LA ESTADÍSTICA EN CÓMIC

LARRY GONICK
Y WOOLLCOTT SMITH

editorial
Zendrera Zariquiey

♦ CONTENIDO ♦

♦ Capítulo 1 ♦
¿QUÉ ES LA ESTADÍSTICA?

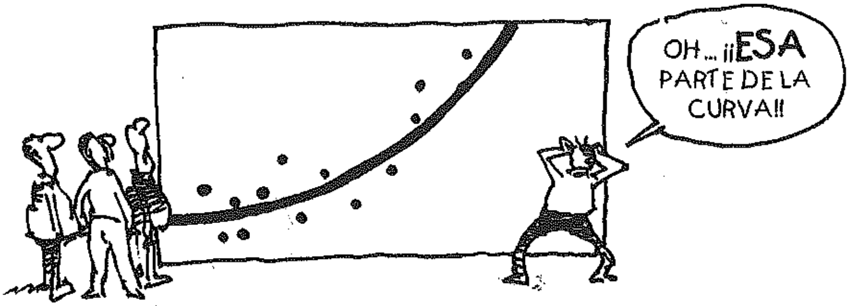
VAGAMOS POR LA VIDA TOMANDO DECISIONES BASADAS EN UNA INFORMACIÓN INCOMPLETA...

LA MAYORÍA DE NOSOTROS VIVIMOS CÓMODOS CON CIERTO NIVEL DE INCERTIDUMBRE.

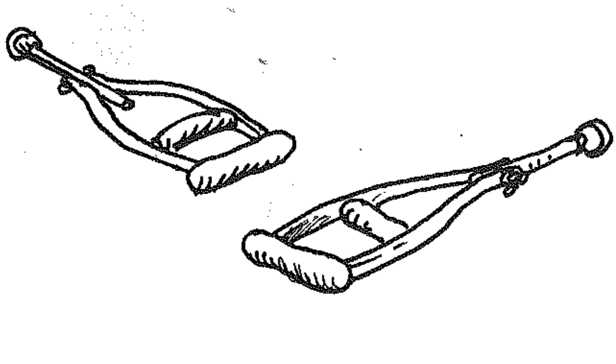
LO ESPECIAL DE LA ESTADÍSTICA, PARA SER PRECISOS, ES SU HABILIDAD DE **CUANTIFICAR** LA INCERTIDUMBRE. ESTO PERMITE A LOS ESTADÍSTICOS HACER **AFIRMACIONES CATEGÓRICAS** CON UNA SEGURIDAD TOTAL SOBRE EL NIVEL DE INCERTIDUMBRE.

¡NO ES SÓLO CUESTIÓN DE PEDIR UNA SOPA! LA ESTADÍSTICA TAMBIÉN TRATA ASUNTOS DE **VIDA O MUERTE**...

POR EJEMPLO, EN 1986, LA LANZADERA ESPACIAL **CHALLENGER** EXPLOTÓ CON SIETE ASTRONAUTAS DENTRO. LA DECISIÓN DE LANZAR LA NAVE A UNA TEMPERATURA DE −2 ºC SE TOMÓ SIN REALIZAR UN SIMPLE ANÁLISIS SOBRE LA FIABILIDAD DE LOS DATOS A BAJAS TEMPERATURAS.



UN EJEMPLO MÁS POSITIVO ES EL DE LA **VACUNA DE SALK CONTRA LA POLIO**. EN 1954, SE PROBÓ LA VACUNA EN 400.000 NIÑOS CON UN RIGUROSO CONTROL PARA EVITAR RESULTADOS SESGADOS. UN BUEN ANÁLISIS ESTADÍSTICO DE LOS RESULTADOS ESTABLECIÓ LA EFICACIA DE LA VACUNA, Y ACTUALMENTE LA POLIO ESTÁ CASI ERRADICADA.



PARA REALIZAR ESTAS HAZAÑAS DE PRESTIDIGITACIÓN MATEMÁTICA, LOS ESTADÍSTICOS SE BASAN EN TRES DISCIPLINAS QUE ESTÁN ESTRECHAMENTE RELACIONADAS:

El análisis de datos,
LA RECOPILACIÓN, ORGANIZACIÓN Y RESUMEN DE LOS DATOS;

La probabilidad,
LAS LEYES DEL AZAR DENTRO Y FUERA DEL CASINO;

La inferencia estadística,
LA CIENCIA QUE EXTRAE CONCLUSIONES ESTADÍSTICAS A PARTIR DE DATOS CONCRETOS BASÁNDOSE EN EL CÁLCULO DE PROBABILIDADES.

EN ESTE LIBRO, TRATAREMOS LAS TRES DISCIPLINAS Y LAS VEREMOS APLICADAS A UNA AMPLIA VARIEDAD DE SITUACIONES DEL MUNDO ACTUAL EN LAS QUE LA ESTADÍSTICA JUEGA UN PAPEL CLAVE.


EN EL CAPÍTULO 2 VEREMOS UN SIMPLE CONJUNTO DE DATOS, EL PESO DE UN GRUPO DE ESTUDIANTES UNIVERSITARIOS.


EN EL CAPÍTULO 3 ESTUDIAREMOS LAS LEYES DE LA PROBABILIDAD EN SU LUGAR DE NACIMIENTO, UN ANTRO DE JUEGO.
± 3% EN EL SIGLO XVII


LOS CAPÍTULOS 4 Y 5 ENSEÑAN A DESCRIBIR EL MUNDO CON **MODELOS DE PROBABILIDAD** UTILIZANDO EL CONCEPTO DE **VARIABLE ALEATORIA**.
¡VAYA! ¡QUÉ SÍMBOLO!

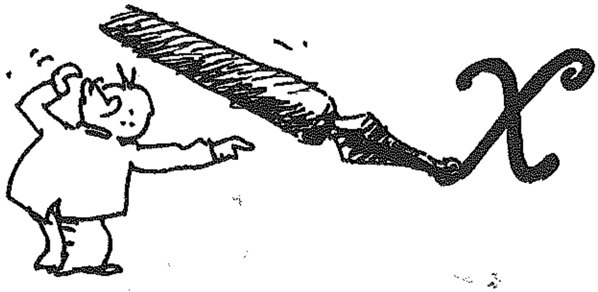

EL CAPÍTULO 6 PRESENTA UNO DE LOS PROCEDIMIENTOS ESENCIALES DE LA ESTADÍSTICA, TOMAR **MUESTRAS** DE UNA POBLACIÓN.


A PARTIR DEL CAPÍTULO 7 MOSTRAMOS CÓMO HACER INFERENCIA ESTADÍSTICA EN CAMPOS TAN COTIDIANOS COMO LOS **SONDEOS DE OPINIÓN, EL CONTROL DE CALIDAD INDUSTRIAL, LAS PRUEBAS MÉDICAS, LOS PROGRAMAS DE SEGUIMIENTO PARA LA PROTECCIÓN MEDIOAMBIENTAL, LA DISCRIMINACIÓN RACIAL Y LA LEY.**
EN RESUMEN, ¡EN TODAS PARTES!

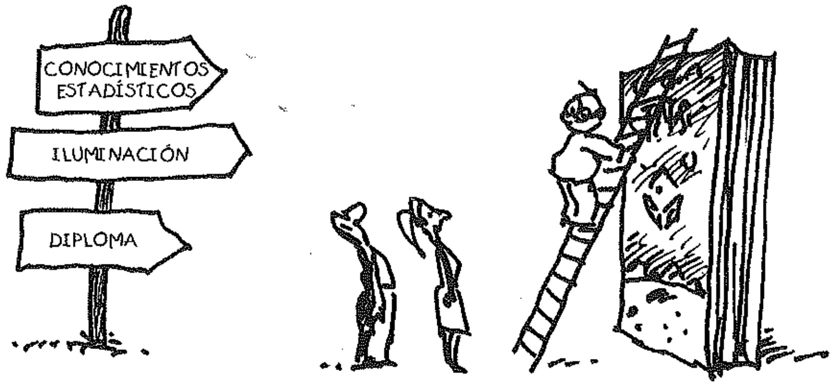
POR ÚLTIMO, CUANDO SE HABLA DE ESTA DISCIPLINA, RESULTA DIFÍCIL NO MENCIONAR ALGO MÁS: LA AMPLIA **DESCONFIANZA** EN LA ESTADÍSTICA DEL MUNDO ACTUAL. NO HAY QUIEN NO HAYA OÍDO HABLAR DE «ESTADÍSTICAS AMAÑADAS». Y EN LA VIDA COTIDIANA ES CASI IMPOSIBLE ENCONTRAR BUENOS ANÁLISIS ESTADÍSTICOS. ¡QUÉ LE VAMOS A HACER!*

*EL LIBRO «LYING WITH STATISTICS» ES MUY POPULAR EN LOS EE.UU.; EN ÉSTE SE DA CUENTA DE FORMAS FRAUDULENTAS DE USAR LA ESTADÍSTICA. [N.T.]

NUESTRA HUMILDE OPINIÓN ES QUE NO SERÍA MALA IDEA **APRENDER UN POCO MÁS SOBRE EL TEMA**... Y POR ESO HEMOS ESCRITO ESTE LIBRO.



EN LOS SIGUIENTES CAPÍTULOS INTENTAMOS PRESENTAR LOS ELEMENTOS DE LA ESTADÍSTICA DE LA FORMA MÁS GRÁFICA E INTUITIVA POSIBLE. LO ÚNICO QUE NECESITAS PARA LEER ESTE LIBRO ES UN POCO DE **PACIENCIA**, ALGO DE **RAZONAMIENTO**, Y CIERTA TOLERANCIA AL **ÁLGEBRA**. O COMO MÍNIMO, ¡AL MENOS **UNO** DE ESTOS REQUISITOS!



♦ Capítulo 2 ♦
ESTADÍSTICA DESCRIPTIVA

LOS **DATOS** SON LA MATERIA PRIMA DE LOS ESTADÍSTICOS, LOS NÚMEROS QUE UTILIZAMOS PARA INTERPRETAR LA REALIDAD. EN TODO PROBLEMA ESTADÍSTICO HAY QUE RECOPILAR, DESCRIBIR Y ANALIZAR DATOS, O AL MENOS **PENSAR** EN LA RECOPILACIÓN, LA DESCRIPCIÓN Y EL ANÁLISIS DE LOS MISMOS.

