.

Media geométrica =  $\sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \dots \cdot x_i^{n_i}}$       $n = \sum_{j=1}^i n_j$

$$M_g = \sqrt{\frac{1}{2} \cdot \frac{1}{32}} = \sqrt{\frac{1}{64}} = \frac{1}{8}$$

O SEA, QUE LA CONCLUSIÓN QUE DEBÍAMOS SACAR DE LA APLICACIÓN DE LA MEDIA GEOMÉTRICA, COMO MÁS ACERTADA QUE LA ARITMÉTICA, SERÍA EN AQUELLAS POBLACIONES CON CRECIMIENTO NO PROPORCIONAL, PODRÍAMOS DECIR CON CRECIMIENTO EXPONENCIAL.

$$\text{Media}_{\text{geométrica}} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \sqrt[n]{\prod_{j=1}^n x_j}$$

$$M_g = \sqrt{\frac{1}{2} \cdot \frac{1}{32}} = \sqrt{\frac{1}{64}} = \frac{1}{8}$$


ES UNA BUENA TEORÍA, VERBIGRACIA: LA POBLACIÓN DE UNA CIUDAD ES EN EL AÑO 1990 DE 10.000 PERSONAS Y EN EL AÑO 2000 DE 80.000, TENDREMOS QUE SUPONER HA IDO CRECIENDO PROGRESIVAMENTE.



1990

2000

¿CUÁNTOS HABITANTES TENDRÍA EN EL AÑO 1995 (MITAD DEL PERIODO)?

$$\bar{p} = \frac{10.000 + 80.000}{2} = 45.000$$

AQUÍ, EN ESTE EJEMPLO, LA MEDIA GEOMÉTRICA PUEDE ACERCARSE A LA REALIDAD MÁS QUE LA MEDIA ARITMÉTICA:

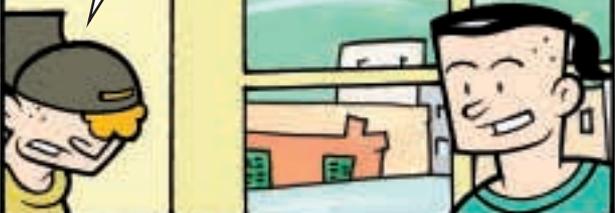
$$p_{mg} = \sqrt{10.000 \times 80.000} = 28.284'27$$


AHORA VEO DOS COSAS, UNA, QUE ES MÁS ACERTADA LA MEDIA GEOMÉTRICA EN ESTE CASO PORQUE CADA AÑO AUMENTARÁ MÁS, Y NO LA MISMA CANTIDAD TODOS LOS AÑOS COMO PRESUPONE LA ARITMÉTICA Y LA SEGUNDA... ¡SE ME HA OLVIDADO!



¡AH!... ¡YA!.. QUE SON VALORES REPRESENTATIVOS, YA QUE NO PODRÍAN JAMÁS, EN EL AÑO 1995 EXISTIR 0'27 HOMBRES.

¡VALE A MEDIAS! ... ¿TIENES MÁS OCURRENCIAS?



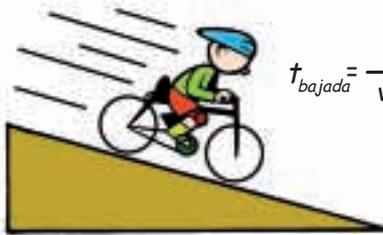






TIEMPO QUE TARDA EN SUBIR:

$$t_{\text{subir}} = \frac{\text{espacio}}{\text{velocidad}} = \frac{30 \text{ Kms}}{30 \frac{\text{Kms}}{\text{hora}}} = 1 \text{ hora}$$



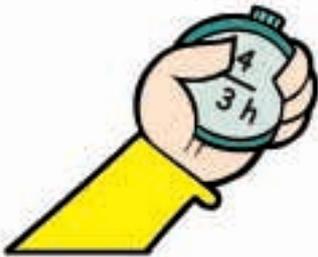
TIEMPO QUE TARDA EN BAJAR:

$$t_{\text{bajada}} = \frac{\text{espacio}}{\text{velocidad}} = \frac{30 \text{ Kms}}{90 \frac{\text{Kms}}{\text{hora}}} = \frac{30}{90} = \frac{3}{9} = \frac{1}{3} \text{ hora}$$



TOTAL DE TIEMPO TARDADO:

$$t_{\text{total}} = 1 + \frac{1}{3} = \frac{4}{3} \text{ hora}$$



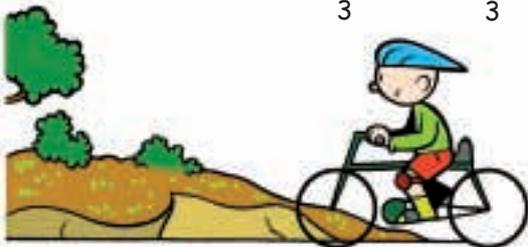
ESPACIO TOTAL RECORRIDO:

$$e_{\text{total}} = 30 + 30 = 60 \text{ hora}$$



VELOCIDAD MEDIA:

$$v_{\text{media}} = \frac{\text{espacio}}{\text{velocidad}} = \frac{60 \text{ Kms}}{\frac{4}{3} \text{ hora}} = \frac{60}{\frac{4}{3}} = \frac{60 \cdot 3}{4 \cdot 1} = \frac{180}{4} = \frac{90}{2} = 45 \text{ Kms / hora}$$



¡PUES NO HABÍA SALIDO BIEN!



LA COSA ES COMPRESIBLE, YA QUE DECÍA BIEN BINOMIO, AQUÍ LA FÓRMULA A APLICAR ES LA DE LA MEDIA ARMÓNICA:

$$x_{armónica} = \frac{n}{\sum \frac{1}{x_i}}$$

$$v_{media} = \frac{2}{\frac{1}{30} + \frac{1}{90}} = \frac{2}{\frac{3+1}{90}} = \frac{2 \cdot 90}{4} = \frac{180}{4} = 45 \text{ Kms / hora}$$

¡EXTRAORDINARIO!... PERO.... EN ESTE EJERCICIO LOS DOS TRAYECTOS ERAN IGUALES.... ¿SALDRÍA TAMBIÉN SI NO LO FUERAN...?

NO, EN ESTE CASO TENDRÍAMOS QUE PONDERAR, ES DECIR, UTILIZAR LOS CORRESPONDIENTES PESOS.

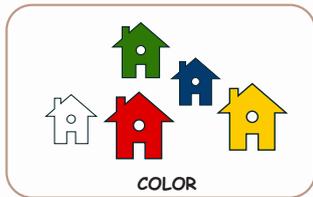
POR EJEMPLO: UN CICLISTA RECORRE UNA DISTANCIA DE 50 KMS; LOS PRIMEROS 10 KMS A UNA VELOCIDAD DE 30KMS/HORA, LOS SIGUIENTES 40 KMS A UNA VELOCIDAD DE 60 KMS/HORA. ¿CUÁL SERÁ SU VELOCIDAD MEDIA?

$$v_{media} = \frac{10 + 40}{\frac{1}{30} \cdot 10 + \frac{1}{60} \cdot 40} = \frac{50}{\frac{10}{30} + \frac{40}{60}} = \frac{50}{\frac{1}{3} + \frac{2}{3}} = \frac{50}{\frac{1}{2} + \frac{1}{2}} = 50 \text{ Kms / hora}$$

DEJADME A MÍ.

MUY BIEN, PUES LO HICE POR FÍSICA Y LOS RESULTADOS SON COINCIDENTES.





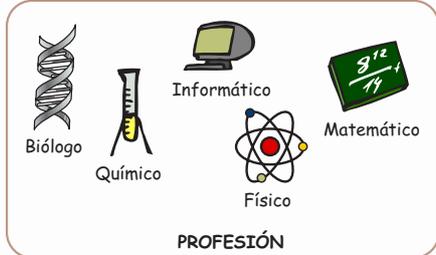
COLOR



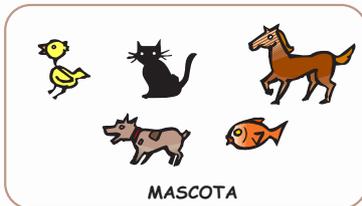
NACIONALIDAD



BEBIDA



PROFESIÓN



MASCOTA

PARA QUEDAR BIEN DIRÍAMOS QUE BUSCAMOS UN VECTOR DE PARÁMETROS, QUE AÚN DESCONOCEMOS Y QUE VENDRÍA A DETERMINAR SUS CINCO COMPONENTES.



POSTERIORMENTE CREAMOS UNA TABLA DE DOBLE ENTRADA, Y EN CADA CELDA, IREMOS INDICANDO EN ROJO FUERTE AQUELLA POSIBILIDAD QUE SE CONVIERTA EN SEGURA, POR LA LECTURA DE LAS CLAVES DEL ACERTIJO: Y EN NEGRO LA QUE ES IMPOSIBLE DE SER, PORQUE LO DICEN LAS CLAVES O PORQUE SE LE HA ASIGNADO A OTRO.

ASÍ CUANDO EN UNA CELDA TENAMOS CUATRO ASIGNACIONES EN NEGRO, SABREMOS QUE TIENE QUE SER SEGURO EL QUE NOS FALTA.



PRIMERO, LA CLAVE "NUEVE"



SEGUNDO, LA CLAVE "CATORCE"



TERCERO, LA CLAVE "OCHO"



RESULTANDO LA SIGUIENTE TABLA:

ROJO... SÍ NEGRO... NO	Casa 1ª	Casa 2ª	Casa 3ª	Casa 4ª	Casa 5ª
Color	Azul	<b>AZUL</b> 🏠	Azul	Azul	Azul
Nacionalidad	<b>NORUEGO</b> 🇳🇴	Noruego	Noruego	Noruego	Noruego
Bebida	Leche	Leche	<b>LECHE</b> 🥛	Leche	Leche
Profesión	-	-	-	-	-
Mascota	-	-	-	-	-



## CAPÍTULO 3



RONALD AYLMER FISHER,  
LONDRES (1890-1962)



GEORGE SNEDECOR,  
TENNESSEE (1881-1974)



¿ADÓNDE NOS LLEVAS?

¿A QUÉ VIENE TANTO MISTERIO?

ATENCIÓN CHICOS...



¡UUUAAU!



YA QUE NOS DEJAN USAR ESTA SALA DE NUEVAS TECNOLOGÍAS, APROVECHEMOS PARA RESOLVER PRÁCTICAMENTE ALGUNAS DE LAS CUESTIONES ANTERIORES.

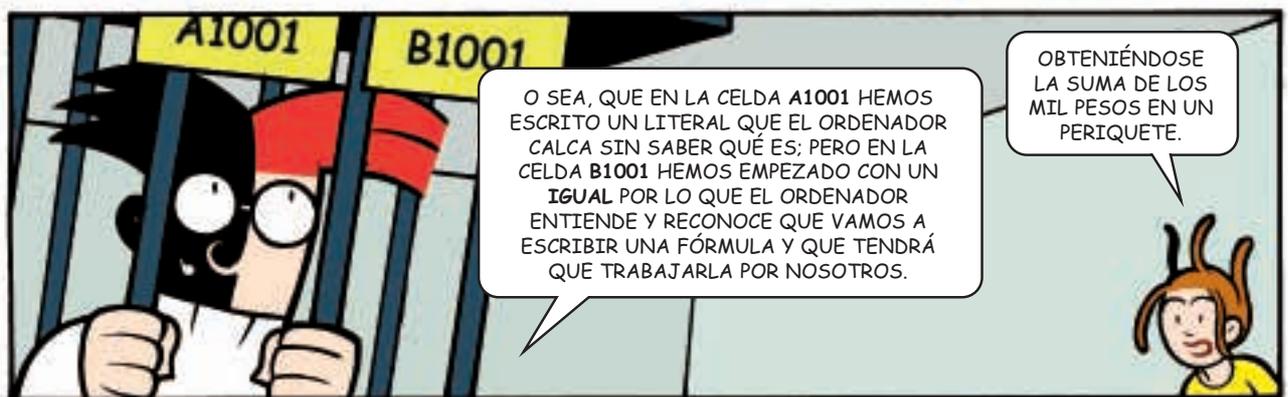
EMPECEMOS CON LA SERIE DE LOS PESOS DE LOS RECIÉN NACIDOS.



YO YA LOS TENGO COPIADOS EN UNA HOJA EXCEL® EN LA COLUMNA B, FILAS DESDE 1 A 1.000, ¿VEIS?

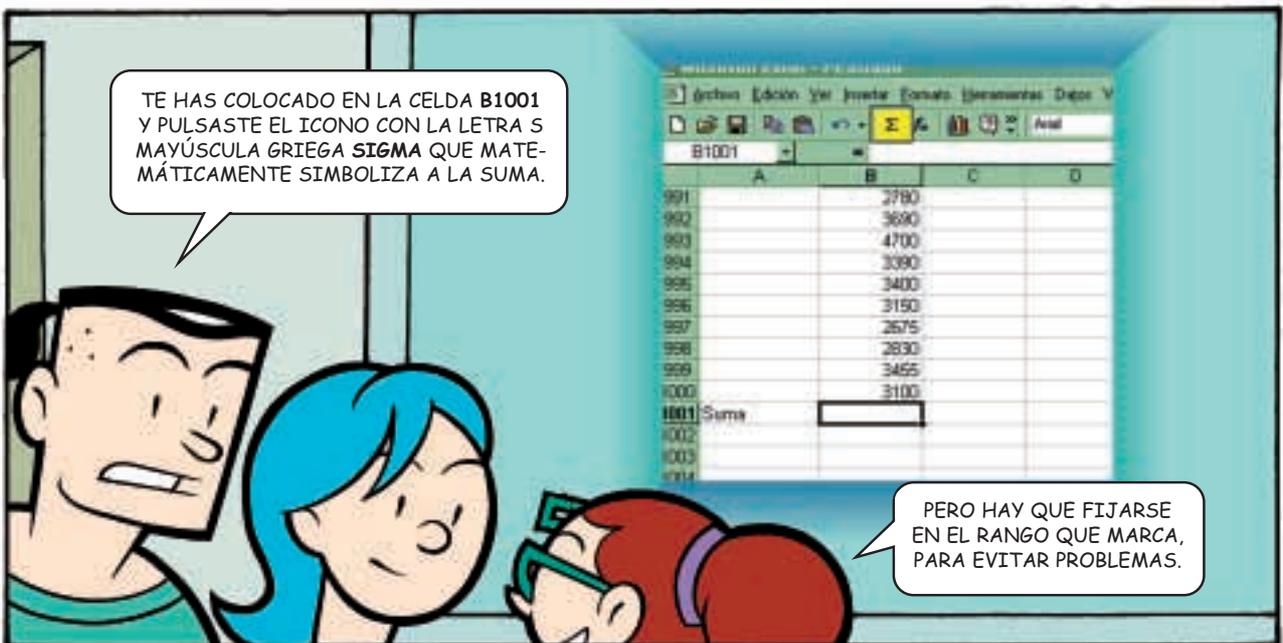
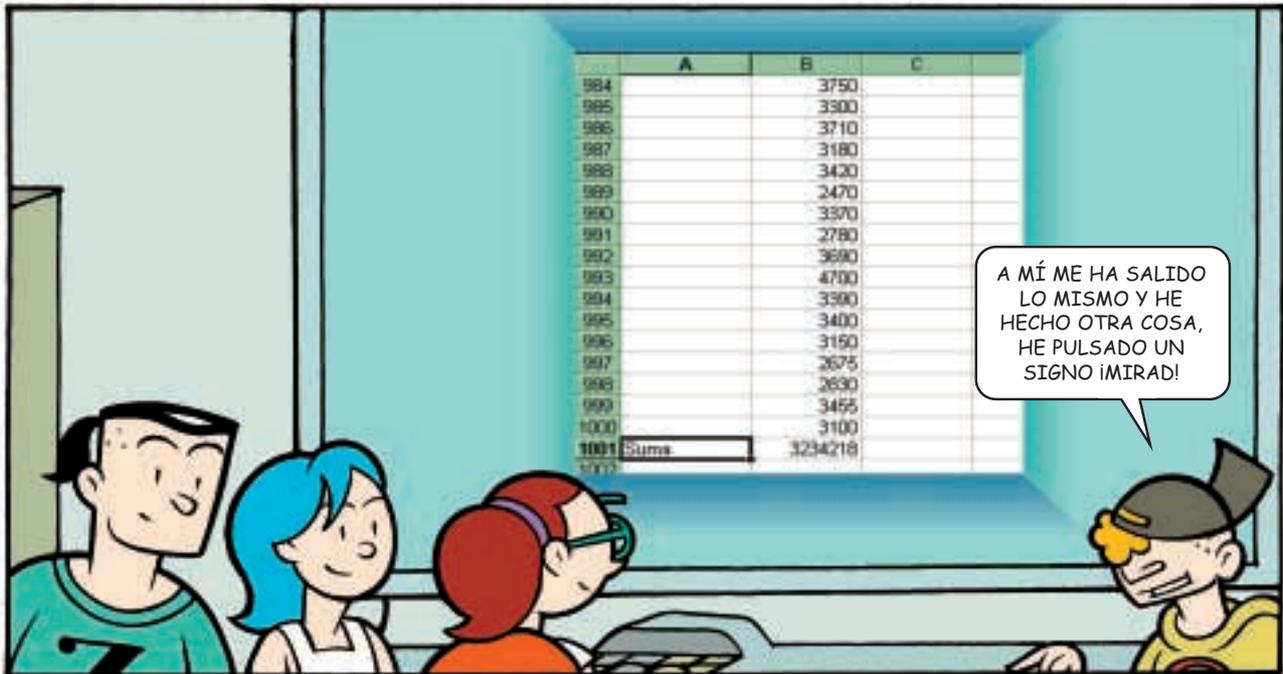