ESTE CAPÍTULO SE CENTRA EN LA **DESCRIPCIÓN** DE DATOS. ¿CÓMO PODEMOS REPRESENTARLOS DE FORMA ÚTIL? ¿CÓMO DESCUBRIR LAS ESTRUCTURAS INTERNAS DE UN MONTÓN DE NÚMEROS DESNUDOS? ¿CÓMO SE PUEDE RESUMIR LA FORMA BÁSICA DE LOS DATOS?

BIEN, PARA DESCRIBIR DATOS, LO PRIMERO QUE NECESITAMOS SON DATOS QUE DESCRIBIR. ASÍ QUE, ¡VAMOS A RECOPILAR UNOS CUANTOS!

AQUÍ TENEMOS UNOS DATOS REALES: COMO PARTE DE UN EXPERIMENTO DE CLASE, SE RECOGIERON LOS PESOS DE 92 ESTUDIANTES DE PENNSYLVANIA. LOS RESULTADOS FUERON ÉSTOS:*

*EL PESO ESTÁ DADO EN LIBRAS. 1 LIBRA = 0.454 KG. [N.T.]

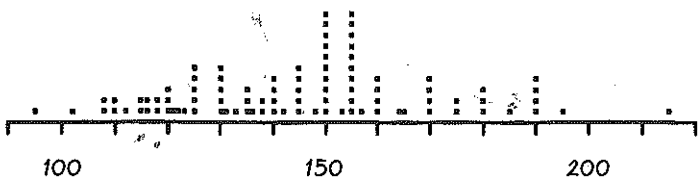
HOMBRES

140 145 160 190 155 165 150 190 195 138 160 155 153 145 170 175 175 170 180 135
170 157 130 185 190 155 170 155 215 150 145 155 155 150 155 150 180 160 135 160
130 155 150 148 155 150 140 180 190 145 150 164 140 142 136 123 155

MUJERES

140 120 130 138 121 125 116 145 150 112 125 130 120 130 131 120 118 125 135 125
118 122 115 102 115 150 110 116 108 95 125 133 110 150 108

PARA EMPEZAR, DIBUJAMOS UN **DIAGRAMA DE PUNTOS**: UN PUNTO POR CADA ESTUDIANTE SOBRE SU PESO CORRESPONDIENTE:



PESO EN LIBRAS

AQUÍ ENCONTRAMOS UN **PROBLEMA**: LA CONCENTRACIÓN EN ***150* Y *155*** LIBRAS. LOS ESTUDIANTES SUELEN DECIR SU PESO EN **MÚLTIPLOS DE CINCO LIBRAS**. EN SITUACIONES REALES COMO ÉSTA, REDONDEAR ASÍ LOS DATOS PUEDE LLEGAR A OCULTAR LAS ESTRUCTURAS GENERALES DE UN CONJUNTO DE DATOS... PERO, DE MOMENTO, LO PASAREMOS POR ALTO.

LOS DATOS SE PUEDEN RESUMIR EN UNA **TABLA DE FRECUENCIAS**. DIVIDIMOS LA LÍNEA DE NÚMEROS EN INTERVALOS IGUALES Y CONTAMOS EL NÚMERO DE PESOS QUE HAY ANOTADOS EN CADA INTERVALO. LA **FRECUENCIA** ES EL NÚMERO DE ANOTACIONES DE CADA INTERVALO. LA **FRECUENCIA RELATIVA** ES LA PROPORCIÓN DE PESOS DE CADA INTERVALO, ES DECIR, LA FRECUENCIA DIVIDIDA ENTRE EL TOTAL DE ESTUDIANTES.

INTERVALO DE CLASES	PUNTO MEDIO	FRECUENCIA	FRECUENCIA RELATIVA
87,5-102,4	95	2	0,022
102,5-117,4	110	9	0,098
117,5-132,4	125	19	0,206
132,5-147,4	140	17	0,185
147,5-162,4	155	27	0,293
162,5-177,4	170	8	0,087
177,5-192,4	185	8	0,087
192,5-207,4	200	1	0,011
207,5-222,4	215	1	0,011
TOTAL		92	1,000

NOTA: HEMOS INTENTADO QUE LOS LÍMITES DE LOS INTERVALOS NO COINCIDAN CON LOS MÚLTIPLOS DE CINCO LIBRAS. ASÍ EVITAMOS EL SESGO EN LOS DATOS QUE HAN FACILITADO LOS ESTUDIANTES.

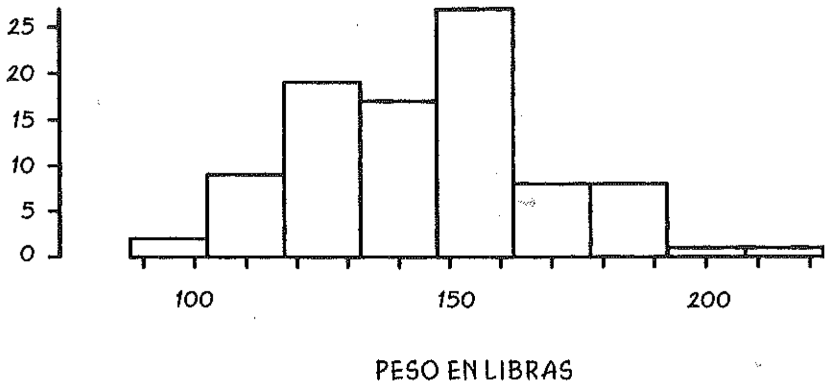
DIRECTRICES PARA ESTABLECER LOS INTERVALOS DE CLASE:

1) UTILIZA INTERVALOS IGUALES, QUE TENGAN SU PUNTO MEDIO EN NÚMEROS APROPIADOS.

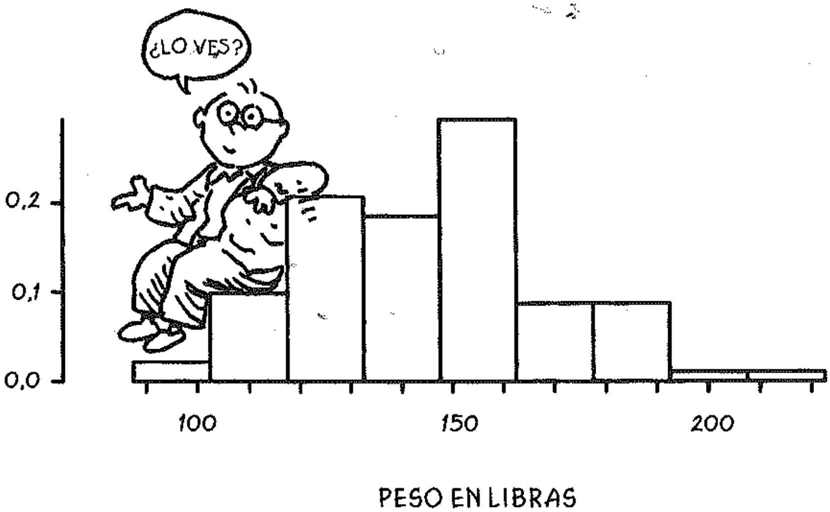
2) CUANDO HAYA POCOS DATOS, UTILIZA POCOS INTERVALOS.

3) CUANDO LOS DATOS SEAN MUY NUMEROSOS, ¡UTILIZA TAMBIÉN MUCHOS INTERVALOS!

CON LA TABLA DE FRECUENCIAS MOSTRAMOS CUÁNTAS OBSERVACIONES HAY «ALREDEDOR» DE CADA VALOR. ESTA INFORMACIÓN TAMBIÉN SE PUEDE REPRESENTAR CON UN DIAGRAMA DE BARRAS QUE SE LLAMA **HISTOGRAMA**. CADA BARRA CUBRE UN INTERVALO Y TIENE SU PUNTO MEDIO EN EL CENTRO. LA ALTURA DE UNA BARRA REPRESENTA LA CANTIDAD DE PUNTOS, U OBSERVACIONES, DE CADA INTERVALO.



TAMBIÉN PODEMOS DIBUJAR UN **HISTOGRAMA DE FRECUENCIAS RELATIVAS**, REPRESENTANDO ÉSTAS EN FUNCIÓN DEL PESO. VISUALMENTE ES EL MISMO, SÓLO CAMBIA LA ESCALA VERTICAL.



EL ESTADÍSTICO **JOHN TUKEY** INVENTÓ UNA FORMA RÁPIDA PARA RESUMIR LOS DATOS Y MANTENER A LA VEZ TODAS LAS OBSERVACIONES INDIVIDUALES. LO LLAMÓ GRÁFICO DE **TALLOS Y HOJAS**.

EN ESTE CASO, EL TALLO ES UNA COLUMNA NUMÉRICA EN LA QUE SE REPRESENTA EL PESO DE DIEZ EN DIEZ LIBRAS, OMITIENDO EL ÚLTIMO DÍGITO.

9
10
11
12
13
14
15
16
17
18
19
20
21

LUEGO SE AÑADE EL ÚLTIMO DÍGITO DE CADA PESO EN LA LÍNEA CORRESPONDIENTE:

TALLO : HOJAS
9 :
10 :
11 : 628
12 : 0155005
13 : 080015
14 : 05
15 : 0
16 :
17 :
18 :
19 :
20 :
21 :

CUANDO SE HAN AÑADIDO TODOS LOS DATOS, EL DIAGRAMA TIENE ESTE ASPECTO:

9 : 5
10 : 288
11 : 628855060
12 : 01553005525
13 : 8500850600153
14 : 05505580502
15 : 50537055055050500500
16 : 050004
17 : 055000
18 : 0500
19 : 00500
20 :
21 : 5]

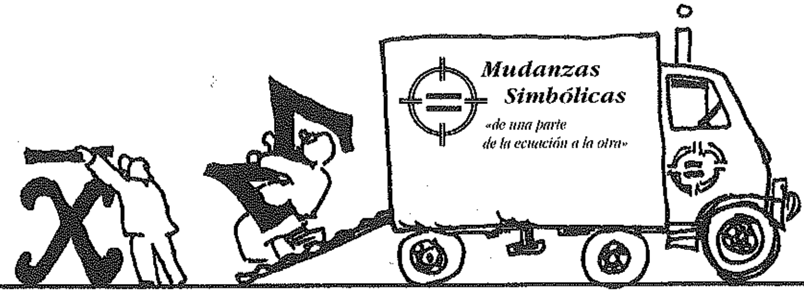
POR ÚLTIMO, SE ORDENAN LAS «HOJAS».