PERFECTO, AHORA PODEMOS HALLAR LA SUMA DE LAS MIL OBSERVACIONES, ACTUEMOS ASÍ:

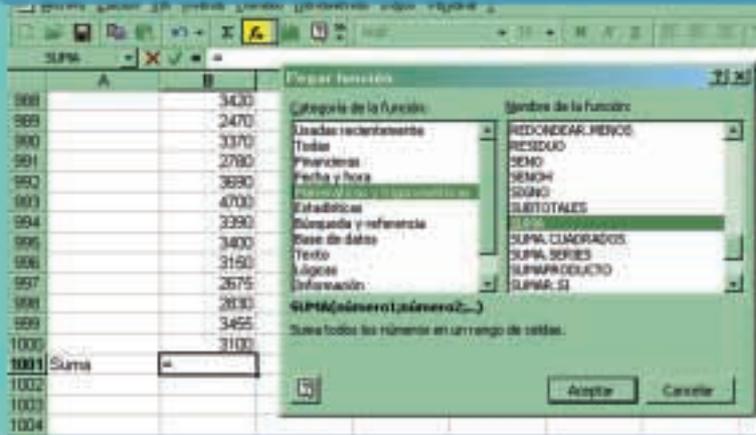


O SEA, QUE EN LA CELDA A1001 HEMOS ESCRITO UN LITERAL QUE EL ORDENADOR CALCA SIN SABER QUÉ ES; PERO EN LA CELDA B1001 HEMOS EMPEZADO CON UN IGUAL POR LO QUE EL ORDENADOR ENTIENDE Y RECONOCE QUE VAMOS A ESCRIBIR UNA FÓRMULA Y QUE TENDRÁ QUE TRABAJARLA POR NOSOTROS.

OBTENIÉNDOSE LA SUMA DE LOS MIL PESOS EN UN PERIQUETE.



PUES, YO HE IDO POR OTRO CAMINO; FIJAOS:



POR ESTE CAMINO, DE LAS FUNCIONES, TENDREMOS QUE IR PARA RESOLVER MUCHOS PROBLEMAS ESTADÍSTICOS.



PARA HALLAR LA MEDIA, TENDRÍAMOS QUE DIVIDIR ESTA SUMA ENTRE EL NÚMERO DE OBSERVACIONES, EN ESTE CASO 1.000.



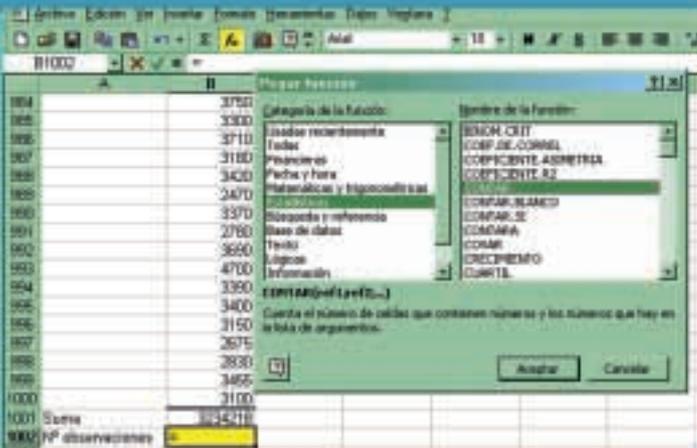
PERO HABRÁ CASOS EN QUE NO SEPAMOS CUÁNTAS OBSERVACIONES HAY.



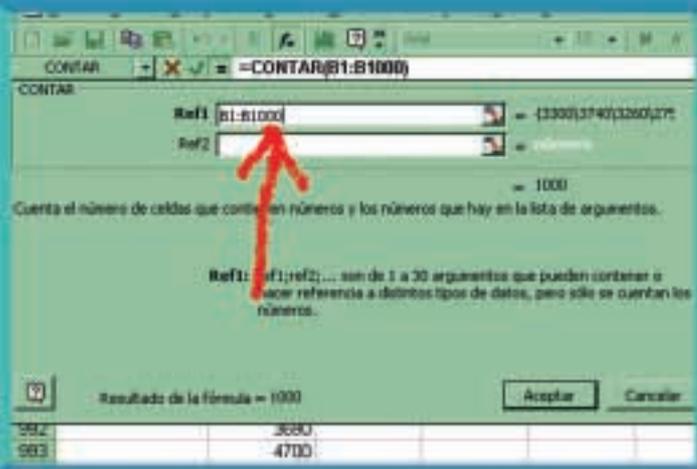
MENUDO LATAZO CONTARLAS SI SON MUCHAS, ME ESTOY PONIENDO MALITO.



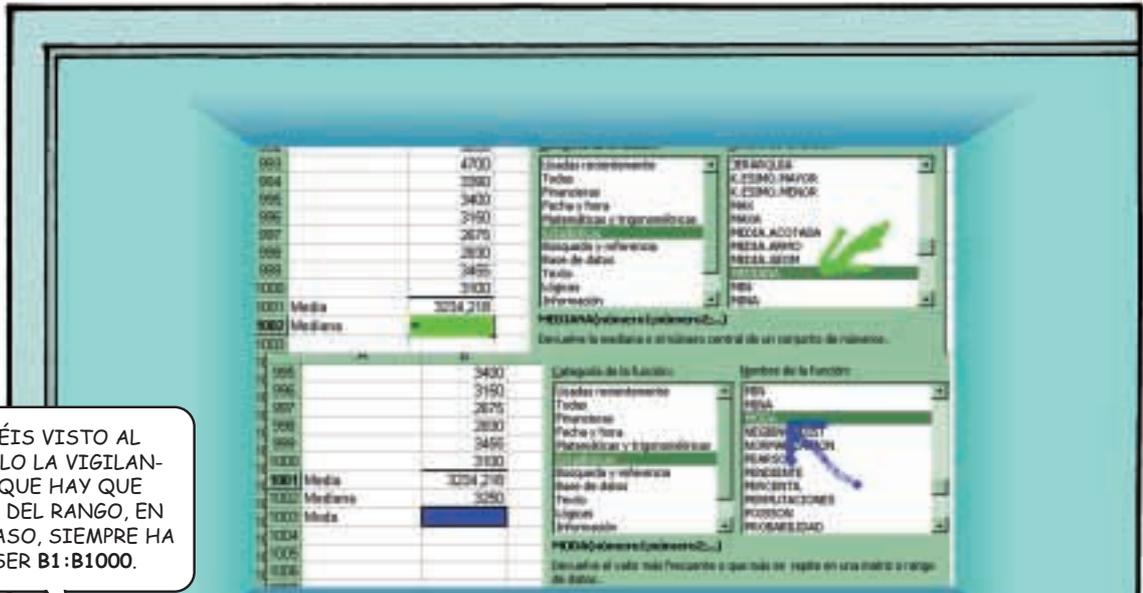
POR ESO YO HE HECHO...



Y AHORA ACEPTAMOS TENIENDO MUCHO CUIDADO CON EL RANGO A CONTAR, QUE SI NO, CUENTA TAMBIÉN LA CELDA DE LA SUMA.

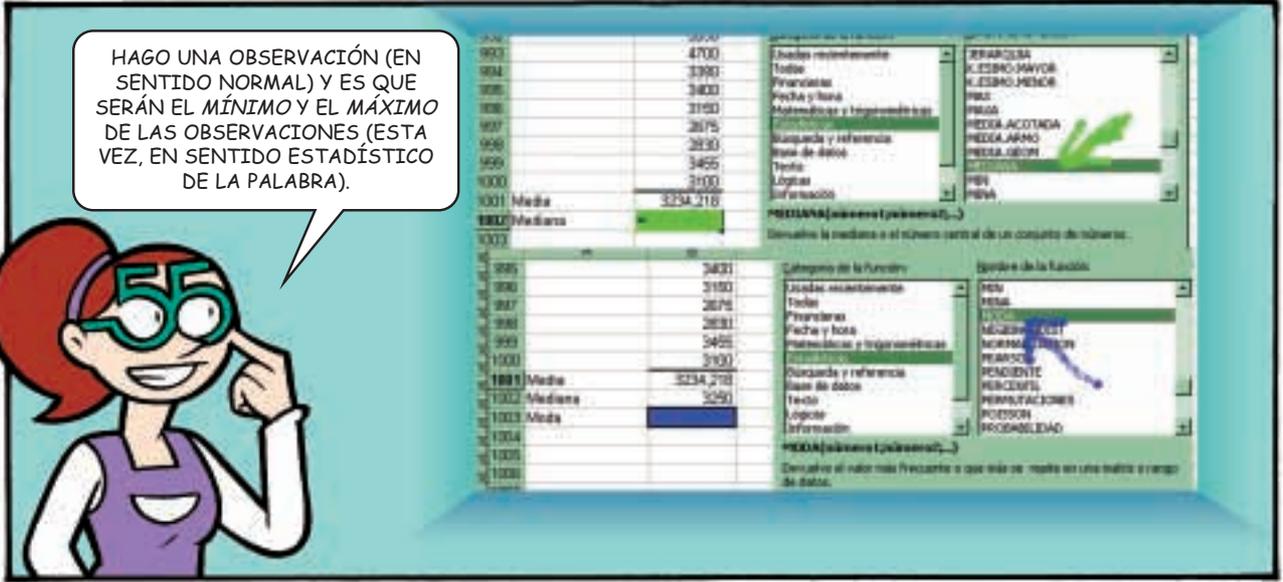






HABÉIS VISTO AL HACERLO LA VIGILANCIA QUE HAY QUE TENER DEL RANGO, EN ESTE CASO, SIEMPRE HA DE SER B1:B1000.

PUES AHORA PODÍAMOS ENCONTRAR EL VALOR MÁS PEQUEÑO Y EL MÁS GRANDE DE TODAS LAS OBSERVACIONES.



HAGO UNA OBSERVACIÓN (EN SENTIDO NORMAL) Y ES QUE SERÁN EL MÍNIMO Y EL MÁXIMO DE LAS OBSERVACIONES (ESTA VEZ, EN SENTIDO ESTADÍSTICO DE LA PALABRA).



YO TAMBIÉN HE DE HACER UNA OBSERVACIÓN...

MODAS, PUEDE HABER VARIAS, Y LA HOJA SÓLO DA LA PRIMERA QUE ENCUENTRA.

O SEA, EN ESTE CASO LA MODA ES 3.200 PORQUE ES EL DE MÁXIMA FRECUENCIA CON 40 OBSERVACIONES, PERFECTO.



PERO SUPONGAMOS, SÓLO ES UN SUPONER...

VAMOS, EN PLAN FORMAL, UNA HIPÓTESIS.



LA HIPÓTESIS DE QUE DOS VALORES, EL 3.200 Y EL 3.111 TUVIERAN LOS DOS LA MÁXIMA FRECUENCIA, 40 OBSERVACIONES, HABRÍA DOS MODAS.



Y SEGÚN DICE GAUSS, LA HOJA DE CÁLCULO SÓLO NOS DARÍA UNO, EL PRIMERO QUE ENCONTRARA.

NOS DARÍA EL 3.111, EL OTRO LO TENDRÍAIS QUE ENCONTRAR.



¿CÓMO QUE LO TENDRÍAIS...?

¿Y TÚ?

¡MUJER! YO YA HE DESCUBIERTO UNA MODA... VOSOTROS EL RESTO...



HAGAMOS EN PRIMER LUGAR UNA PANTALLA CON LAS FÓRMULAS DE CÁLCULO MANUAL.

$$\text{Var}(X) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

ó

$$\text{Var}(X) = \frac{\sum_{i=1}^n x_i^2}{n} - (\bar{x})^2$$

$$\text{Cuasi var}(X) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

ó

$$\text{Cuasi var}(X) = \frac{n}{n-1} \left\{ \frac{\sum_{i=1}^n x_i^2}{n} - (\bar{x})^2 \right\}$$





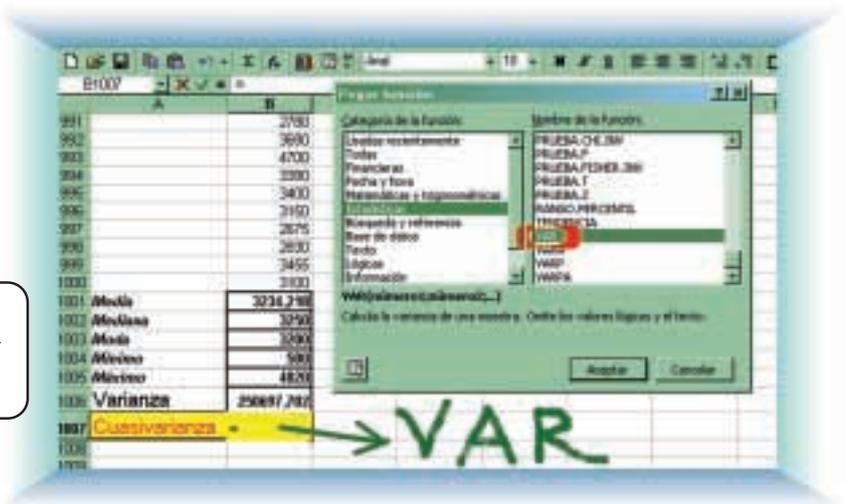
EMPECEMOS, LA VARIANZA SE HALLA DE ESTA FORMA.



¡EH! ¡EH! CREO QUE DESCUBRÍ LA CUASI-VARIANZA Y ES SIMPATIQUÍSIMO.



SI ES ASÍ, ME APUNTO.

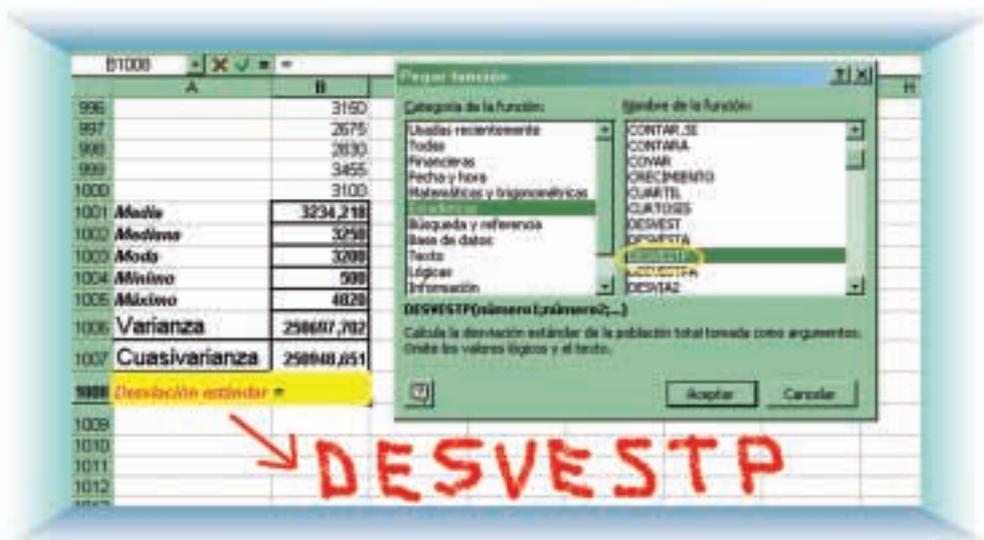




DE IGUAL FORMA HALLAREMOS LA DESVIACIÓN TÍPICA Y LA CUASIDESVIACIÓN TÍPICA, QUE SON LAS RESPECTIVAS RAÍCES CUADRADAS.

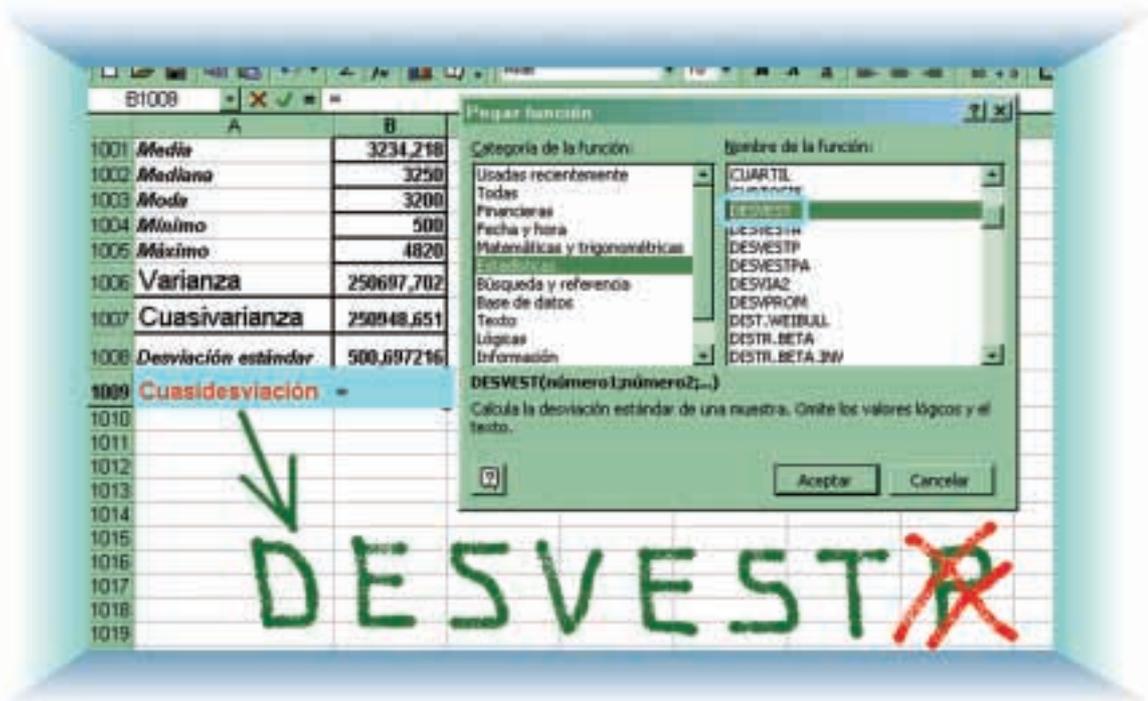
$$\text{Desviación típica o estándar} = \sqrt{\text{Varianza}}$$

$$\text{Cuasidesviación típica o estándar} = \sqrt{\text{Cuasi varianza}}$$





CUIDADO CON EL RANGO, DEL QUE QUEREMOS OBTENER LOS ESTADÍSTICOS ES SIEMPRE.....B1:B1000  
 OJO CADA VEZ QUE ACEPTAMOS UNA FUNCIÓN!





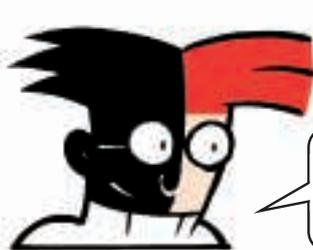
Media	3.234,2180
Mediana	3.250,0000
Moda	3.200,0000
Mínimo	500,0000
Máximo	4.820,0000
Varianza	250.697,7025
Cuasivarianza	250.984,6511
Desviación estándar	500,6972
Cuasidesviación	500,9478

**PESOS AL NACER**  
1.000 observaciones





ROJO... SÍ NEGRO... NO	Casa 1ª	Casa 2ª	Casa 3ª	Casa 4ª	Casa 5ª
Color	Roja/ Azul	<b>AZUL</b> 🏠	Azul	Azul	Azul
Nacionalidad	<b>NORUEGO</b> 🇳🇴	Británico/Noruego	Noruego	Noruego	Noruego
Bebida	Leche	Zumo de pomelo Leche	<b>LECHE</b> 🥛	Leche	Leche
Profesión	-	Biólogo	Informático	-	-
Mascota	-	-	-	-	-



AHORA SÍ QUE PODEMOS TOMARNOS UN DESCANSO HASTA EL PRÓXIMO CAPÍTULO.

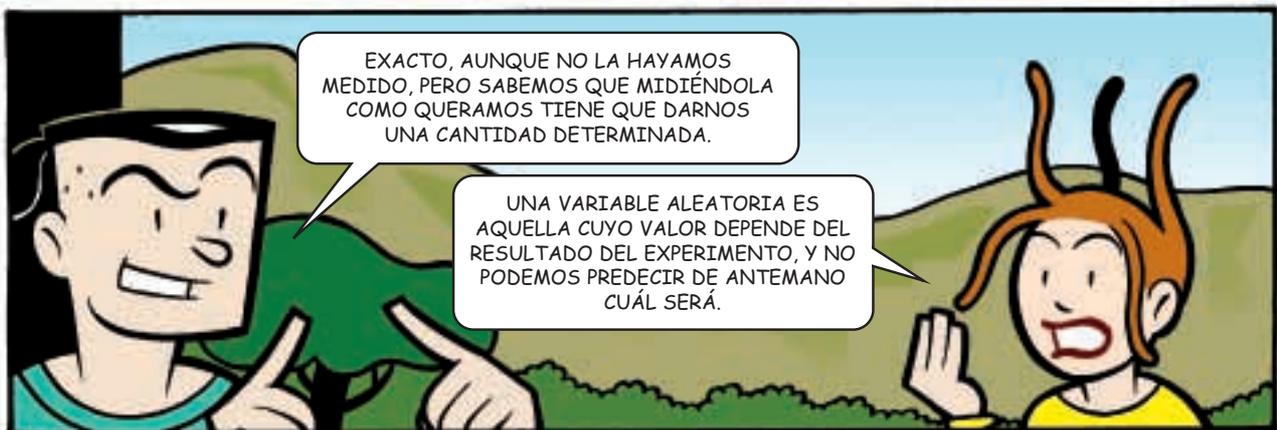


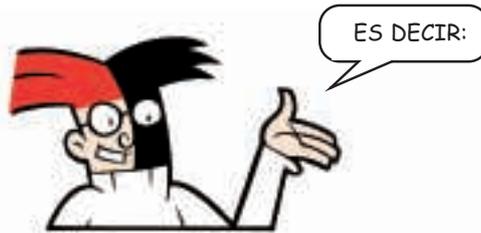
# CAPÍTULO 4



GERTRUDE MARY COX, DAYTON, IOWA (1900-1978)







Variable determinista:  $X = \text{"mi peso ahora"}$   
 $X = 41,350 \text{ Kilogramos}$

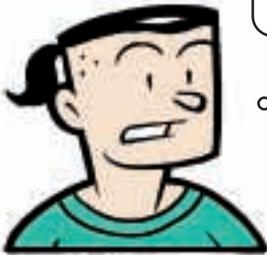
Variable aleatoria:  $X = \text{"sacar 6 al lanzar un dado"}$   
 $X = \{0, 1\}$



ES MUY IMPORTANTE LA APOSTILLA QUE HA HECHO GRÁFICA:  
**LANZAR UN DADO DOS VECES = LANZAR DOS DADOS A LA VEZ**  
 QUE DESPUÉS OBSERVAREMOS CON DETALLE; PERO AHORA  
 NOS FIJAREMOS EN QUE UNIDO A LA VARIABLE ALEATORIA  
 TENEMOS LO QUE EN PRINCIPIO LLAMAREMOS "UNA IDEA  
 DE CANTIDAD DE POSIBILIDAD".



EL RESULTADO DE UNA TIRADA ES INDEPENDIENTE  
 DE LA OTRA; Y AL TIRAR DOS DADOS LOS RESULTA-  
 DOS SON INDEPENDIENTES EL UNO DEL OTRO.



AHORA SÍ QUE LA HEMOS LIADO...



ESTUDIEMOS DESPACIO  
 ESTA CUESTIÓN CON LA  
 AYUDA DE GRÁFICA.

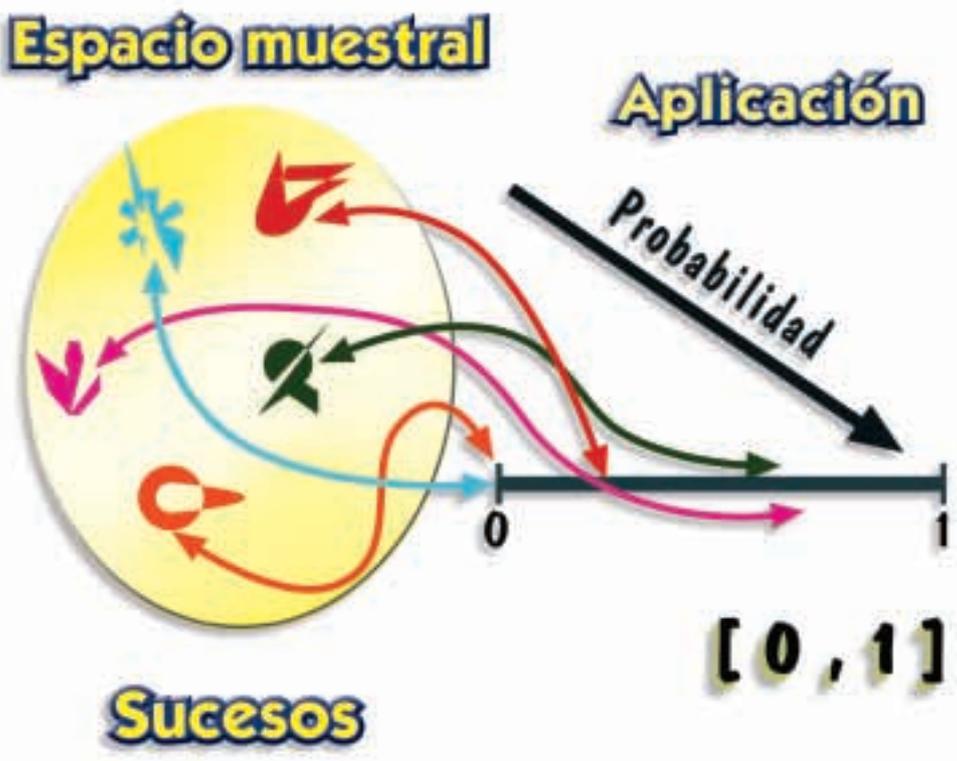
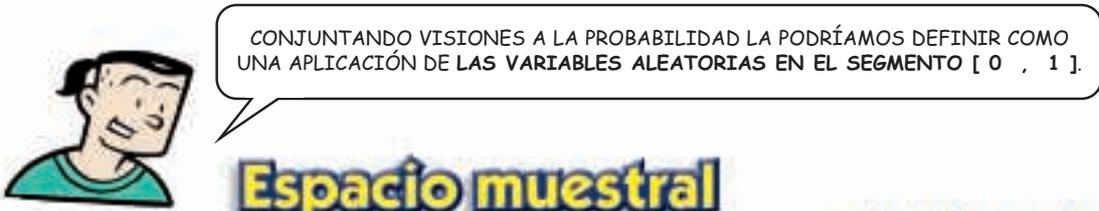
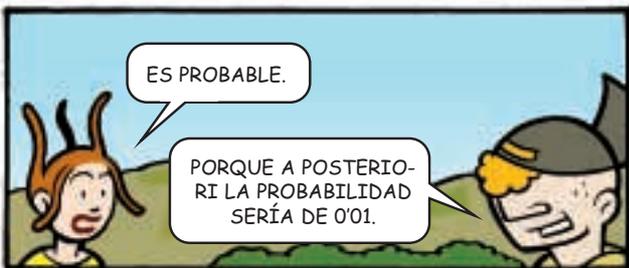
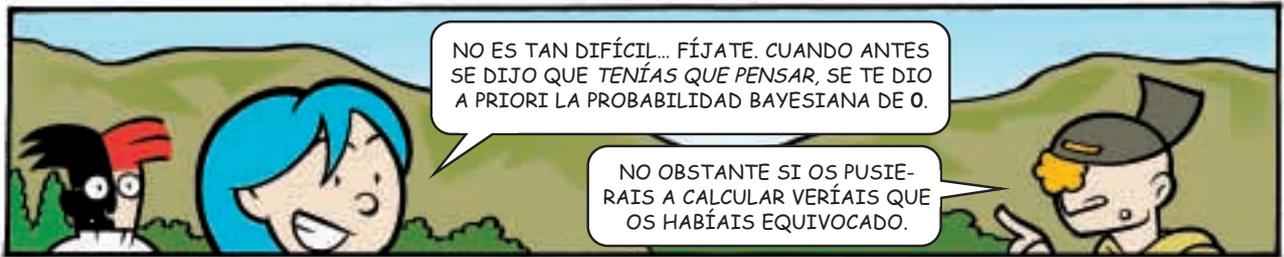
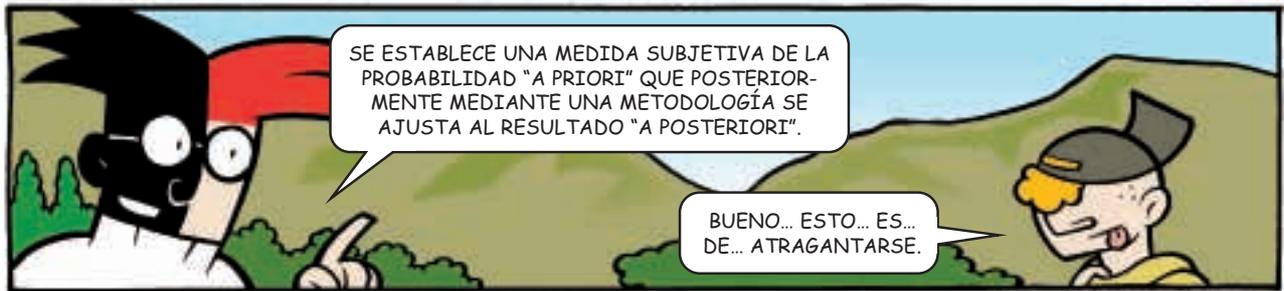
Variable:	Resultado:	Posibilidad:
<b>Determinista</b> x = "Área de un cuadrado lado 1 m."	1 m. Cuadrado	Seguro
<b>Aleatoria</b> X = "resultado del lanzamiento de un dado"	1	Posible
	2	Posible
	3	Posible
	4	Posible
	5	Posible
	6	Posible