9 : 5
10 : 288
11 : 002556688
12 : 00012355555
13 : 0000013555688
14 : 00002555558
15 : 000000000355555555557
16 : 000045
17 : 000055
18 : 0005
19 : 00005
20 :
21 : 5

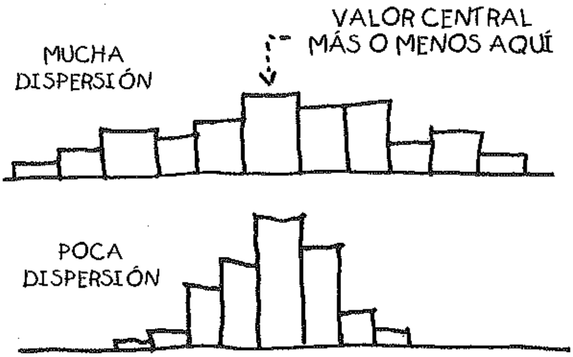
¡TODOS ESOS CEROS Y CINCOS DEMUESTRAN EL SESGO EN LA INFORMACIÓN QUE HAN PROPORCIONADO LOS ESTUDIANTES!

RESUMEN NUMÉRICO

AHORA PASAMOS DE LOS GRÁFICOS A LAS FÓRMULAS. NUESTRO OBJETIVO ES CONSEGUIR CÁLCULOS SIMPLES DE LAS CARACTERÍSTICAS BÁSICAS DE UN CONJUNTO DE DATOS.



TODO CONJUNTO DE DATOS TIENE DOS PROPIEDADES PRINCIPALES: EL **VALOR CENTRAL**, O **TÍPICO**, Y LA **DISPERSIÓN** DE ESE VALOR. PUEDES HACERTE UNA IDEA CON ESTOS HISTOGRAMAS HIPOTÉTICOS.



SE PUEDE AVANZAR MUCHO CON POCA NOTACIÓN. SUPÓN QUE HEMOS HECHO UNA SERIE DE OBSERVACIONES... n OBSERVACIONES, PARA SER EXACTOS. ENTONCES ESCRIBIMOS

$$x_1, x_2, x_3, \ldots x_n$$

PARA CADA UNO DE LOS VALORES QUE HEMOS OBSERVADO. DE ESTA FORMA, n ES EL NÚMERO TOTAL DE DATOS, Y x_4, POR EJEMPLO, ES EL VALOR DEL CUARTO DATO.

LA **TABLA** ES UNA FORMA DE ORDENAR LOS DATOS:

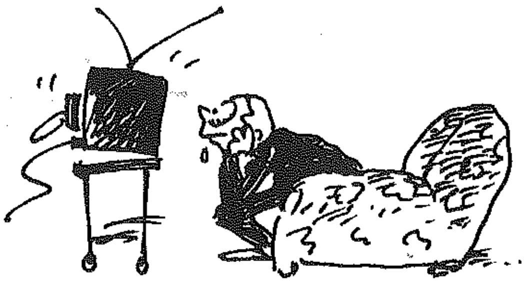
OBSERVACIÓN	1	2	3	4	....	n
VALOR DEL DATO	x_1	x_2	x_3	x_4	....	x_n.

UN PEQUEÑO CONJUNTO DE $n = 5$ DATOS FACILITA LAS OPERACIONES. SUPÓN, POR EJEMPLO, QUE PREGUNTAMOS A CINCO PERSONAS CUÁNTAS HORAS DE **TELEVISIÓN** VEN A LA SEMANA, Y CONFECCIONAMOS LA SIGUIENTE TABLA:

OBSERVACIÓN	1	2	3	4	5
VALOR DEL DATO	5	7	3	38	7

ENTONCES $x_1 = 5$, $x_2 = 7$, $x_3 = 3$, $x_4 = 38$, Y $x_5 = 7$.

¿CUÁL ES EL «CENTRO» DE ESTOS DATOS? HAY DIFERENTES FORMAS DE CALCULARLO, PERO SÓLO VEREMOS DOS.



LA MEDIA

LA **MEDIA** SE REPRESENTA CON EL SÍMBOLO $\bar{x}$, Y SE OBTIENE DIVIDIENDO LA SUMA DE TODOS LOS DATOS ENTRE EL NÚMERO DE OBSERVACIONES:

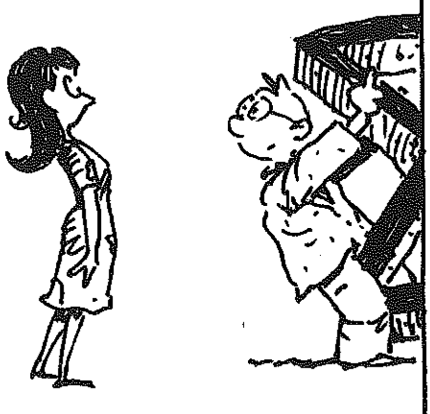
$$\bar{x} = \frac{\text{SUMA DE LOS DATOS}}{n}$$

$$= \frac{x_1 + x_2 + \dots + x_n}{n}$$

EN NUESTRO EJEMPLO:

$$\bar{x} = \frac{5 + 7 + 3 + 38 + 7}{5} = \frac{60}{5}$$

$$= 12 \text{ HORAS}$$



ASÍ QUE, VAMOS A REPETIRLO. LA MEDIA DE UN CONJUNTO DE DATOS x_i ES

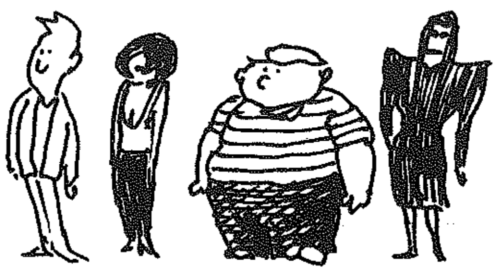
$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{O BIEN} \quad \sum_{i=1}^{n} \frac{x_i}{n}$$

EN EL CASO DE NUESTROS 92 ALUMNOS DE PENNSYLVANIA, EL PESO MEDIO ES

$$\sum_{i=1}^{92} \frac{x_i}{92} = \frac{13.354}{92}$$

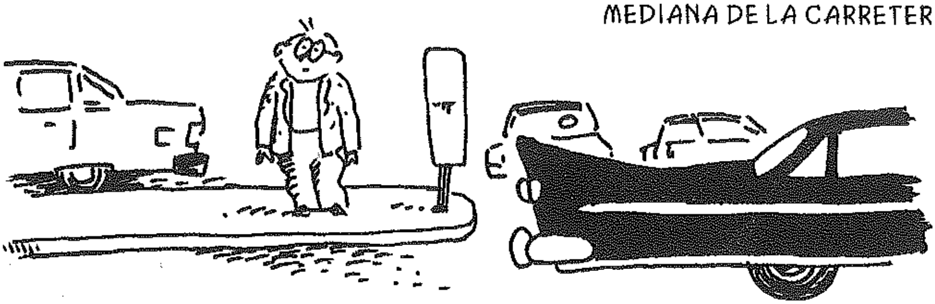
$$=$$

145,15 LIBRAS



LA **MEDIANA**

ES OTRO TIPO DE CENTRO: EL «PUNTO MEDIO» DE LOS DATOS, IGUAL QUE LA MEDIANA DE LA CARRETERA.



PARA ENCONTRAR LA MEDIANA DE UN CONJUNTO DE DATOS, ORDENAMOS LOS DATOS DE MENOR A MAYOR. LA MEDIANA ES EL VALOR QUE QUEDA EN EL CENTRO.

3 5 7 7 38

MEDIANA

SI EL NÚMERO DE OBSERVACIONES ES **PAR**, EN CUYO CASO NO HAY NINGÚN PUNTO CENTRAL, HACEMOS LA MEDIA DE LOS DOS VALORES QUE QUEDAN EN EL CENTRO. ASÍ QUE SI LOS DATOS SON

3 5 7 7

ESPACIO CENTRAL

HACEMOS LA MEDIA DE 5 Y 7: $\frac{5+7}{2} = 6$

ESTO NOS DA UNA REGLA GENERAL: ORDENAR LOS DATOS DE MENOR A MAYOR.

SI EL NÚMERO DE DATOS ES **IMPAR**, LA MEDIANA ES EL VALOR CENTRAL.

SI EL NÚMERO DE DATOS ES **PAR**, LA MEDIANA ES LA MEDIA DE LOS DOS DATOS MÁS CERCANOS AL CENTRO.

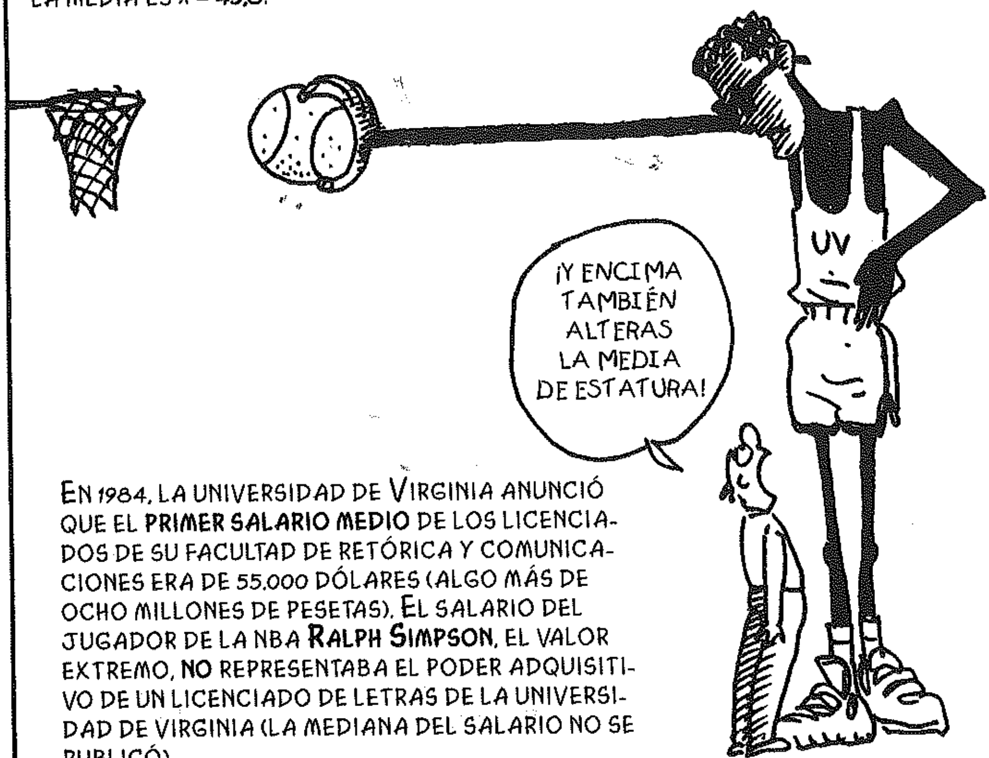
PARA ENCONTRAR LA MEDIANA DE LOS PESOS DE LOS ESTUDIANTES $n = 92$, PODEMOS UTILIZAR EL GRÁFICO DE TALLOS Y HOJAS ORDENADO: CUENTA HASTA LA OBSERVACIÓN NÚMERO 46. LA MEDIANA ES

$$\frac{x_{46} + x_{47}}{2} = \frac{145 + 145}{2}$$

$$= 145 \text{ LIBRAS}$$

9 : 5
10 : 288
11 : 002556688
12 : 00012355555
13 : 0000013555688
14 : 000025555558
15 : 00000000003555555555557
16 : 000045
17 : 000055
18 : 0005
19 : 00005
20 :
21 : 5

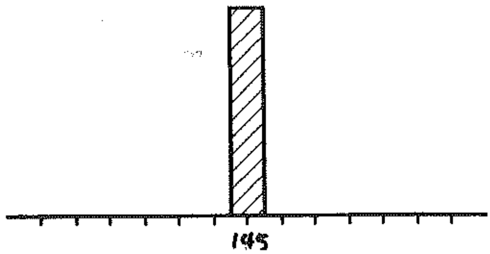
¿POR QUÉ EXISTEN DIFERENTES FORMAS DE CALCULAR EL CENTRO? CADA UNA TIENE SUS VENTAJAS. POR EJEMPLO, LA **MEDIANA** NO SE VE AFECTADA POR LOS DATOS MÁS ALEJADOS DEL CENTRO, LOS VALORES EXTREMOS QUE NO SON TÍPICOS DEL CONJUNTO DE DATOS. SUPONGAMOS QUE EN NUESTRO PEQUEÑO GRUPO DE TELESPECTADORES HAY UNA PERSONA QUE VE 200 HORAS DE TELEVISIÓN A LA SEMANA. NUESTROS DATOS SERÍAN ENTONCES 3, 5, 7, 7, 200. LA MEDIANA, 7, SIGUE SIENDO LA MISMA, ¡PERO AHORA LA MEDIA ES $\bar{x} = 45{,}8$!



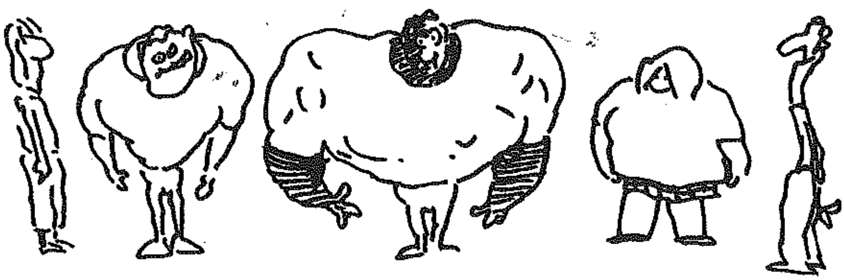
EN 1984, LA UNIVERSIDAD DE VIRGINIA ANUNCIÓ QUE EL **PRIMER SALARIO MEDIO** DE LOS LICENCIADOS DE SU FACULTAD DE RETÓRICA Y COMUNICACIONES ERA DE 55.000 DÓLARES (ALGO MÁS DE OCHO MILLONES DE PESETAS). EL SALARIO DEL JUGADOR DE LA NBA **RALPH SIMPSON**, EL VALOR EXTREMO, **NO** REPRESENTABA EL PODER ADQUISITIVO DE UN LICENCIADO DE LETRAS DE LA UNIVERSIDAD DE VIRGINIA (LA MEDIANA DEL SALARIO NO SE PUBLICÓ).

MEDIDAS DE DISPERSIÓN

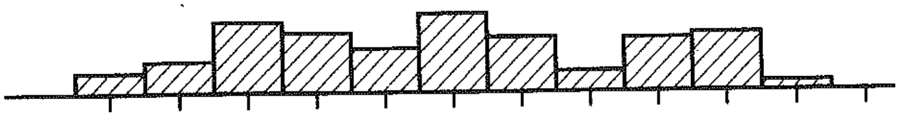
ADEMÁS DE CONOCER EL PUNTO CENTRAL DE UN CONJUNTO DE DATOS, TAMBIÉN QUEREMOS DESCRIBIR LA **DISPERSIÓN**, ES DECIR, A CUÁNTA DISTANCIA DEL CENTRO SE ENCUENTRAN LOS DATOS. POR EJEMPLO, SI TODOS LOS ESTUDIANTES PESARAN **EXACTAMENTE** 145 LIBRAS, NO HABRÍA DISPERSIÓN. NUMÉRICAMENTE, LA DISPERSIÓN SERÍA **CERO** Y EL HISTOGRAMA SERÍA MUY DELGADITO.



PERO SI MUCHOS DE LOS ESTUDIANTES PESARAN O BIEN MUCHO Y/O MUY POCO, VERÍAMOS OBVIAMENTE QUE HAY DISPERSIÓN. POR EJEMPLO, SI EL **EQUIPO DE FÚTBOL** HUBIERA PARTICIPADO EN EL EXPERIMENTO...



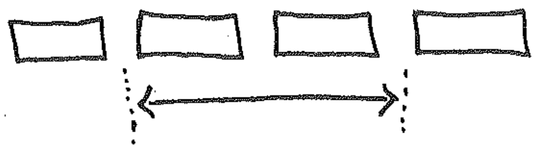
EL HISTOGRAMA SERÍA MÁS EXTENSO, ALGO ASÍ:



TAMBIÉN HAY VARIAS FORMAS DE MEDIR LA DISPERSIÓN. UNA ES EL

RECORRIDO INTERCUARTÍLICO

SE TRATA DE DIVIDIR LOS DATOS EN CUATRO GRUPOS IGUALES Y OBSERVAR LA DISTANCIA QUE SEPARA LOS GRUPOS EXTREMOS.



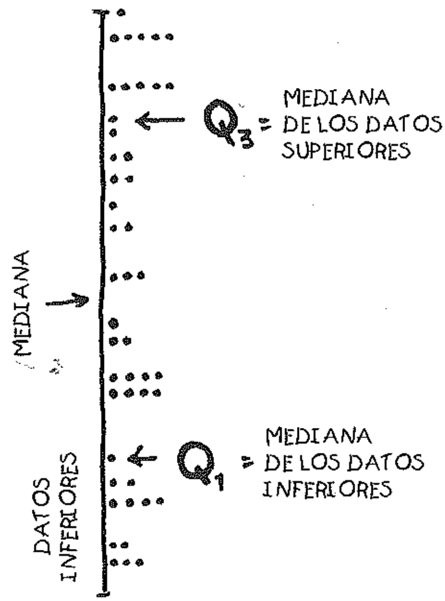
ÉSTA ES LA RECETA:

1) ORDENA LOS DATOS NUMÉRICAMENTE.

2) DIVIDE LOS DATOS POR LA MEDIANA EN DOS GRUPOS IGUALES (SI LA MEDIANA COINCIDE CON UN DATO, INCLÚYELO EN LOS DOS GRUPOS).

3) CALCULA LA MEDIANA DEL GRUPO INFERIOR. ÉSE ES EL PRIMER CUARTIL, O Q_1.

4) LA MEDIANA DEL GRUPO SUPERIOR ES EL TERCER CUARTIL, O Q_3.



EL RECORRIDO INTERCUARTÍLICO (IQR) ES LA DISTANCIA (O DIFERENCIA) QUE HAY ENTRE ELLOS:

$$IQR = Q_3 - Q_1$$

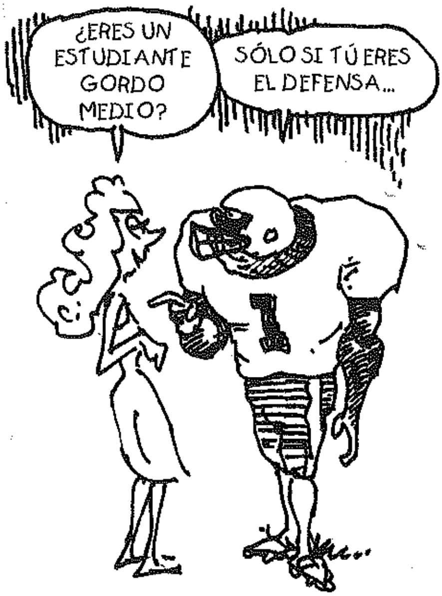
ESTOS SON LOS DATOS DE LOS PESOS, EN LOS QUE HEMOS DESTACADO LOS PUNTOS MEDIOS DEL GRUPO INFERIOR Y DEL SUPERIOR:

```
 9 : 5
10 : 288
11 : 002556688
12 : 00012355555
13 : 0000013555688
14 : 00002555558
15 : 0000000000355555555557
16 : 000045
17 : 000055
18 : 0005
19 : 00005
20 :
21 : 5
```

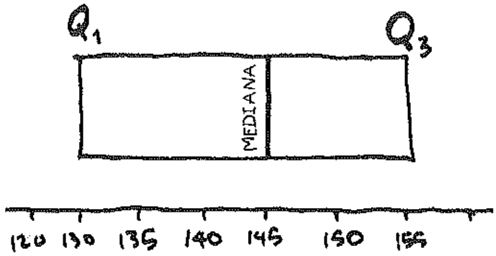
Y VEMOS QUE

$$\text{IQR} = 156 - 125$$
$$= 31 \text{ LIBRAS}$$

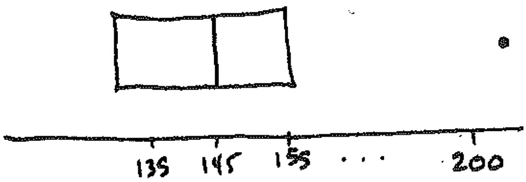
ES DECIR, LA DIFERENCIA ENTRE LA MEDIANA DE LOS ESTUDIANTES QUE PESAN MUCHO Y LA MEDIANA DE LOS ESTUDIANTES QUE PESAN POCO.