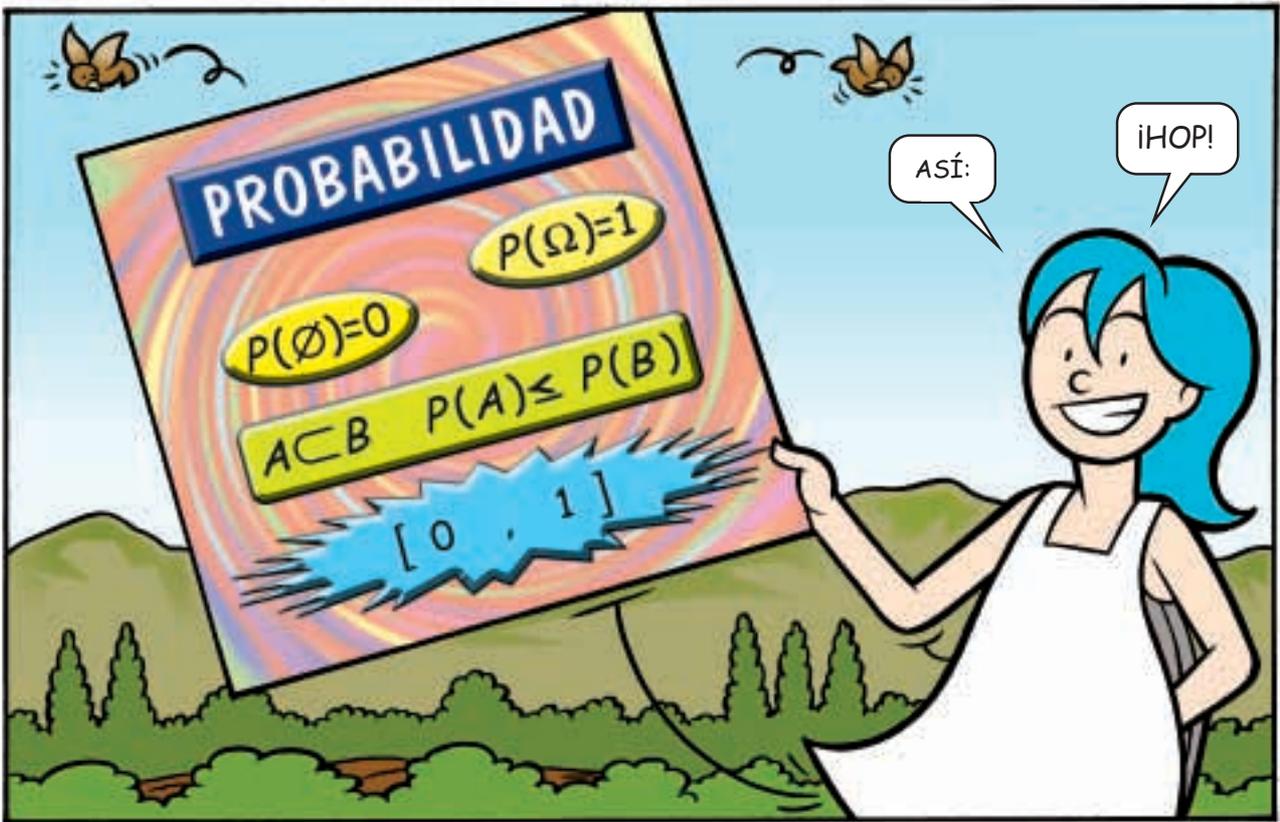


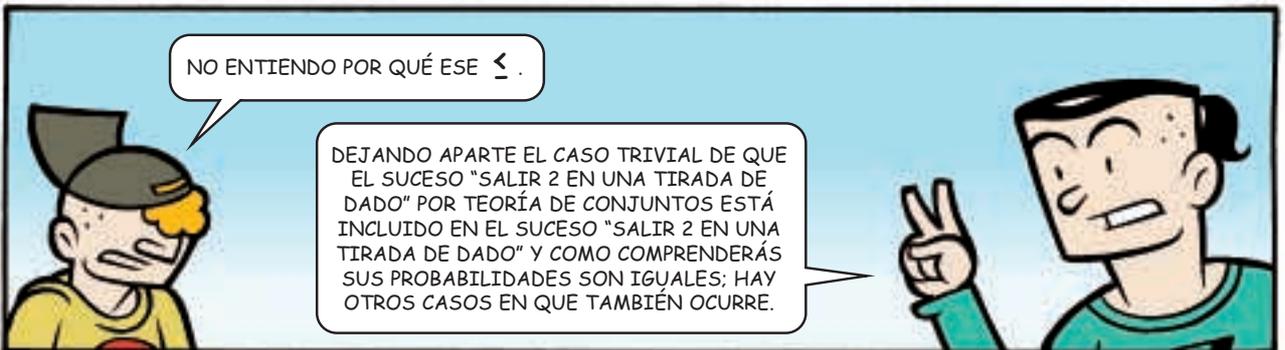












 DÉJAME A MÍ:

y como verás fácilmente  $P\{\text{salir } 2\} < P\{\text{salir } 2 \text{ ó } 3\}$

puesto que:  $\frac{1}{6} < \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$

en cambio:  $\{\text{salir } 2\} \subset \{\text{salir } 2 \text{ ó } 7\}$

$P\{\text{salir } 2\} = P\{\text{salir } 2 \text{ ó } 7\}$

puesto que:  $\frac{1}{6} = \frac{1}{6} + 0$



CREO QUE PODRÍAMOS REPASAR LOS SUCESOS MEDIANTE GRÁFICOS Y HACERNOS UNA IDEA DE SU APLICACIÓN CON LA PROBABILIDAD.

**Espacio Muestral**

S01	S02	S03	S04	S05	S06	S07
S08	S09	S10	S11	S12	S13	S14
S15	S16	S17	S18	S19	S20	S21
S22	S23	S24	S25	S26	S27	S28
S29	S30	S31	S32	S33	S34	S35
S36	S37	S38	S39	S40	S41	S42

**Sucesos elementales equiprobables**

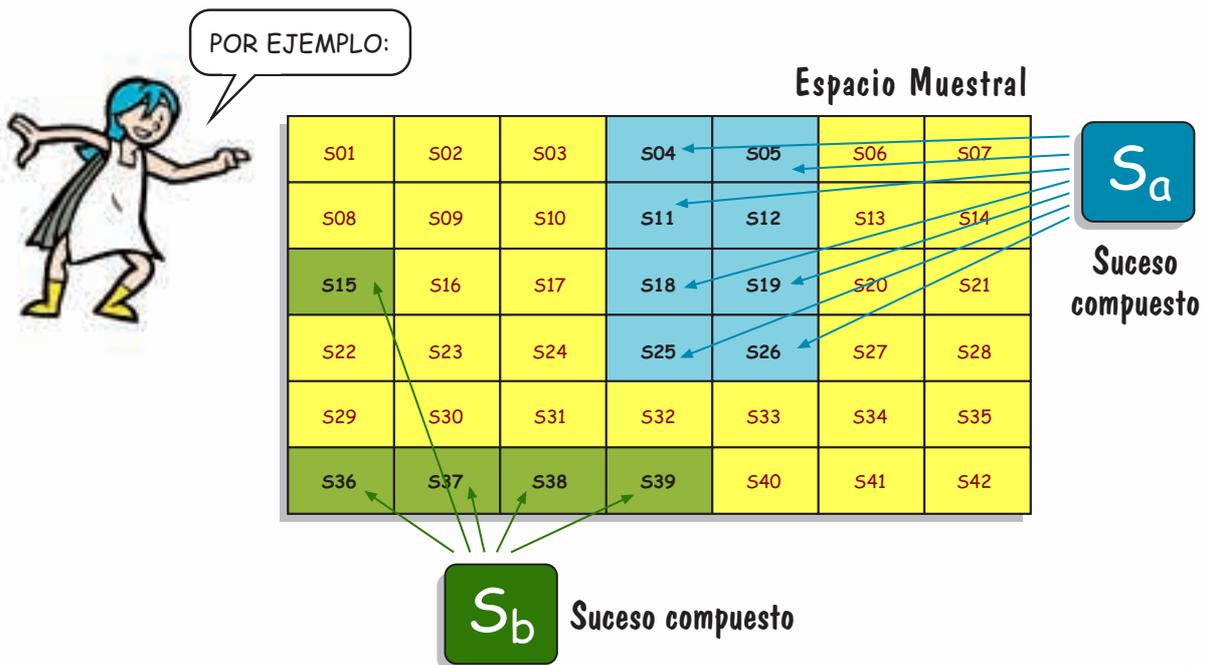
S01	S02	S03	S04	S05	S06	S07
S08	S09	S10	S11	S12	S13	S14
S15	S16	S17	S18	S19	S20	S21
S22	S23	S24	S25	S26	S27	S28
S29	S30	S31	S32	S33	S34	S35
S36	S37	S38	S39	S40	S41	S42

**Sucesos elementales no equiprobables**



SON DOS ESPACIOS MUESTRALES, CONJUNTO DE TODOS LOS SUCESOS ELEMENTALES.

EN EL PRIMERO, TODOS LOS SUCESOS TIENEN LA MISMA PROBABILIDAD DE SALIR.



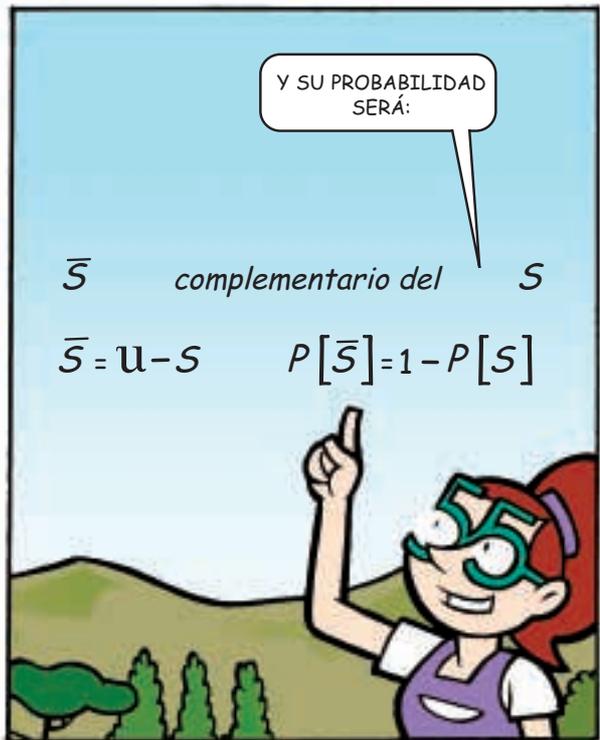
EL SUCESO "A" ESTÁ FORMADO POR LA UNIÓN DE VARIOS SUCESOS ELEMENTALES.

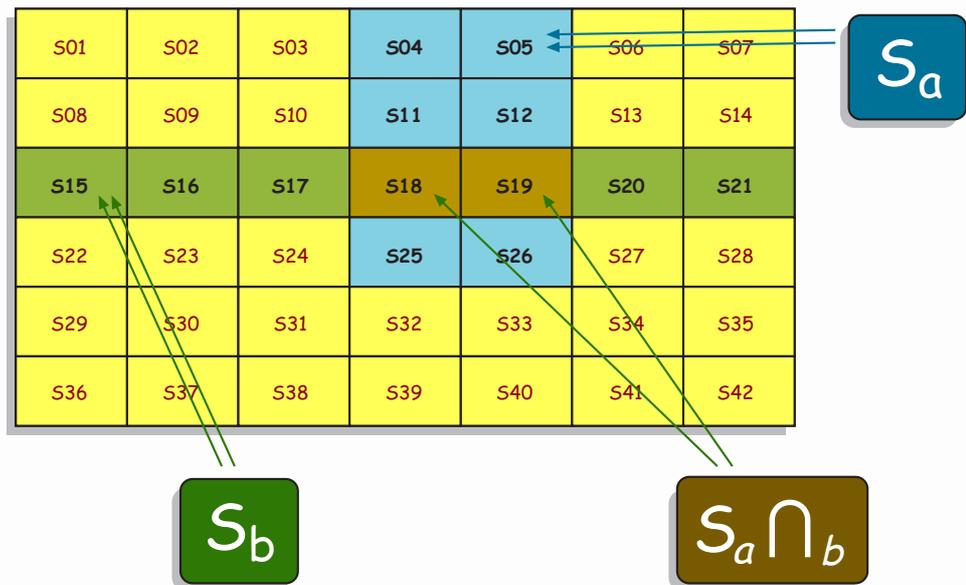
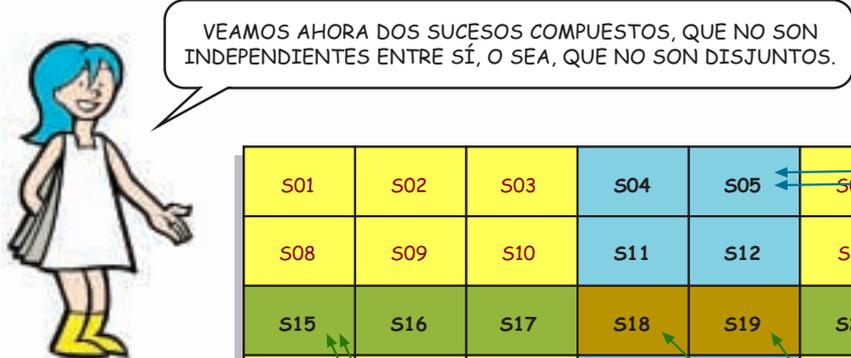


$$s_a = (s_4 \cup s_5 \cup s_{11} \cup s_{12} \cup s_{18} \cup s_{19} \cup s_{25} \cup s_{26})$$

ASÍ COMO EL SUCESO "B".

$$s_b = (s_{15} \cup s_{36} \cup s_{37} \cup s_{38} \cup s_{39})$$

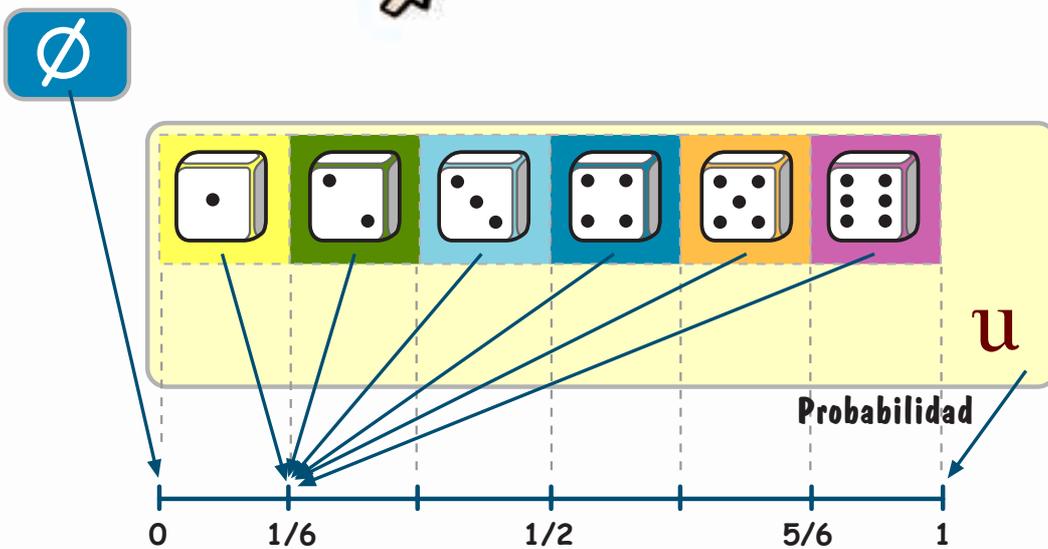




$$s_a = (s_4 \cup s_5 \cup s_{11} \cup s_{12} \cup s_{18} \cup s_{19} \cup s_{25} \cup s_{26})$$

$$s_b = (s_{15} \cup s_{16} \cup s_{17} \cup s_{18} \cup s_{19} \cup s_{20} \cup s_{21})$$

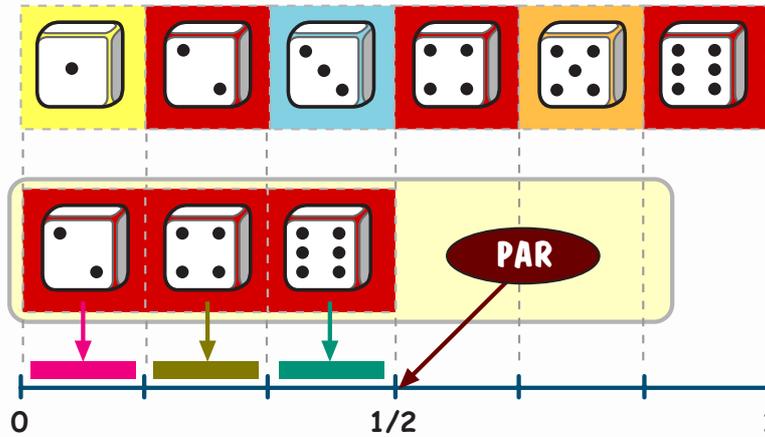
$$s_a \cap b = s_a \cap s_b = (s_{18} \cup s_{19})$$





HAGAMOS LO MISMO CON EL SUCESO COMPUUESTO:

SACAR PAR AL LANZAR UN DADO.



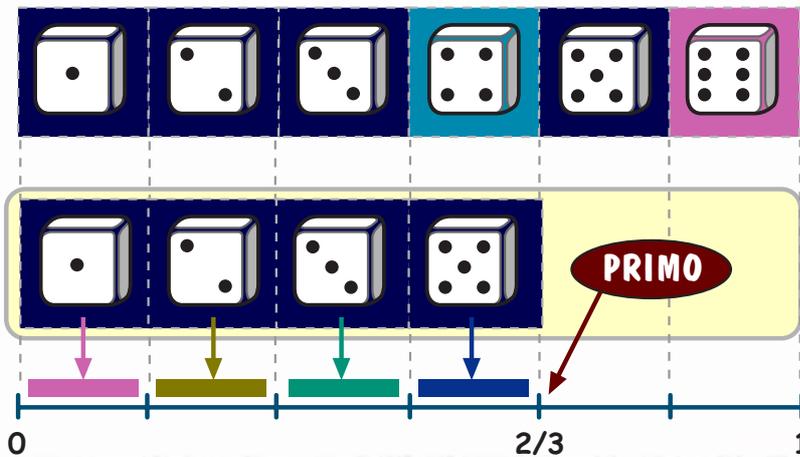
$$S_{\text{par}} = (S_2 \cup S_4 \cup S_6)$$

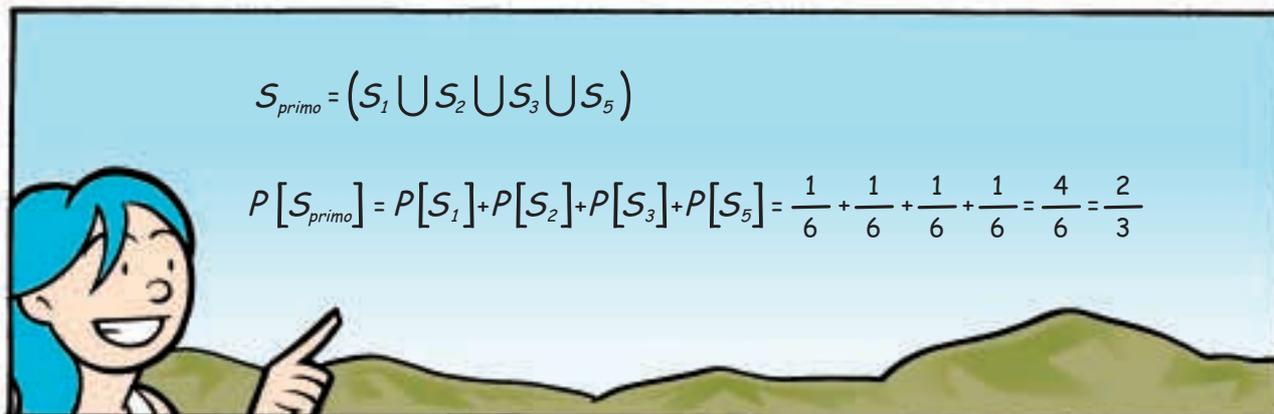
$$P[S_{\text{par}}] = P[S_2] + P[S_4] + P[S_6] = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

VEAMOS UN CASO MÁS COMPLICADITO:



SACAR UN NÚMERO PRIMO AL LANZAR UN DADO.





PAR						
PRIMO						

Y ENTONCES RESULTA:

$$s_{par} = (s_2 \cup s_4 \cup s_6) \quad P[s_{par}] = \frac{1}{2}$$

$$s_{primo} = (s_1 \cup s_2 \cup s_3 \cup s_5) \quad P[s_{primo}] = \frac{2}{3}$$

$$s_{par} \cup s_{primo} = (s_1 \cup s_2 \cup s_3 \cup s_4 \cup s_5 \cup s_6) \quad P[s_{par} \cup s_{primo}] = 1$$

$$P[s_{par} \cup s_{primo}] \neq \frac{1}{2} + \frac{2}{3}$$

$$s_{par} \cap s_{primo} = (s_2) \quad P[s_{par} \cap s_{primo}] = \frac{1}{6}$$

$$P[s_{par} \cap s_{primo}] \neq \frac{1}{2} * \frac{2}{3}$$

ES IMPORTANTE ANOTAR QUE CUANDO LOS SUCESOS NO SON DISJUNTOS...

LA PROBABILIDAD DEL SUCESO UNIÓN NO ES LA SUMA DE LAS PROBABILIDADES.  
LA PROBABILIDAD DEL SUCESO INTERSECCIÓN NO ES EL PRODUCTO DE LAS PROBABILIDADES.



YO QUERÍA QUE VIERAIS LO QUE HE PREPARADO PARA MIS TRANSPARENCIAS, INDICANDO QUE UNA VARIABLE DETERMINISTA, COMO LO ES EL SUELDO DE UN DÍA DE TRABAJO, PUEDE CONVERTIRSE EN ALEATORIA.

## Variables aleatorias

TU SALARIO CONSISTIRÁ EN 20 € FIJOS AL DÍA MÁS 5 € POR HORA TRABAJADA.

**v.d.** SUPONIENDO QUE UN DÍA TRABAJAS 8 HORAS, ¿CUÁL ES EL SALARIO QUE TE CORRESPONDE?

**SALARIO = 20 + 8 X 5 = 60 €**



¿CUÁL ES EN ESTA SEGUNDA SITUACIÓN EL SUELDO?

AÑADAMOS AHORA UNA CONDICIÓN: CADA DÍA SE LANZA UNA MONEDA. SI SALE CARA SE AÑADEN 10 € Y SI SALE CRUZ SE RESTAN 10 €.

**v.a.** → **Salario**

A gold coin is shown with '70 €' on the top half and '50 €' on the bottom half, with a double-headed arrow indicating a range or choice between the two values.

A MÍ ME GUSTARÍA RESALTAR QUE A PESAR DE QUE LOS EJEMPLOS PRIMEROS SON DE JUEGO PARA LAS VARIABLES ALEATORIAS (DADOS, MONEDAS, CARTAS, ETC.), LOS USAMOS POR SU FACILIDAD TANTO PROBABILÍSTICA COMO GRÁFICA.

Illustrations of two dice, two gold coins (one with '55'), and a stack of playing cards including the 4 of diamonds and the Ace of hearts.



## Variables aleatorias





ROJO... SÍ NEGRO... NO	Casa 1ª	Casa 2ª	Casa 3ª	Casa 4ª	Casa 5ª
Color	Roja/Azul	<b>AZUL</b> 🏠	Azul/Verde	Azul	Azul
Nacionalidad	<b>NORUEGO</b> 🇳🇴	Británico/Noruego	Noruego/Danés	Noruego	Noruego
Bebida	Leche/té	Zumo de pomelo Leche/café	<b>LECHE</b> 🥛	Leche	Leche
Profesión	Matemático	Biólogo	Informático	-	-
Mascota	-	-	-	-	-



## CAPÍTULO 5



ANDREI NIKOLAEVICH  
KOLMOGOROV, MOSCÚ (1903-1987)







EJEMPLOS, AHÍ VAN... LA CLASIFICACIÓN DE UNA PELÍCULA:

ABURRIDÍSIMA, ABURRIDA, REGULAR, DIVERTIDA, DESTERNILLANTE.

LA MAYORÍA DE LAS CALIFICACIONES DE ATRIBUTOS:

MALO, REGULAR, BUENO.

PERO YO HE VISTO QUE A VECES SE LE PONEN NÚMEROS A ESTAS VARIABLES.

SÍ, TIENES RAZÓN. PERO ESOS GUARISMOS NO TIENEN VALOR DE MAGNITUD NUMÉRICA.

POR EJEMPLO, PARA INDICAR HOMBRE Y MUJER, ALGUNAS TABLAS MARCAN RESPECTIVAMENTE "0" Y "1", PERO ESTO NO QUIERE DECIR QUE LOS CHICOS NO VALGAN NADA Y LAS CHICAS TENGAN UNA CALIFICACIÓN DE SÓLO UNO.

SON SÓLO SÍMBOLOS, DE FACILIDAD OPERATIVA.

PERO CON ESTAS VARIABLES CUALITATIVAS NOMINALES NO PODREMOS POR EJEMPLO SACAR LA MEDIA, PUES LOS "0" Y "1" SON FICTICIOS.

SERÍA ABSURDO DECIR POR EJEMPLO QUE LA MEDIA DE UN GRUPO DE CHICAS Y CHICOS ES 0'45, QUE NO TIENE NINGÚN SENTIDO.

PERO SÍ QUE PODREMOS HALLAR EL PORCENTAJE DE CADA UNO DE ELLOS.

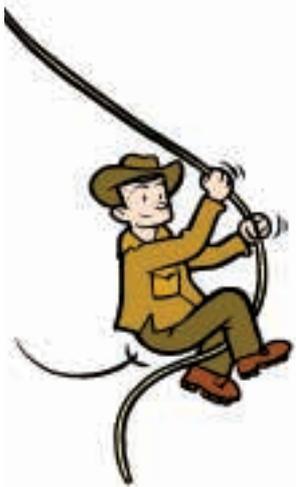


EJERCICIOS.

	¿Te gusta?	Frecuencia absoluta	Frecuencia relativa	Porcentaje
El cine de aventuras	Nada	30	0,1000	10,00%
	Poco	45	0,1500	15,00%
	Regular	65	0,2167	21,67%
	Mucho	90	0,3000	30,00%
	Con locura	70	0,2333	23,33%
	<b>Total =</b>	<b>300</b>	<b>1</b>	<b>100,00%</b>

EN LAS VARIABLES CUALITATIVAS ORDINALES; LOS NÚMEROS QUE SE INDICAN, ME SUPONGO QUE SON DEL MISMO TIPO QUE EN LAS VARIABLES NOMINALES.

LA TALLA CERO EN ROPA DE BEBÉ ES UTILIZADA POR ALGUNAS MARCAS, SIN QUE ELLO TRAIGA COMO CONSECUENCIA QUE NO TE DEN NADA CUANDO LA COMPRES.



	¿Te gusta?	Valores	Frecuencia absoluta	Frecuencia relativa	X*fr
El cine de aventuras	Nada	1	30	0,1000	0,1000
	Poco	2	45	0,1500	0,3000
	Regular	3	65	0,2167	0,6500
	Mucho	4	90	0,3000	1,2000
	Con locura	5	70	0,2333	1,1667
	<b>Total=</b>		<b>300</b>	<b>Media=</b>	<b>3,4167</b>

COMO SE VE, LA MEDIA NO TIENE VALOR SIGNIFICATIVO, NOS DICE QUE PARECEN DECANTARSE POR LAS ÚLTIMAS RESPUESTAS, PERO...



AHORA ME DOY CUENTA DE QUE ESOS NÚMEROS SON FICTICIOS, PUES YO HUBIERA PODIDO ASIGNAR 0; 1; 2; 3; 4, POR EJEMPLO U OTRA CUALQUIERA, CON TAL DE QUE FUERAN DISTINTAS ENTRE SÍ.



HASTA PODÍAS PONER 5, 4, 3, 2, 1; Y ADEMÁS EN LA CLASIFICACIÓN DE ARRIBA "POCO" NO ES EL DOBLE QUE "MUCHO" AUNQUE "4" SEA EL DOBLE QUE "2".





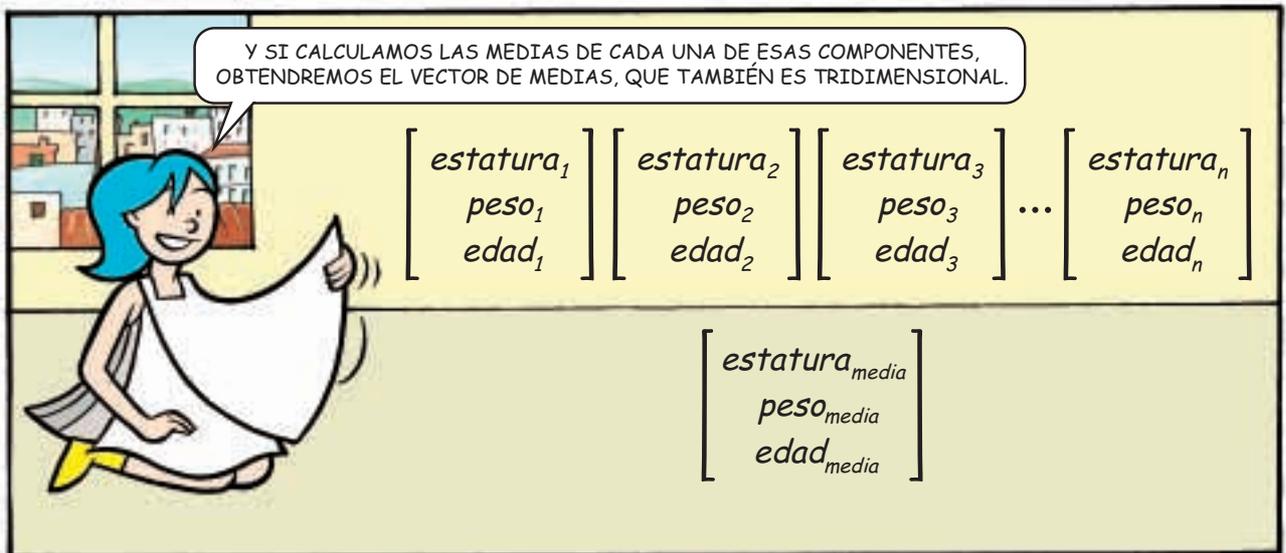


OBSERVACIÓN UNIVARIANTE:  
"LA EDAD".



OBSERVACIÓN MULTIVARIABLE:  
"LA EDAD, EL PESO,  
EL SUELDO SEMANAL,  
EL GÉNERO, SU  
NÚMERO EN LA  
LISTA DE CLASE".





X	f. Absoluta	f. Relativa	X*fr	X al cuadrado	X*X*fr
1.380.000	6	0,20	276.000	1.904.400.000.000	380.880.000.000
1.400.000	9	0,30	420.000	1.960.000.000.000	588.000.000.000
1.480.000	12	0,40	592.000	2.190.400.000.000	876.160.000.000
1.520.000	3	0,10	152.000	2.310.400.000.000	231.040.000.000
	<b>30</b>	<b>Media=</b>	<b>1.440.000</b>	<b>Suma=</b>	<b>2.076.080.000.000</b>

$$Media = \sum x_i fr_i = 1.440.000$$

$$Varianza = \sum x_i^2 fr_i - \bar{x}^2 = 2.480.000.000$$



DIVIDAMOS LOS VALORES DE X POR 10.000 A VER QUÉ PASA.

X	f. Absoluta	f. Relativa	X*fr	X al cuadrado	X*X*fr
138	6	0,20	28	19.044	3.809
140	9	0,30	42	19.600	5.880
148	12	0,40	59	21.904	8.762
152	3	0,10	15	23.104	2.310
	<b>30</b>	<b>Media=</b>	<b>144</b>	<b>Suma=</b>	<b>20.761</b>

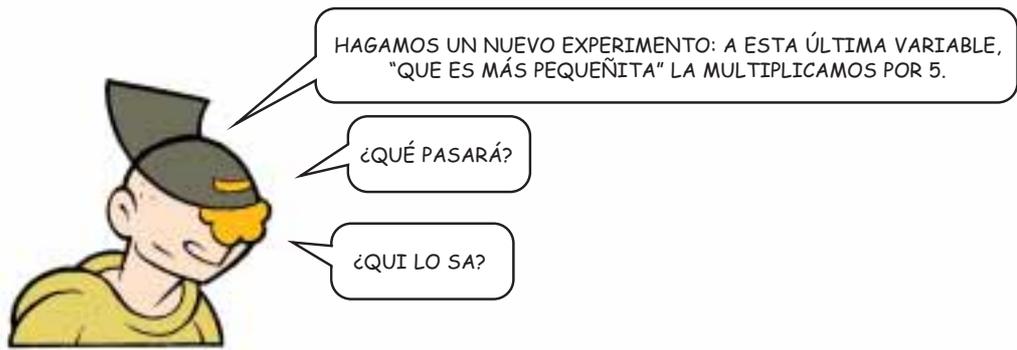
$$Media = \sum x_i fr_i = 144$$

$$Varianza = \sum x_i^2 fr_i - \bar{x}^2 = 24,80$$



HAGAMOS UN CUADRO DE LO QUE HA PASADO:

	Variable antigua	Variable antigua dividida por 10.000
<b>Media</b>	1.440.000	Sale dividida por 10.000 = <b>144</b>
<b>Varianza</b>	2.480.000.000	Sale dividida por 10.000 al cuadrado = <b>24,80</b>



X	f. Absoluta	f. Relativa	X*fr	X al cuadrado	X*X*fr
690	6	0,20	138	476.100	95.220
700	9	0,30	210	490.000	147.000
740	12	0,40	296	547.600	219.040
760	3	0,10	76	577.600	57.760
	<b>30</b>	<b>Media=</b>	<b>720</b>	<b>Suma=</b>	<b>519.020</b>
		<b>Varianza=</b>	<b>620,00</b>		



	Variable anterior	Variable anterior multiplicada por 5
<b>Media</b>	144	Sale multiplicada por 5 = 720
<b>Varianza</b>	24'80	Sale multiplicada por 5 al cuadrado = 620



X	f. Absoluta	f. Relativa	X*fr	X al cuadrado	X*X*fr
715	6	0,20	143	511.225	102.245
725	9	0,30	218	525.625	157.688
765	12	0,40	306	585.225	234.090
785	3	0,10	79	616.225	61.623
	<b>30</b>	<b>Media=</b>	<b>745</b>	<b>Suma=</b>	<b>555.645</b>
		<b>Varianza=</b>	<b>620,00</b>		



EN CAMBIO, AQUÍ LA VARIANZA NO CAMBIA.