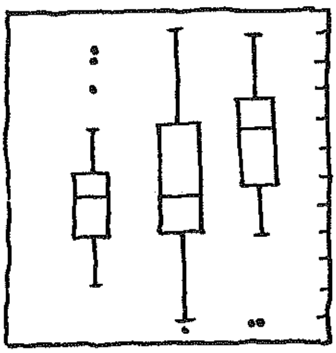
JOHN TUKEY INVENTÓ OTRO TIPO DE REPRESENTACIÓN PARA MOSTRAR EL IQR, EL GRÁFICO DE CAJA. LOS EXTREMOS DE LA CAJA SON LOS CUARTILES Q_1 Y Q_3. LA MEDIANA SE DIBUJA DENTRO DE LA CAJA.



SI UN PUNTO SE ENCUENTRA A MÁS DE 1,5 IQR DE LOS EXTREMOS DE LA CAJA, SE CONSIDERA QUE ES UNA **OBSERVACIÓN ATÍPICA**, Y SE REPRESENTA INDIVIDUALMENTE.



POR ÚLTIMO, EXTENDEMOS LÍNEAS HASTA LOS PUNTOS MÁS ALEJADOS QUE NO SON OBSERVACIONES ATÍPICAS (ES DECIR, QUE SE ENCUENTRAN A MENOS DE 1,5 IQR DE LOS CUARTILES).

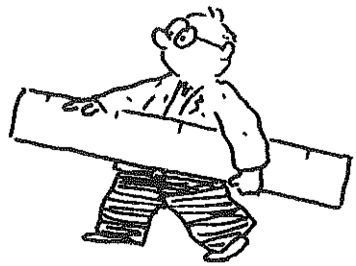


EL GRÁFICO CAJA ES MUY ÚTIL, EN ESPECIAL PARA REPRESENTAR LAS DIFERENCIAS ENTRE GRUPOS.

LA MEDIDA ESTÁNDAR DE LA DISPERSIÓN ES LA

DESVIACIÓN TÍPICA (TAMBIÉN DESVIACIÓN ESTÁNDAR)

A DIFERENCIA DEL IQR, QUE SE CALCULA A PARTIR DE LAS MEDIANAS, LA DESVIACIÓN TÍPICA MIDE LA DISPERSIÓN DE LOS DATOS DESDE LA **MEDIA**. UNA FORMA INTUITIVA DE VERLA ES COMO LA DISTANCIA MEDIA ENTRE LOS DATOS Y LA MEDIA $\bar{x}$...

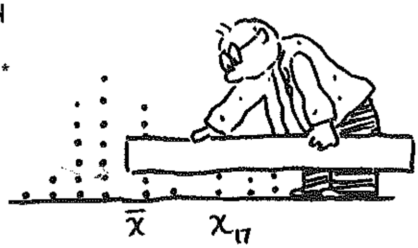


SIN EMBARGO, EN ESTA OCASIÓN UTILIZAMOS LAS DISTANCIAS ELEVADAS AL **CUADRADO**. O SEA, SI LA DISTANCIA AL CUADRADO ENTRE EL PUNTO x_i Y $\bar{x}$ ES $(x_i - \bar{x})^2$, ENTONCES

$$\text{LA DISTANCIA CUADRÁTICA MEDIA} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

POR MOTIVOS TÉCNICOS, SE UTILIZA $n-1$ EN EL DENOMINADOR EN LUGAR DE n, Y DEFINIMOS ENTONCES LA **VARIANZA MUESTRAL** s^2:*

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



* TAMBIÉN ES CORRECTA VARIANCIA. [N.T.]

EN EL CONJUNTO DE DATOS {3 5 7 7 38}, DONDE $\bar{x} = 12$ Y $n = 5$, LA VARIANZA SE CALCULA ASÍ:

$$s^2 = \frac{(3-12)^2 + (5-12)^2 + (7-12)^2 + (7-12)^2 + (38-12)^2}{(5-1)}$$

$$= \frac{81 + 49 + 25 + 25 + 676}{4}$$

$$= 214$$

SIN EMBARGO, LA MEDIDA DE LA DISPERSIÓN DEBERÍA MEDIRSE EN LAS MISMAS **UNIDADES** QUE EL CONJUNTO INICIAL DE DATOS. EN EL EJEMPLO DE LOS PESOS, LA VARIANZA s^2 SE CALCULA EN LIBRAS **AL CUADRADO**... ¡UY!

LO MÁS LÓGICO ES HACER LA RAÍZ CUADRADA, Y ESO ES PRECISAMENTE LO QUE HACEMOS PARA DEFINIR LA...:

DESVIACIÓN TÍPICA

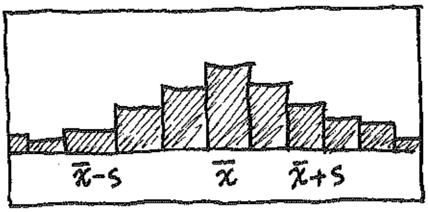
$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

QUE EN NUESTRO PEQUEÑO CONJUNTO DE DATOS ES

$$s = \sqrt{214} = 14{,}63$$

INCLUSO CUANDO EL CONJUNTO DE DATOS ES PEQUEÑO, ¡LOS CÁLCULOS PUEDEN SER AGOTADORES! EN LA ACTUALIDAD BASTA CON APRETAR EL BOTÓN s DE LA CALCULADORA, O CONSULTAR SU VALOR CON LA AYUDA DE UN PAQUETE INFORMÁTICO.

Las propiedades de $\overline{X}$ y S

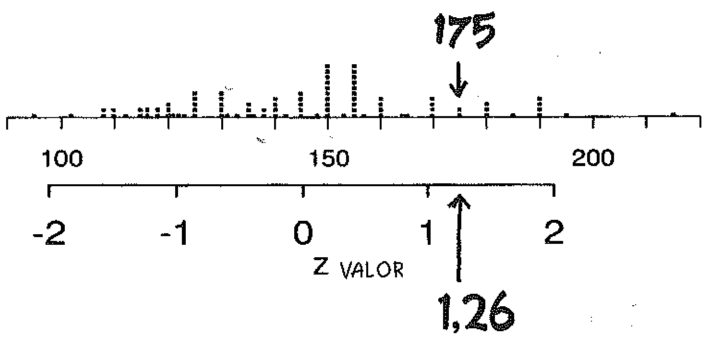


LA MEDIA Y LA DESVIACIÓN TÍPICA SON MUY ÚTILES PARA RESUMIR LAS PROPIEDADES DE **HISTOGRAMAS** BASTANTE **SIMÉTRICOS, SIN OBSERVACIONES ATÍPICAS**, O SEA, HISTOGRAMAS CON FORMA DE MONTAÑA.

A MENUDO RESULTA ÚTIL SABER CUÁNTAS DESVIACIONES TÍPICAS DISTA UN PUNTO DE LA MEDIA. ENTONCES DEFINIMOS z, O VALORES ESTANDARIZADOS, COMO LA DISTANCIA DESDE $\bar{x}$ **POR DESVIACIÓN TÍPICA**.

$$z_i = \frac{x_i - \bar{x}}{s} \quad \text{PARA CADA } i.$$

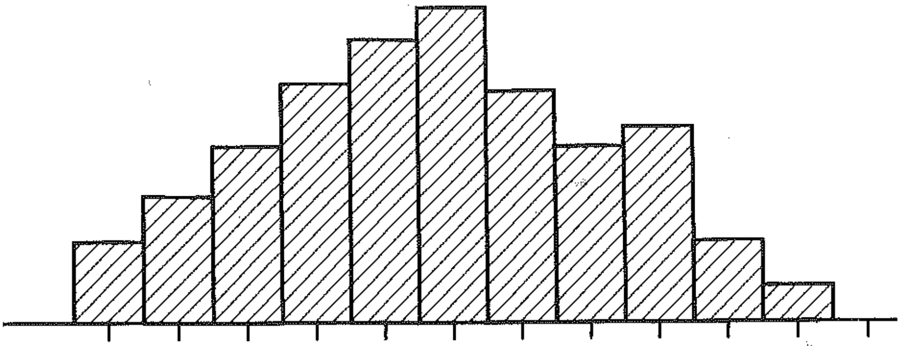
UNA z DE +2 QUIERE DECIR QUE LA OBSERVACIÓN SE ENCUENTRA **DOS DESVIACIONES TÍPICAS POR ENCIMA DE LA MEDIA**. EN LOS DATOS DE LOS PESOS DE LOS ESTUDIANTES ($\bar{x} = 145,2$ Y $s = 23,7$), PODEMOS REPRESENTAR LOS DATOS SIMULTÁNEAMENTE EN EL EJE x DEL PRINCIPIO Y UN EJE z.



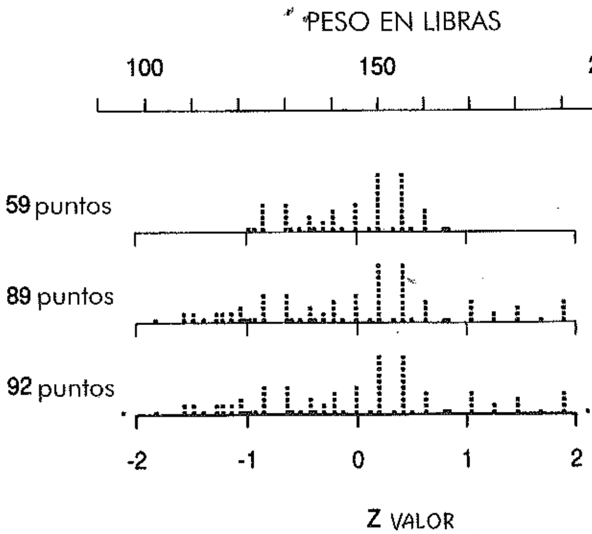
UN ESTUDIANTE QUE PESE 175 LIBRAS TIENE UNA z DE $\frac{175-145,2}{23,7} = 1,26$

una REGLA EMPÍRICA:

EN LOS CONJUNTOS DE DATOS CASI SIMÉTRICOS, CON FORMA DE MONTAÑA, ALREDEDOR DE UN **68%** DE LOS DATOS SE ENCUENTRA A MENOS DE **UNA** DESVIACIÓN TÍPICA DE LA MEDIA, Y EL **95%** ESTÁ A MENOS DE **DOS** DESVIACIONES TÍPICAS DE LA MEDIA.



SI MIRAMOS LOS PESOS, ESTA REGLA EMPÍRICA FUNCIONA BASTANTE BIEN: UN **64%** (= 59/92) DE LOS PESOS ESTÁN A MENOS DE UNA DESVIACIÓN TÍPICA DE LA MEDIA, Y UN **97%** (= 89/92) ESTÁN A MENOS DE DOS DESVIACIONES TÍPICAS DE LA MEDIA.



Y AHORA TOCA DESCANSAR DE TANTO NÚMERO.

¡HEMOS APRENDIDO MUCHO EN UN SOLO CAPÍTULO! EMPEZAMOS CON UN MONTÓN DE NÚMEROS DESORDENADOS, Y AHORA YA TENEMOS:

1) DIFERENTES FORMAS DE REPRESENTARLOS.

2) DOS CONCEPTOS DIFERENTES DEL CENTRO DE LOS DATOS, LA MEDIANA Y LA MEDIA.

3) DOS FORMAS DE CALCULAR LA DISPERSIÓN DE LOS DATOS ALREDEDOR DEL CENTRO.

4) HISTOGRAMAS EN FORMA DE MONTAÑA Y *z*, UNA VARIABLE QUE INDICA A CUÁNTAS DESVIACIONES TÍPICAS DE LA MEDIA SE ENCUENTRA UNA OBSERVACIÓN.

AHORA, PARA INVESTIGAR CON MÁS PROFUNDIDAD EL COMPORTAMIENTO DE LOS DATOS, VAMOS A DAR UN PEQUEÑO PASEO POR EL REINO DE LA **ALEATORIEDAD**... UNA TIERRA EN LA QUE TODO FUNCIONA **SIEMPRE** A LARGO PLAZO, Y DONDE NO HAY MÁS LEY QUE LA DEL **CASINO**...

♦ Capítulo 3 ♦

LA PROBABILIDAD

EN LA VIDA, NADA ES SEGURO. EN TODAS NUESTRAS ACCIONES, CALCULAMOS SIEMPRE LAS POSIBILIDADES DE UN BUEN RESULTADO, TANTO EN EL MUNDO DE LOS NEGOCIOS COMO EN LA MEDICINA O EL CLIMA. SIN EMBARGO, EN LA HISTORIA DE LA HUMANIDAD, LA **PROBABILIDAD**, EL ESTUDIO FORMAL DE LAS LEYES DEL AZAR, SE HA UTILIZADO PARA UNA SOLA COSA: **EL JUEGO**.

NADIE SABE CUÁNDO SE INVENTÓ EL JUEGO. COMO MÍNIMO SE REMONTA A TIEMPOS TAN ANCESTRALES COMO EL **ANTIGUO EGIPTO**, CUANDO HOMBRES Y MUJERES JUGABAN CON «**ASTRÁGALOS**» HECHOS CON LAS TABAS DE ANIMALES.

EL EMPERADOR ROMANO **CLAUDIO** (10 A. DE C. - 54 D. DE C.) ESCRIBIÓ EL PRIMER TRATADO SOBRE EL JUEGO. POR DESGRACIA, EL LIBRO *CÓMO GANAR A LOS DADOS* NO SE HA CONSERVADO.

LOS DADOS, TAL Y COMO LOS CONOCEMOS EN LA ACTUALIDAD, SE HICIERON MUY POPULARES EN LA EDAD MEDIA, A TIEMPO PARA QUE UN CALAVERA DEL RENACIMIENTO, **CHEVALIER DE MERE**, PROPUSIERA UN ENIGMA MATEMÁTICO:

Chevalier razonó que la probabilidad de obtener una tirada ganadora era la misma en los dos juegos:

La probabilidad de un seis $= \frac{1}{6}$

La media en cuatro tiradas $= 4 \cdot \left(\frac{1}{6}\right) = \frac{2}{3}$

La probabilidad de un doble seis en una tirada $= \frac{1}{36}$

La media en 24 tiradas $= 24 \cdot \left(\frac{1}{36}\right) = \frac{2}{3}$

Entonces, ¿¿por qué **perdía más a menudo con la segunda apuesta**??

De Mere le planteó la pregunta a su amigo el genio **Blaise Pascal** (*1623-1666*).

A pesar de que Pascal había renunciado a las matemáticas por considerarlas una forma de deleite sexual (!!), aceptó estudiar el problema de De Mere.

Pascal escribió a su compañero, también genio, **Pierre de Fermat**, y en el transcurso de unas cuantas cartas, los dos ya habían desarrollado la teoría de la probabilidad en su forma moderna (sin viñetas, claro).

DEFINICIONES BÁSICAS

MIENTRAS NUESTROS JUGADORES ECHAN UNA PARTIDA, NOSOTROS JUGAREMOS A SER CIENTÍFICOS Y ANALIZAREMOS LOS RESULTADOS:

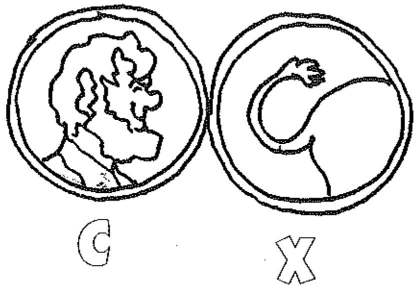
UN **experimento aleatorio** ES EL PROCESO DE OBSERVAR EL RESULTADO DE UN SUCESO CASUAL.

LOS **resultados elementales** SON TODOS LOS POSIBLES RESULTADOS INDIVISIBLES DEL EXPERIMENTO ALEATORIO.

EL **espacio muestral** ES EL CONJUNTO O EL COMPENDIO DE TODOS LOS RESULTADOS ELEMENTALES.

SI SE LANZA UNA MONEDA AL AIRE, POR EJEMPLO, EL **EXPERIMENTO ALEATORIO** CONSISTE EN TOMAR NOTA DE LOS RESULTADOS...

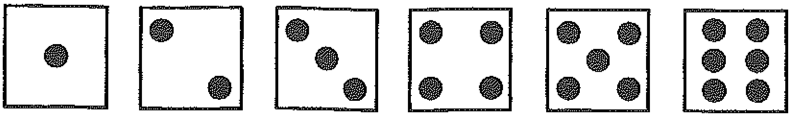
LOS **RESULTADOS ELEMENTALES** SON CARA O CRUZ...



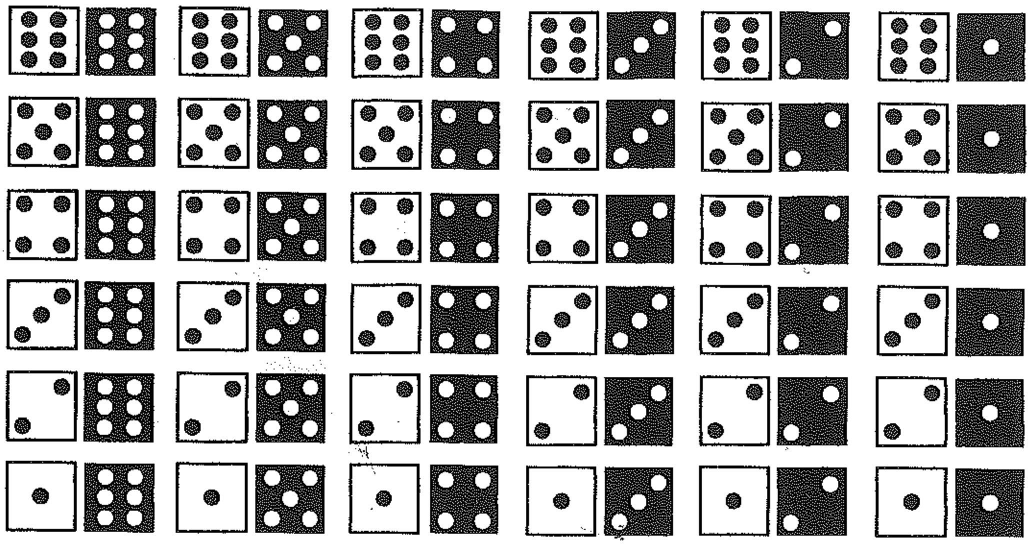
Y EL **ESPACIO MUESTRAL** SE ESCRIBE

$$\{C, X\}$$

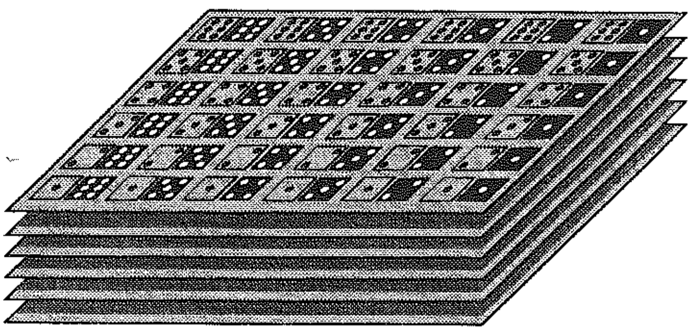
EL ESPACIO MUESTRAL DE UNA TIRADA CON **UN SOLO DADO** ES ALGO MAYOR.