	Variable anterior	Variable anterior más 25
<b>Media</b>	720	Sale aumentada en 25 = 745
<b>Varianza</b>	620	Se mantiene, no cambia = 620

VENGA...¡VENGA! Y SI A ESTA ÚLTIMA LE RESTÁRAMOS 100.



X	f. Absoluta	f. Relativa	X*fr	X al cuadrado	X*X*fr
615	6	0,20	123	378.225	75.645
625	9	0,30	188	390.625	117.188
665	12	0,40	266	442.225	176.890
685	3	0,10	69	469.225	46.923
	<b>30</b>	<b>Media=</b>	<b>645</b>	<b>Suma=</b>	<b>416.645</b>
		<b>Varianza=</b>	<b>620,00</b>		

	Variable anterior	Variable anterior menos 100
<b>Media</b>	745	Sale disminuida en 100 = 645
<b>Varianza</b>	620	Se mantiene, no cambia = 620



ROJO... SÍ NEGRO... NO	Casa 1ª	Casa 2ª	Casa 3ª	Casa 4ª	Casa 5ª
Color	AMARILLA 🏠	AZUL 🏠	ROJA 🏠	VERDE 🏠	BLANCA 🏠
Nacionalidad	NORUEGO 🇳🇴	Británico/Noruego Sueco	BRITÁNICO 🇬🇧	Noruego/Británico Danés	Noruego/Británico
Bebida	Té/Leche Chocolate/Café	Café/Leche	LECHE 🥛	CAFÉ ☕	Café/leche
Profesión	BIÓLOGO 🧬	Biólogo/Físico	Informático Matemático Biólogo	Informático Químico/Biólogo	Químico/Biólogo
Mascota	Perro/Pájaro Caballo	CABALLO 🐎	Perro/Caballo	Gato/Caballo	Gato/Caballo

# CAPÍTULO 6



JOHN WILDER TUKEY, NEW BEDFORD  
MASSACHUSETTS (1915-2000)



Media	3.234,2180
Mediana	3.250,0000
Moda	3.200,0000
Mínimo	500,0000
Máximo	4.820,0000
Varianza	250.697,7025
Cuasivarianza	250.984,6511
Desviación estándar	500,6972
Cuasidesviación	500,9478

**PESOS AL NACER**  
1.000 observaciones



YO QUIERO AÑADIR OTRAS, PUES ME SERVIRÁN PARA UNAS EXPERIENCIAS GRÁFICAS QUE HE TRABAJADO.

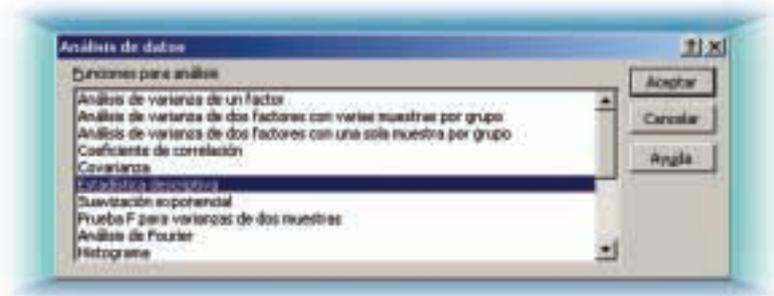
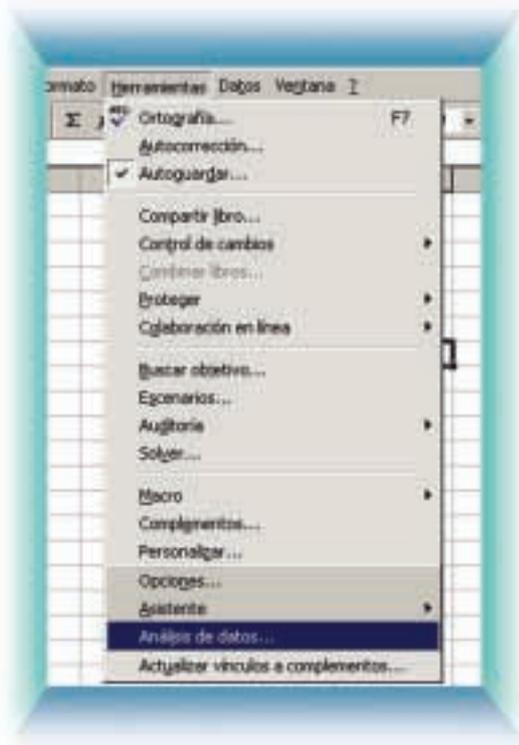
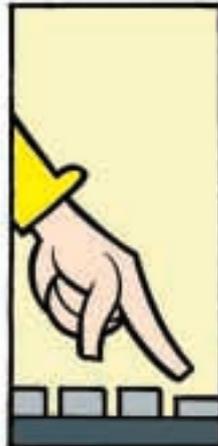
<b>Mediana=</b>	Primer Cuartil=	2950
	Segundo Cuartil=	<b>3250</b>
<b>Recorrido intercuartílico=</b>	Tercer Cuartil=	3570
	Tercer-Primer Cuartil=	<b>620</b>



ANTES DE SEGUIR, PODEMOS RECORDAR QUE EN LA HOJA DE CÁLCULO EXISTE UNA HERRAMIENTA, QUE OBTIENE DIRECTA Y CONJUNTAMENTE MUCHAS DE ESTAS MEDIDAS.



CREO QUE ES EN ANÁLISIS DE DATOS.



PESOS AL NACER



Media	3234,218
Error típico	15,84135888
Mediana	3250
Moda	3200
Desviación estándar	500,9477529
Varianza de la muestra	250948,6511
Curtosis	2,203375205
Coefficiente de asimetría	-0,612646309
Rango	4320
Mínimo	500
Máximo	4820
Suma	3234218
Cuenta	1000



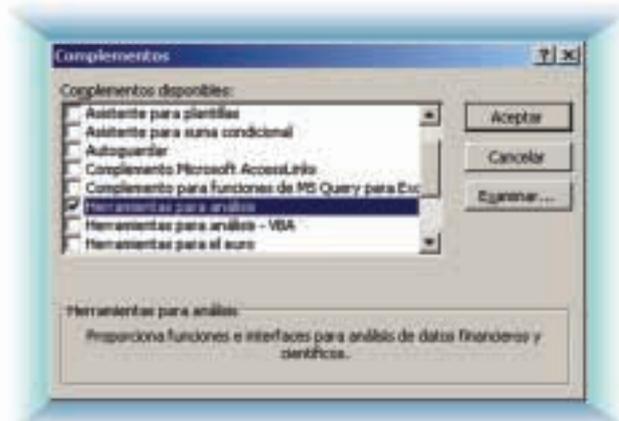
TENEMOS QUE OBSERVAR AQUÍ QUE EL ERROR TÍPICO ES LA DESVIACIÓN ESTÁNDAR DE LA MEDIA MUESTRAL (DISTRIBUCIÓN MUESTRAL DE MEDIAS); LA VARIANZA DE LA MUESTRA ES LA CUASIVARIANZA O VARIANZA CORREGIDA....



DEBO SER GAFFE, PUES NO SÓLO NO HE ENTENDIDO ESO DE LA DISTRIBUCIÓN MONSTRUO DE ALGO....., SINO QUE NI SIQUIERA SALE EN MI ORDENADOR ESO DE ANÁLISIS DE DATOS.



TIENES UNA SUERTE... LO QUE PASA ES QUE NO LO TIENES HABILITADO; HAZ LO QUE TE VOY A INDICAR Y VERÁS COMO PUEDES TRABAJAR...



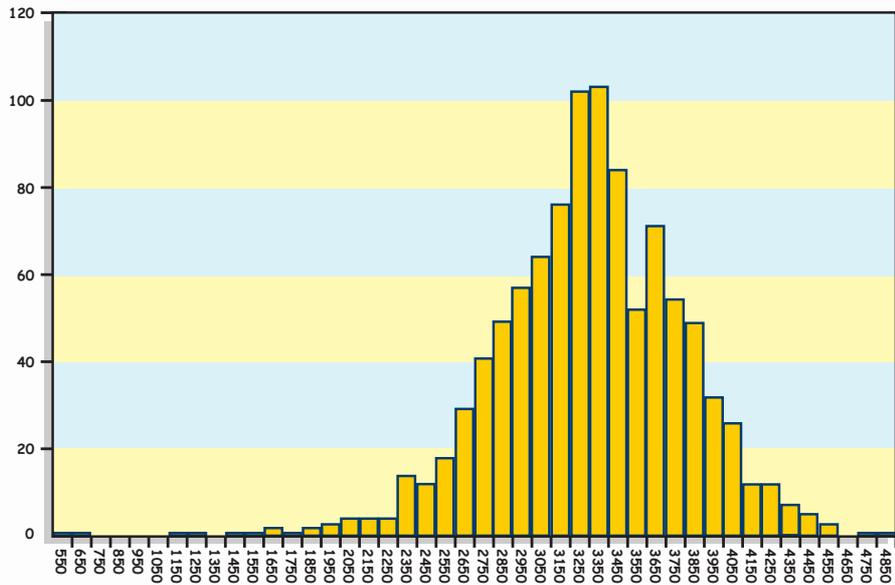


CLASES			
Inferior (incluido)	Superior (excluido)	MARCA	Frecuencia
500	600	550	1
600	700	650	1
700	800	750	0
800	900	850	0
900	1000	950	0
1000	1100	1050	0
1100	1200	1150	1
1200	1300	1250	1
1300	1400	1350	0
1400	1500	1450	1
1500	1600	1550	1
1600	1700	1650	2
1700	1800	1750	1
1800	1900	1850	2
1900	2000	1950	3
2000	2100	2050	4
2100	2200	2150	4
2200	2300	2250	4
2300	2400	2350	14
2400	2500	2450	12
2500	2600	2550	18
2600	2700	2650	29
2700	2800	2750	41
2800	2900	2850	49
2900	3000	2950	57
3000	3100	3050	64
3100	3200	3150	76
3200	3300	3250	102
3300	3400	3350	103
3400	3500	3450	84
3500	3600	3550	52
3600	3700	3650	71
3700	3800	3750	54
3800	3900	3850	49
3900	4000	3950	32
4000	4100	4050	26
4100	4200	4150	12
4200	4300	4250	12
4300	4400	4350	7
4400	4500	4450	5
4500	4600	4550	3
4600	4700	4650	0
4700	4800	4750	1
4800	4900	4850	1

CUYA REPRESENTACIÓN GRÁFICA SERÁ:



FRECUENCIA

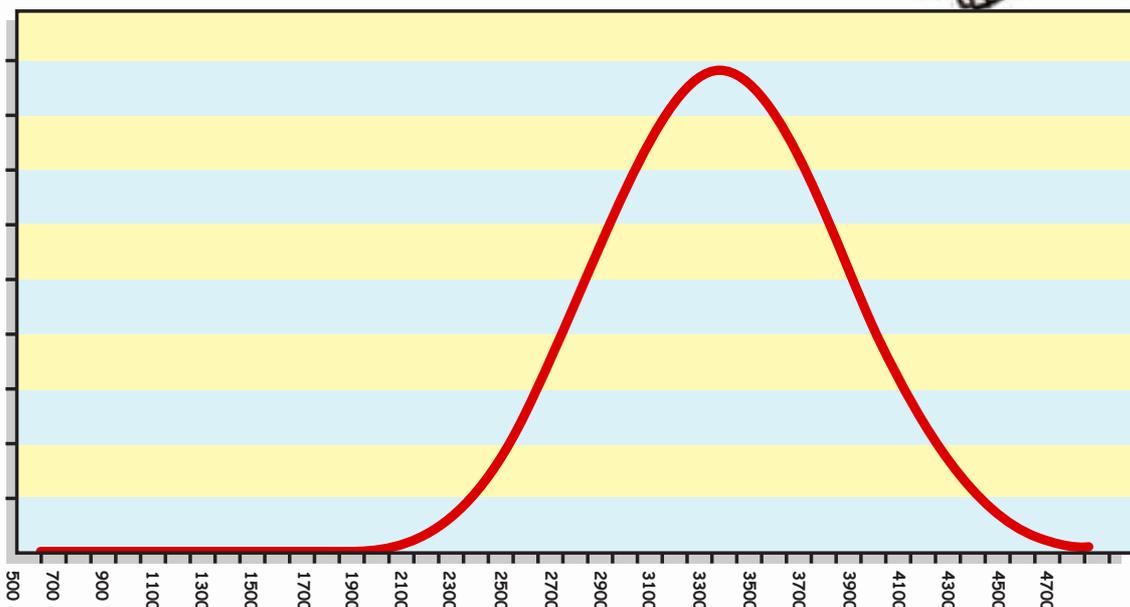


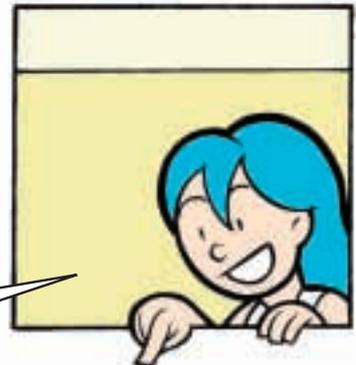
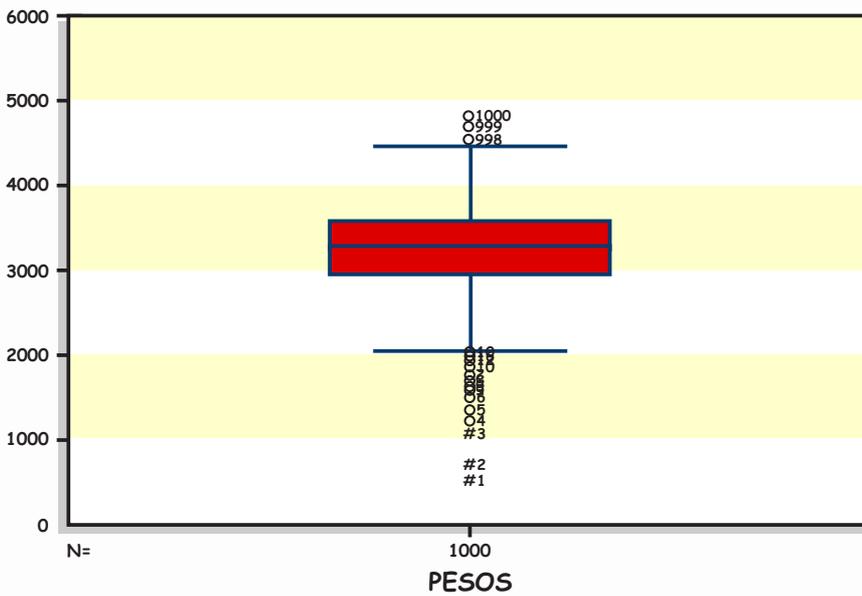
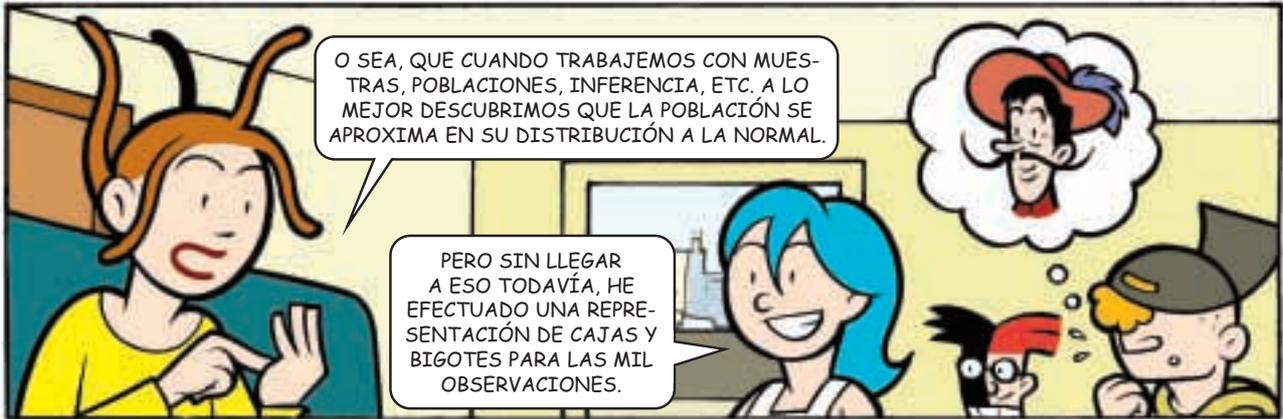
SI HACEMOS LOS CÁLCULOS DE LA MEDIA, POR CLASES, VEMOS QUE NOS SALE UNA APROXIMACIÓN. COMPRENIBLE, PUES ANTES A CADA VALOR LE DÁBAMOS SU VERDADERA MAGNITUD Y AHORA SIEMPRE LE DAMOS EL DE SU MARCA DE CLASE.