Y CON **DOS** DADOS, EL ESPACIO MUESTRAL SERÍA ASÍ (HEMOS DIBUJADO UN DADO BLANCO Y OTRO NEGRO PARA DISTINGUIRLOS):

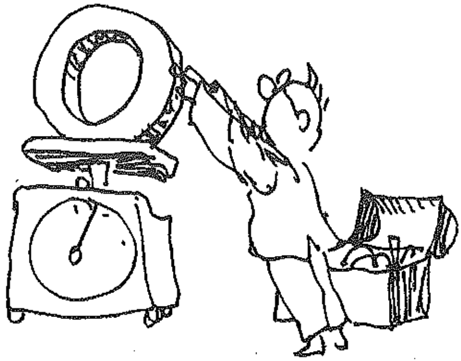


ESTE **ESPACIO MUESTRAL** TIENE **36** (6 × 6) RESULTADOS ELEMENTALES. CON TRES DADOS, EL ESPACIO TENDRÍA 216 ENTRADAS, COMO EN ESTE MONTÓN DE 6 × 6 × 6. ¿Y CON **CUATRO DADOS**?



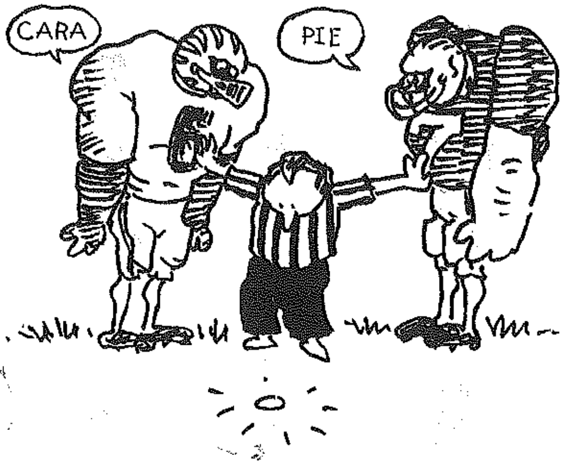
EN ALGÚN MOMENTO TENDREMOS QUE DEJAR DE ENUMERAR Y EMPEZAR A RAZONAR...

VAMOS A IMAGINAR UN EXPERIMENTO ALEATORIO CON n RESULTADOS ELEMENTALES, $O_1, O_2, \ldots O_n$. QUEREMOS ASIGNARLES UN **PESO NUMÉRICO**, O **PROBABILIDAD**, A CADA UNO PARA MEDIR LA POSIBILIDAD DE QUE APAREZCA. ESCRIBIMOS LA PROBABILIDAD DE O_i COMO $P(O_i)$.



POR EJEMPLO, SI SE LANZA UNA MONEDA AL AIRE, HAY TANTAS POSIBILIDADES DE OBTENER CARA COMO CRUZ, Y LES ASIGNAMOS UNA PROBABILIDAD DE 0,5.

$$P(C) = P(X) = 0{,}5$$



CADA RESULTADO APARECE EN LA MITAD DE OCASIONES. ¡PREGUNTA A CUALQUIER **JUGADOR DE FÚTBOL**!

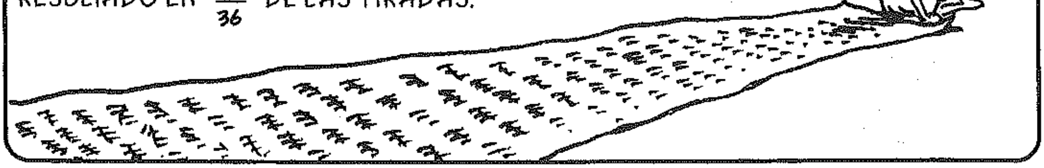
CUANDO SE TIRAN **DOS DADOS**, HAY 36 RESULTADOS ELEMENTALES Y TODOS TIENEN LA MISMA POSIBILIDAD DE APARECER.

ASÍ QUE LA PROBABILIDAD DE CADA UNO ES DE $\frac{1}{36}$.

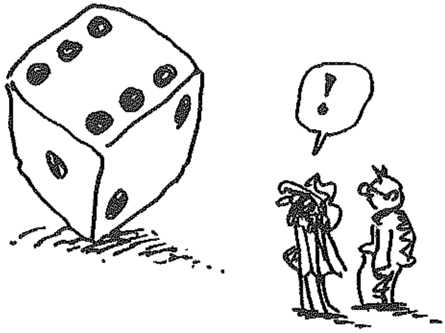
POR EJEMPLO,

$$P(\text{NEGRO } 5, \text{ BLANCO } 2) = \frac{1}{36}$$

ES DECIR: SI TIRAMOS LOS DADOS MUCHAS VECES, A LARGO PLAZO OBTENDREMOS ESE RESULTADO EN $\frac{1}{36}$ DE LAS TIRADAS.



PERO, ¿QUÉ PASA SI NUESTRO JUGADOR HACE TRAMPAS Y TIRA UN DADO TRUCADO? A MODO DE EJEMPLO, SUPONDREMOS QUE EL UNO SALE UN 25% DE LAS VECES (A LARGO PLAZO).



EL ESPACIO MUESTRAL ES EL MISMO QUE EL DE UN DADO SIN TRUCAR

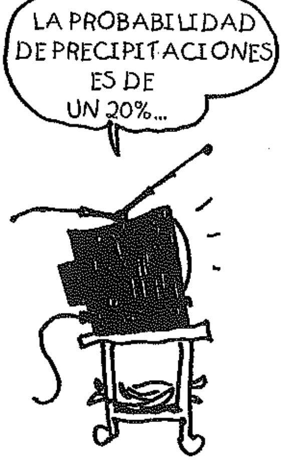
$$\{1, 2, 3, 4, 5, 6\}$$

0,25 0,15 0,15 0,15 0,15 0,15

PERO LAS PROBABILIDADES SON DIFERENTES. AHORA **P(1) = 0,25** Y EL RESTO DE PROBABILIDADES SUMAN EL 0,75 RESTANTE. SI 2, 3, 4, 5 Y 6 TUVIERAN LA MISMA PROBABILIDAD DE SALIR, CADA UNO TENDRÍA UNA PROBABILIDAD DE

$$0{,}15 = \frac{1}{5}\,(0{,}75)$$

EN GENERAL, LOS RESULTADOS ELEMENTALES NO TIENEN POR QUÉ TENER LA MISMA PROBABILIDAD.



¿QUÉ PODEMOS DECIR DE LA PROBABILIDAD DE $P(O_i)$ EN UN EXPERIMENTO ALEATORIO CUALQUIERA? EN PRIMER LUGAR,

$$P(O_i) \geq 0$$

LAS PROBABILIDADES **NUNCA** SON **NEGATIVAS**. UNA PROBABILIDAD DE CERO SIGNIFICA QUE ESE SUCESO NUNCA TENDRÁ LUGAR. UNA PROBABILIDAD POR DEBAJO DE CERO NO TIENE SENTIDO.

EN SEGUNDO LUGAR, SI UN SUCESO ES **SEGURO**, LE ASIGNAMOS UNA PROBABILIDAD DE 1. (A LARGO PLAZO, ESA ES LA PROPORCIÓN DE OCASIONES EN QUE OCURRIRÁ.)

EN CONCRETO, LA **PROBABILIDAD TOTAL DEL ESPACIO MUESTRAL** DEBE SER 1. SI HACEMOS EL EXPERIMENTO, ¡**ALGO** TIENE QUE PASAR!

SI UNIMOS ESTOS DOS ÚLTIMOS PUNTOS, YA TENEMOS LAS **PROPIEDADES CARACTERÍSTICAS DE LA PROBABILIDAD**:

$P(O_i) \geq 0$ — LA PROBABILIDAD **NO** ES **NEGATIVA**

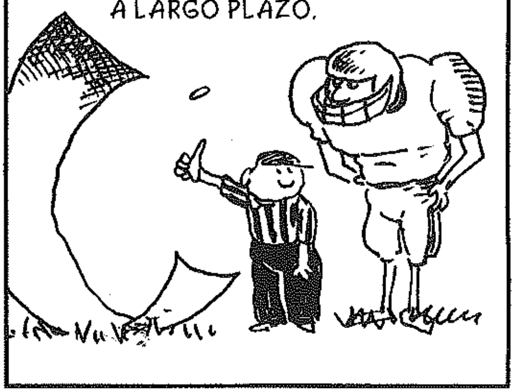
$P(O_1) + P(O_2) + \ldots + P(O_n) = 1$ — LA PROBABILIDAD TOTAL DE LOS RESULTADOS ELEMENTALES ES **UNO**.

IGUAL QUE HARÍA UN BUEN POLÍTICO, HEMOS EVITADO CIERTAS **PREGUNTAS INCÓMODAS** COMO: A) ¿QUÉ **SIGNIFICA** PROBABILIDAD?; Y B) ¿CÓMO ASIGNAMOS UNA PROBABILIDAD A UN RESULTADO?

AQUÍ TENEMOS DIFERENTES FORMAS DE VERLO:

LA PROBABILIDAD **Clásica**: ESTÁ BASADA EN EL JUEGO, LA SUPOSICIÓN FUNDAMENTAL ES QUE EL JUEGO ES JUSTO Y QUE TODOS LOS RESULTADOS ELEMENTALES TIENEN LA MISMA PROBABILIDAD.

LA **Frecuencia Relativa**: CUANDO UN EXPERIMENTO SE **PUEDE** REPETIR, LA PROBABILIDAD DE UN RESULTADO ES LA PROPORCIÓN DE OCASIONES EN LAS QUE APARECE A LARGO PLAZO.

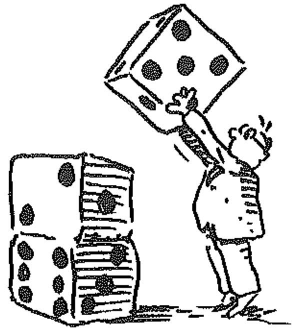


LA PROBABILIDAD **Personal**: LA MAYORÍA DE LOS SUCESOS DE LA VIDA SON **IRREPETIBLES**. LA PROBABILIDAD PERSONAL ES LA **VALORACIÓN PERSONAL QUE HACE UN INDIVIDUO** DE LAS POSIBILIDADES DE OBTENER UN RESULTADO. SI UN JUGADOR CREE QUE UN CABALLO TIENE MÁS DE UN 50% DE POSIBILIDADES DE GANAR, HARÁ LA CORRESPONDIENTE APUESTA.