**Media=** 3.234,2180      **Por clases=** 3296,9000

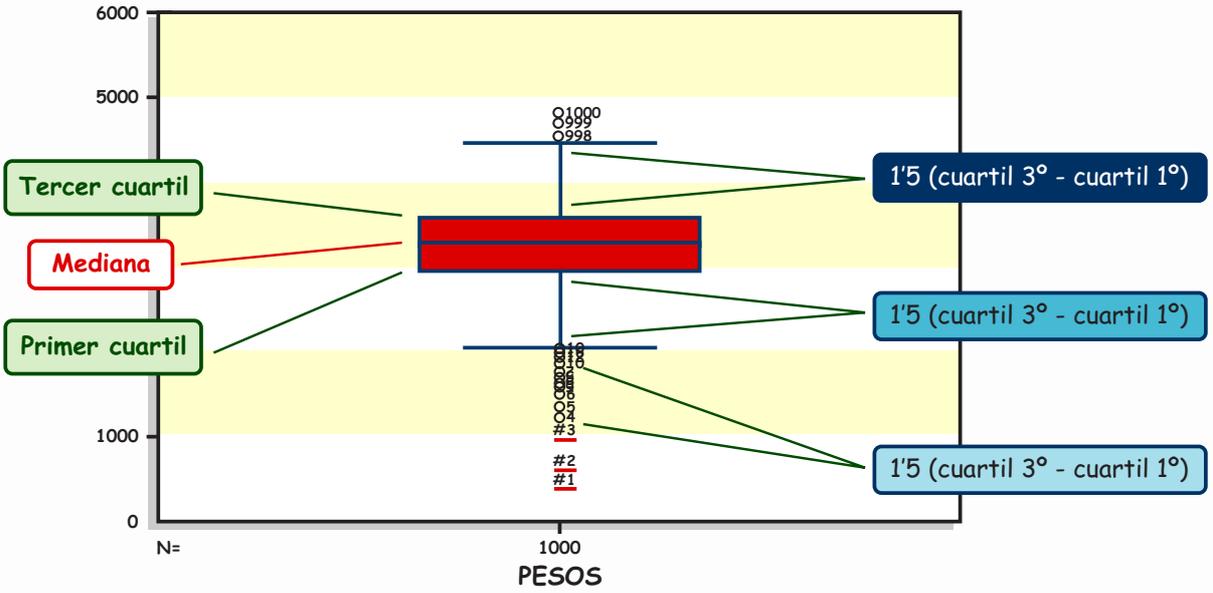


SERÍA CONVENIENTE SOÑAR QUE SI EN VEZ DE HABER TOMADO 1000 OBSERVACIONES, HUBIÉRAMOS TOMADO "TROPECIENTAS-MIL" LA FIGURA IRÍA TOMANDO LA SIGUIENTE FORMA:





BUENO, PONGO OTRA TRANSPARENCIA DE LAS MÍAS, Y DESPUÉS CUENTO CÓMO LO HE ELABORADO.



PASOS A SEGUIR: 1º/ ORDENO LAS OBSERVACIONES DE MENOR A MAYOR.



	A	B	979	979	4200
1	1	600	980	980	4200
2	2	648	981	981	4200
3	3	1056	982	982	4200
4	4	1224	983	983	4250
5	5	1408	984	984	4250
6	6	1470	985	985	4300
7	7	1600	986	986	4300
8	8	1685	987	987	4319
9	9	1700	988	988	4320
10	10	1800	989	989	4320
11	11	1810	990	990	4360
12	12	1860	991	991	4360
13	13	1900	992	992	4360
14	14	1900	993	993	4360
15	15	1980	994	994	4365
16	16	2000	995	995	4470
17	17	2040	996	996	4500
18	18	2040	997	997	4500
19	19	2050	998	998	4700
20	20	2080	999	999	4700
			1000	1000	4820

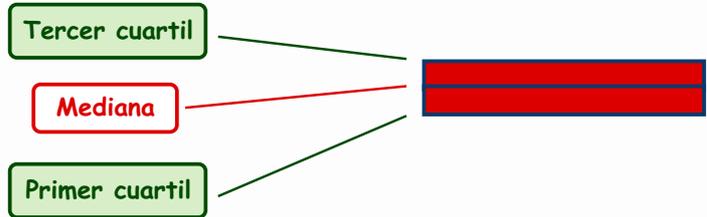
2º/ CALCULO LA MEDIANA Y LOS CUARTILES PRIMERO Y TERCERO



Primer Cuartil=	2950
<b>Mediana=</b>	<b>3250</b>
Tercer Cuartil=	3570



CON ELLO, PUEDO CONSTRUIR LA CAJA:



3º/ CALCULO EL RECORRIDO INTERCUARTÍLICO.

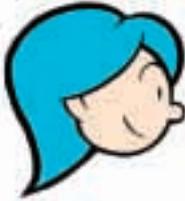
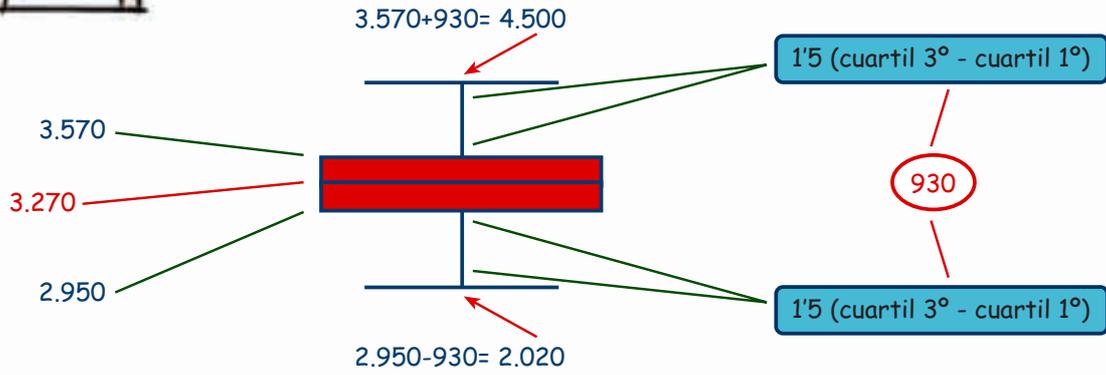


<b>Recorrido intercuartílico=</b>	Tercer-Primer Cuartil=	<b>620</b>
-----------------------------------	------------------------	------------



4º/ ESTA CANTIDAD LA MULTIPLICO POR 1,5.

$620 \text{ por } 1,5 = 930$

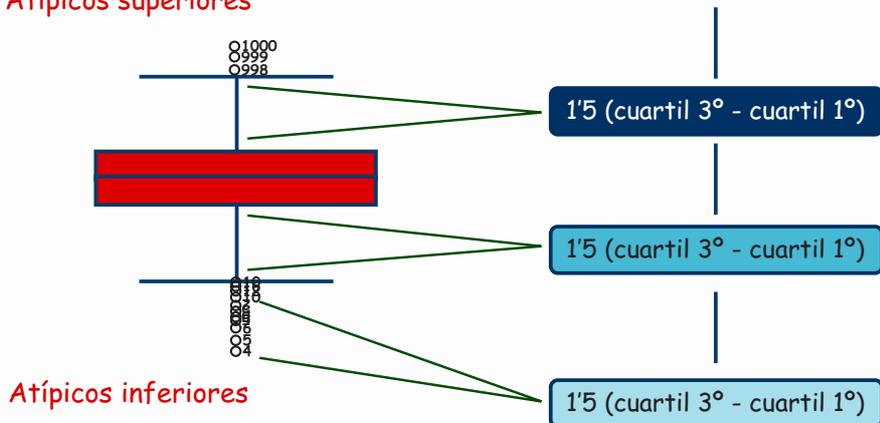


5º/ SE REPRESENTAN CON UN CIRCULITO TODOS Y CADA UNO DE LOS VALORES:

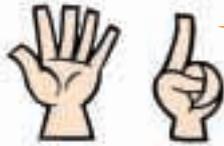
- A/ QUE SE ENCUENTREN ENTRE EL LÍMITE SUPERIOR DEL BIGOTE DE ARRIBA Y "SU VALOR MÁS (OTRA VEZ) 930".
- B/ QUE SE ENCUENTREN ENTRE EL LÍMITE INFERIOR DEL BIGOTE DE ABAJO Y "SU VALOR MENOS 930".

A ESTOS VALORES LES LLAMAREMOS ATÍPICOS POR ARRIBA Y ATÍPICOS POR DEBAJO RESPECTIVAMENTE.

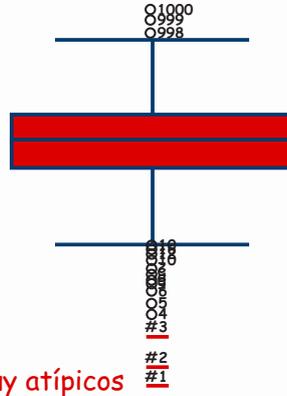
Atípicos superiores



Atípicos inferiores



6º/ SE DIBUJARÁN CON UNA CRUZ O UN SIGNO DISTINTO AL CÍRCULO, TODOS Y CADA UNO DE LOS QUE SUPEREN O SEAN INFERIORES A LOS YA INDICADOS COMO ATÍPICOS, Y LES LLAMAREMOS "MUY ATÍPICOS".



Muy atípicos

UNA OBSERVACIÓN: LOS BIGOTES PUEDEN SER MÁS CORTOS, EN EL CASO DE LOS VALORES INFERIOR O SUPERIOR, BIEN POR ABAJO, POR ARRIBA, BIEN POR AMBOS LADOS, SEAN MAYOR, MENOR RESPECTIVAMENTE QUE EL EXTREMO DEL BIGOTE, QUE NUNCA ESTARÁN FUERA DE LOS LÍMITES QUE MARCARÍAN LOS VALORES MÍNIMO Y MÁXIMO DE LAS OBSERVACIONES.

ASÍ SE VE EN NUESTRO CASO QUE COMO HEMOS ORDENADO LOS PESOS: EL PRIMERO, EL SEGUNDO Y TERCERO SON MUY ATÍPICOS; Y POR EJEMPLO LOS DE LUGAR 998, 999 Y 1000 SON ATÍPICOS SUPERIORES.



1	500	Muy atípicos
2	648	
3	1056	
4	1224	ATÍPICOS
5	1408	
6	1470	
7	1600	
8	1605	
9	1700	
10	1800	
11	1810	
12	1860	
13	1900	
14	1900	
15	1990	
16	2000	
	2020	Bigote
	2950	Caja
	3250	
	3570	
	4500	Bigote
997	4500	
998	4500	
999	4700	
1000	4820	



DE ESTE GRÁFICO SE PUEDEN OBTENER MUCHAS HIPÓTESIS SOBRE LAS OBSERVACIONES, QUE COMO SIEMPRE DEBEMOS CONFIRMAR NUMÉRICAMENTE.

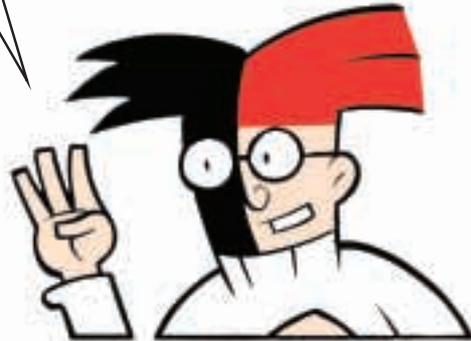


Año	Total	Hombres	Mujeres
1975	11116	5872	5244
1976	11169	5820	5349
1977	10424	5363	5061
1978	10041	5189	4852
1979	10120	5236	4884
1980	9822	4991	4831
1981	9350	4797	4553
1982	9154	4961	4193
1983	8952	4838	4114
1984	8865	4722	4143
1985	8961	4735	4226
1986	8724	4560	4164
1987	8592	4357	4235
1988	8730	4535	4195
1989	8873	4610	4263
1990	8799	4594	4205
1991	8602	4510	4092
1992	8470	4364	4106
1993	7895	3999	3896
1994	7686	3976	3710
1995	7693	3911	3782
1996	7787	3991	3796
1997	8173	4245	3928
1998	8305	4322	3983
1999	8848	4558	4290
2000	9503	4888	4615
2001	9858	4995	4863
2002	10420	5382	5038
2003	10655	5420	5235

AUNQUE EN SU MOMENTO SE ESTUDIARÁN LAS SERIES TEMPORALES, NOSOTROS PODEMOS YA REALIZAR ALGUNA QUE OTRA COSILLA.

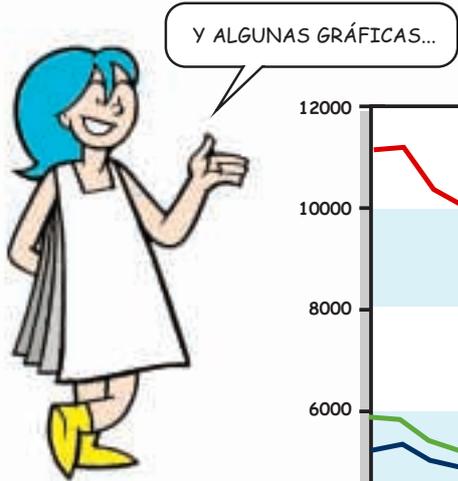


POR LO PRONTO, PODEMOS OBTENER DE CADA UNA DE LAS TRES SERIES UN RESUMEN DE ESTADÍSTICA DESCRIPTIVA.

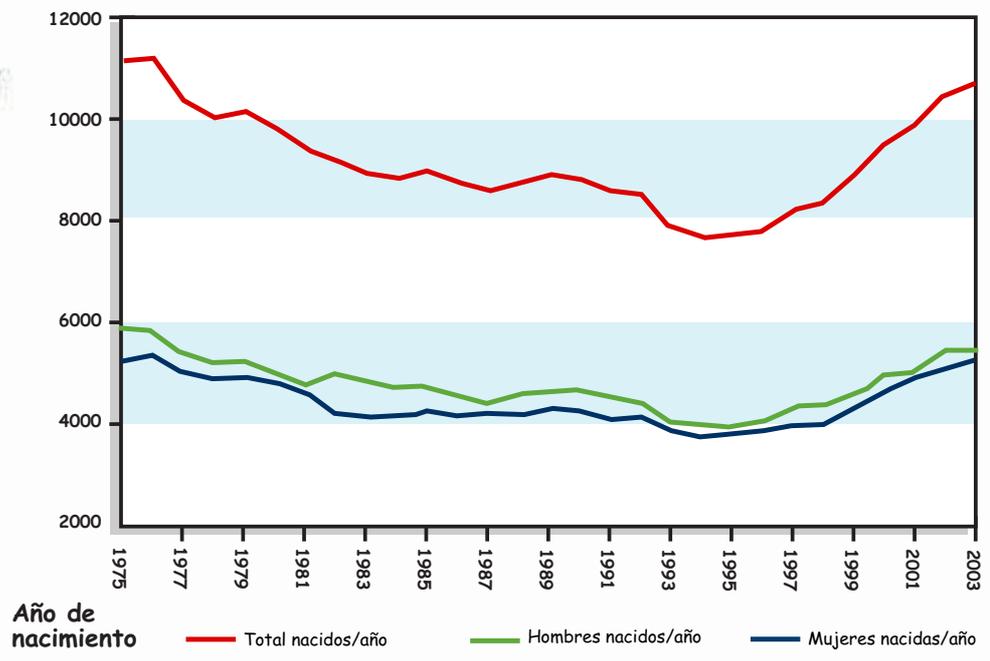


		Total nacidos/año	Hombres nacidos/año	Mujeres nacidas/año
Estadístico	Nº años	29	29	29
	Rango	3483	1961	1639
	Mínimo	7686	3911	3710
	Máximo	11169	5872	5349
	Media	9158,17	4749,69	4408,48
	Desv. típ.	991,74	522,87	482,92
	Varianza	983555,576	273388,650	233214,401
Error típico	Asimetría	,464	,371	,539
	Curtosis	-,584	-,295	-,932
	Media	184,16	97,09	89,68
	Asimetría	,434	,434	,434
	Curtosis	,845	,845	,845

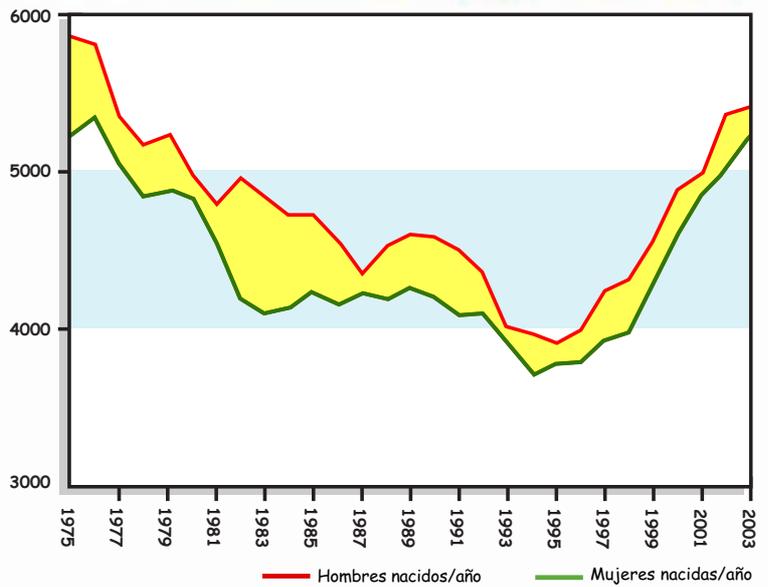
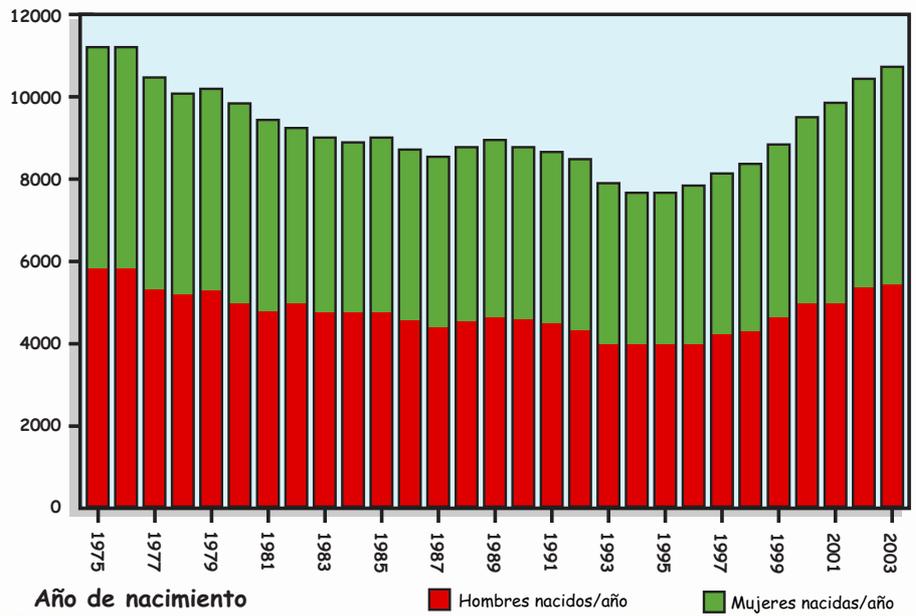




Y ALGUNAS GRÁFICAS...



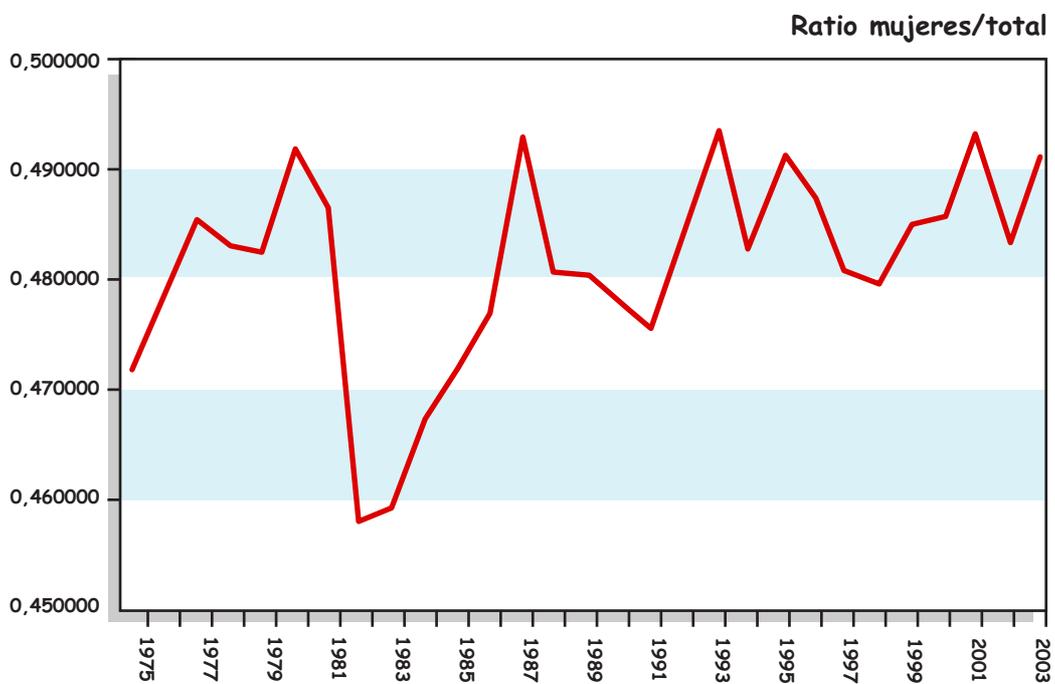
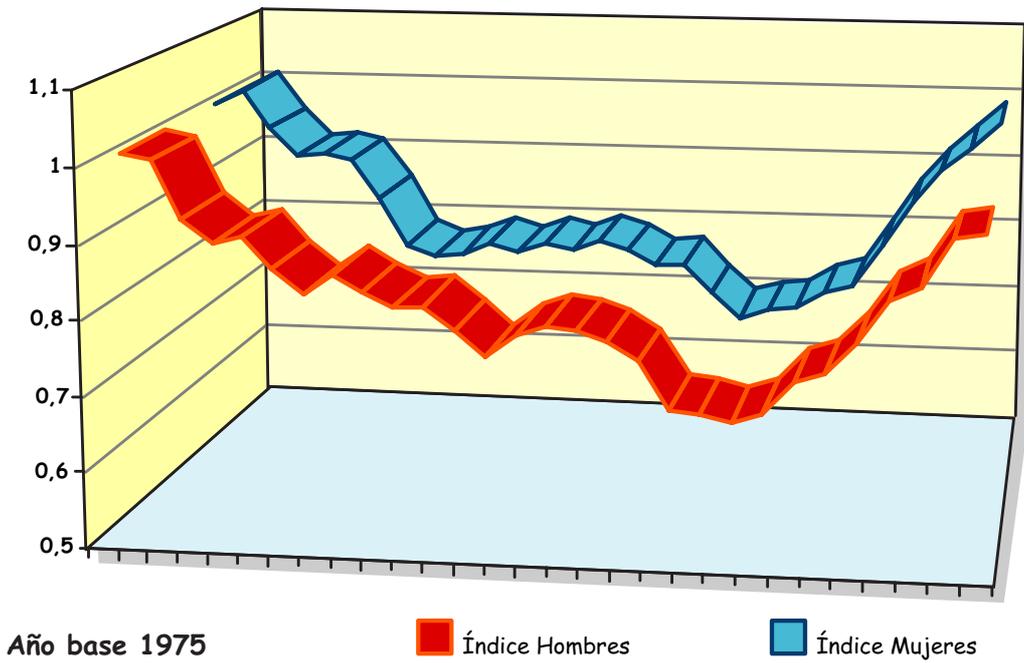
PUES SÍ QUE SE PUEDEN HACER EXPERIENCIAS, PERO VEAMOS OTRA GRÁFICA, QUIERO VER SI LA DESCUBRO.



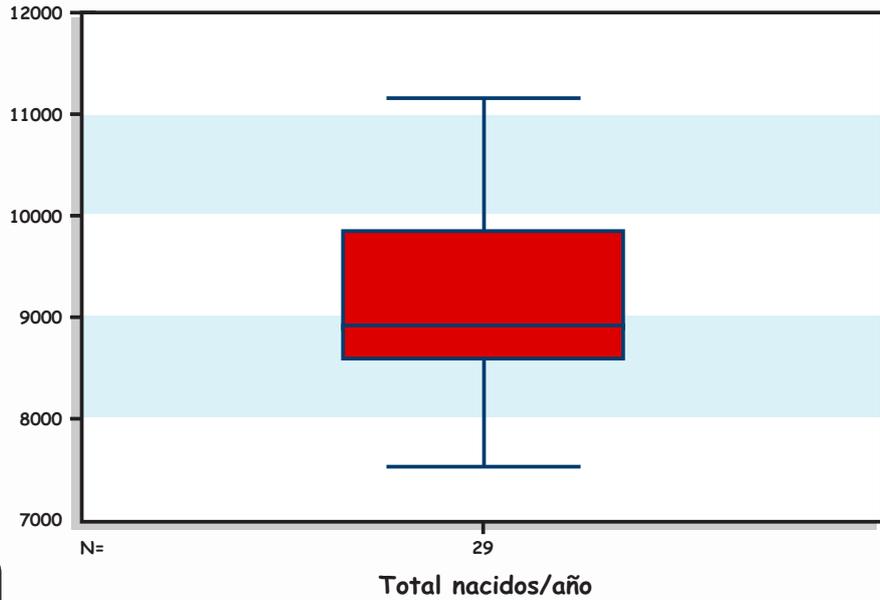


HE REALIZADO UNOS ÍNDICES SIMPLES Y ADEMÁS UN RATIO (O COCIENTE ENTRE LAS MUJERES NACIDAS CADA AÑO Y EL TOTAL DE NACIDOS) QUE QUIERO QUE VEÁIS.

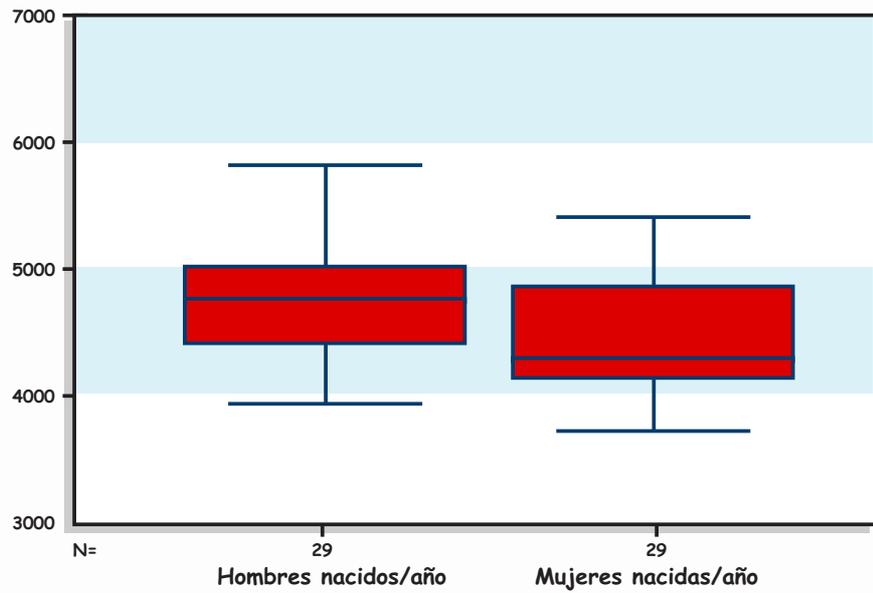
Año	Índices Año base 1975			Ratio Muj./Total
	Índice del total	Índice hombres	Índice mujeres	
1975	1	1	1	0,471752
1976	1,004768	0,991144	1,020023	0,478915
1977	0,937747	0,913317	0,965103	0,485514
1978	0,903293	0,883685	0,925248	0,483219
1979	0,910399	0,891689	0,931350	0,482609
1980	0,883591	0,849966	0,921243	0,491855
1981	0,841130	0,816928	0,868230	0,486952
1982	0,823498	0,844857	0,799580	0,458051
1983	0,805326	0,823910	0,784516	0,459562
1984	0,797499	0,804155	0,790046	0,467343
1985	0,806135	0,806369	0,805873	0,471599
1986	0,784815	0,776567	0,794050	0,477304
1987	0,772940	0,741996	0,807590	0,492900
1988	0,785354	0,772309	0,799962	0,480527
1989	0,798219	0,785082	0,812929	0,480446
1990	0,791562	0,782357	0,801869	0,477895
1991	0,773840	0,768052	0,780320	0,475703
1992	0,761965	0,743188	0,782990	0,484770
1993	0,710237	0,681029	0,742944	0,493477
1994	0,691436	0,677112	0,707475	0,482696
1995	0,692065	0,666042	0,721205	0,491616
1996	0,700522	0,679666	0,723875	0,487479
1997	0,735246	0,722922	0,749047	0,480607
1998	0,747121	0,736035	0,759535	0,479591
1999	0,795970	0,776226	0,818078	0,484855
2000	0,854894	0,832425	0,880053	0,485636
2001	0,886830	0,850647	0,927346	0,493305
2002	0,937388	0,916553	0,960717	0,483493
2003	0,958528	0,923025	0,998284	0,491319



YO VUELVO A LO MÍO.



ES INTERESANTE SACAR ALGUNAS PRECONCLUSIONES, PERO ES MUCHO MEJOR HACERLO SOBRE LOS DIAGRAMAS DE CAJAS Y BIGOTES DE CHICOS Y CHICAS.



OBSERVEMOS QUE NO HA HABIDO UN AÑO EN QUE LA CIFRA DE NACIDOS/AS HAYA SIDO ATÍPICO.



EL 50% INTERMEDIO ES MÁS AMPLIO EN LAS MUJERES QUE EN LOS HOMBRES. LA CAJA DE CHICAS ES MÁS ALTA QUE LA DE CHICOS, AUNQUE ESTÉ MÁS BAJA.





ROJO... SÍ NEGRO... NO	Casa 1ª	Casa 2ª	Casa 3ª	Casa 4ª	Casa 5ª
Color	AMARILLA 	AZUL 	ROJA 	VERDE 	BLANCA 
Nacionalidad	NORUEGO 	DANÉS 	BRITÁNICO 	ALEMÁN 	SUECO 
Bebida	AGUA 	TÉ 	LECHE 	CAFÉ 	ZUMO 
Profesión	BIÓLOGO 	QUÍMICO 	FÍSICO 	MATEMÁTICO 	INFORMÁTICO 
Mascota	GATO 	CABALLO 	PÁJARO 	PEZ 	PERRO 



¡HASTA EL PRÓXIMO CURSO!



FIN





Govern  
de les Illes Balears

**ibae**  
INSTITUT BALEAR  
D'ESTADÍSTICA

Conselleria d'Economia, Hisenda i Innovació > Direcció General d'Economia