UN **OBJETIVISTA** UTILIZA LA DEFINICIÓN CLÁSICA DE PROBABILIDAD O LA DE LA FRECUENCIA RELATIVA. UN **SUBJETIVISTA**, O **BAYESIANO**, APLICA LAS LEYES FORMALES DEL AZAR A SUS PROBABILIDADES PERSONALES, O A LAS NUESTRAS.

OPERACIONES BÁSICAS

HASTA AHORA, SÓLO HEMOS HABLADO DE LA PROBABILIDAD DE LOS RESULTADOS ELEMENTALES. EN TEORÍA, CON ESO BASTARÍA PARA DESCRIBIR UN EXPERIMENTO ALEATORIO, PERO EN LA PRÁCTICA RESULTA UN POCO DIFÍCIL DE MANEJAR. POR EJEMPLO, ALGO TAN NORMAL COMO OBTENER UN SIETE NO ESTÁ CONTEMPLADO EN LOS RESULTADOS ELEMENTALES... ASÍ QUE TENEMOS QUE INTRODUCIR UNA **NUEVA IDEA**:



UN **SUCESO** ES UN CONJUNTO DE RESULTADOS ELEMENTALES. LA PROBABILIDAD DE UN SUCESO ES LA SUMA DE LAS PROBABILIDADES DE LOS RESULTADOS ELEMENTALES DEL CONJUNTO. POR EJEMPLO, ALGUNOS SUCESOS EN LA VIDA DE UN JUGADOR CON DOS DADOS SERÍAN:

DESCRIPCIÓN DEL SUCESO	RESULTADOS ELEMENTALES DEL SUCESO	PROBABILIDAD
A: TIRADA TOTAL SUMA 3	{(1,2), (2,1)}	$P(A) = \frac{2}{36}$
B: TIRADA TOTAL SUMA 6	{(1,5), (2,4), (3,3), (4,2), (5,1)}	$P(B) = \frac{5}{36}$
C: DADO BLANCO CAE EN 1	{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)}	$P(C) = \frac{6}{36}$
D: DADO NEGRO CAE EN 1	{(1,1), (2,1), (3,1), (4,1), (5,1), (6,1)}	$P(D) = \frac{6}{36}$

Lo bello de utilizar sucesos en lugar de resultados elementales es que los podemos combinar para obtener otros distintos utilizando **operaciones lógicas**. Las palabras clave son **Y**, **O** y **NO**.

Es decir, con los sucesos E y F, podemos formar nuevos sucesos:

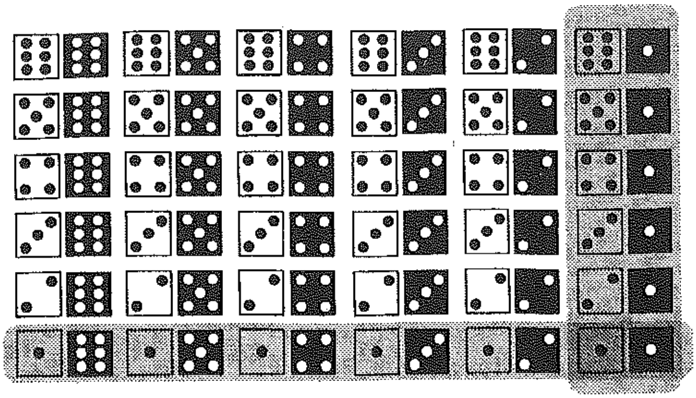
E **Y** F: Tanto el suceso E como el F ocurren.

E **O** F: Ocurre el suceso E, o el F, o los dos.

NO E: El suceso E no ocurre.

Si combinamos las primeras definiciones de probabilidad con estas operaciones lógicas, obtenemos poderosas fórmulas para manipular las probabilidades.

VOLVAMOS AL EJEMPLO DE LOS DADOS. SI C ES EL SUCESO DADO BLANCO = 1, Y SI EL SUCESO D ES DADO NEGRO = 1, ENTONCES:



C O D ES LA ZONA SOMBREADA (DONDE UNO DE LOS DOS DADOS MUESTRA 1).

C Y D ES EL LUGAR EN QUE LAS DOS ZONAS SOMBREADAS SE SUPERPONEN (DONDE LOS DOS DADOS MUESTRAN 1).

ESTO ILUSTRA LA **REGLA DE SUMA**: PARA CUALESQUIERA DOS SUCESOS E, F,

$$P(E \text{ O } F) = P(E) + P(F) - P(E \text{ Y } F)$$

SI SUMAMOS P(E) + P(F), SE DOBLAN LOS RESULTADOS ELEMENTALES QUE COMPARTEN E Y F, ASÍ QUE LUEGO HAY QUE RESTAR LO SOBRANTE, QUE ES P(E Y F).

EN EL EJEMPLO ANTERIOR,

$$P(C \text{ O } D) = \frac{11}{36}$$

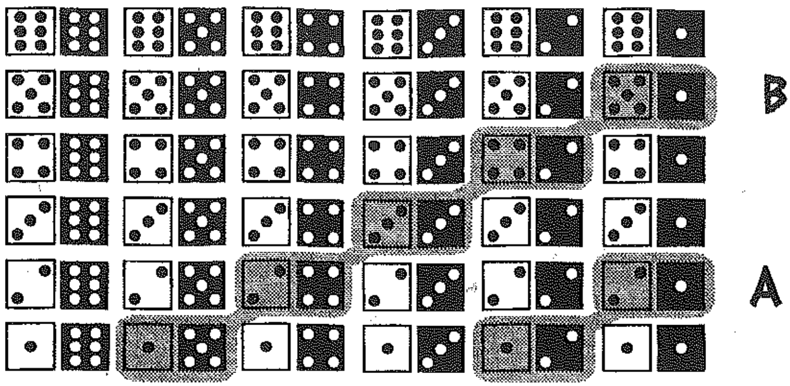
COMO PUEDES VER SI CUENTAS LOS RESULTADOS ELEMENTALES. DE IGUAL FORMA,

$$P(C \text{ Y } D) = \frac{1}{36}$$

Y CONFIRMAMOS LA FÓRMULA:

$$P(C) + P(D) - P(C \text{ Y } D)$$
$$= \frac{6}{36} + \frac{6}{36} - \frac{1}{36} = \frac{11}{36}$$
$$= P(C \text{ O } D)$$

A VECES LA SUPERPOSICIÓN DE **E Y F** ESTÁ VACÍA, Y LOS DOS SUCESOS NO TIENEN RESULTADOS ELEMENTALES EN COMÚN. EN ESE CASO, E Y F SON **EXCLUYENTES** O INCOMPATIBLES Y ENTONCES P(E Y F) = 0. AQUÍ VEMOS LOS SUCESOS EXCLUYENTES **A**, TOTAL DE TIRADA SUMA 3, Y **B**, TOTAL DE TIRADA SUMA 6.



EN EL CASO DE SUCESOS EXCLUYENTES, SE DA UNA **REGLA ESPECIAL DE SUMA**: SI E Y F SON EXCLUYENTES, ENTONCES

P(E O F) = P(E) + P(F)

Y PODEMOS COMPROBAR QUE P(A O B) = $\frac{7}{36} = \frac{2}{36} + \frac{5}{36}$ = P(A)+P(B)

POR ÚLTIMO, LA **REGLA DE RESTA**: PARA CUALQUIER SUCESO E,

P(E) = 1− P(NO E)

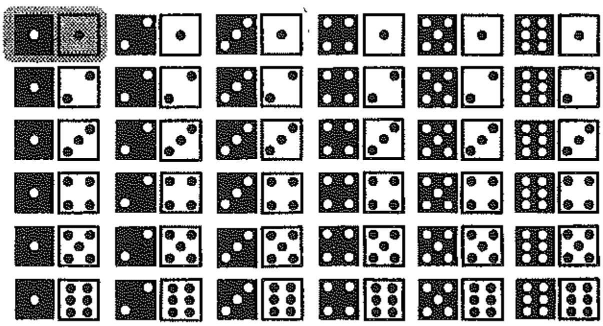
ESTA FÓRMULA RESULTA ÚTIL CUANDO P(NO E) ES MÁS FÁCIL DE CALCULAR QUE P(E). POR EJEMPLO, SI E ES EL SUCESO DE **NO OBTENER** UN DOBLE UNO, ENTONCES EL SUCESO NO E, OBTENER UN DOBLE UNO, TIENE UNA PROBABILIDAD P(NO E) = $\frac{1}{36}$.

ASÍ QUE

P(E) = 1 − P(NO E)

$= 1 - \frac{1}{36}$

$= \frac{35}{36}$

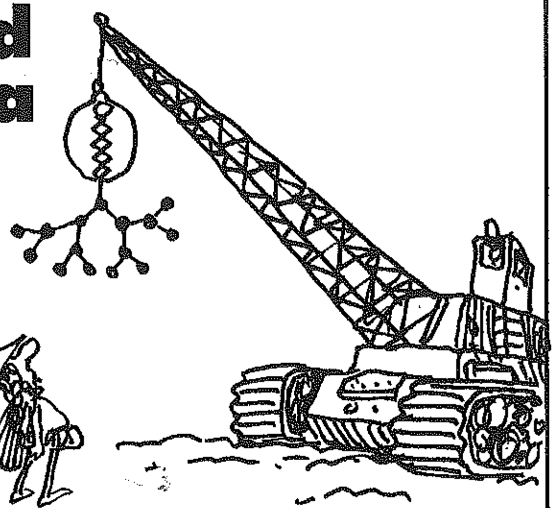


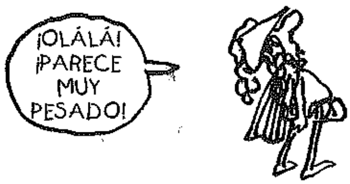
LAS FÓRMULAS QUE HEMOS ESTABLECIDO SON, DE HECHO, ADECUADAS PARA CONTESTAR LA PREGUNTA DE DE MERE, PERO NO SIN DIFICULTAD (PUEDES INTENTAR UTILIZARLAS CON UNA PREGUNTA MÁS SIMPLE: ¿CUÁL ES LA PROBABILIDAD DE SACAR AL MENOS UN SEIS EN DOS TIRADAS CON UN SOLO DADO?). ¡NECESITAMOS MÁS **MAQUINARIA**!

ASÍ QUE PRESENTAMOS LA

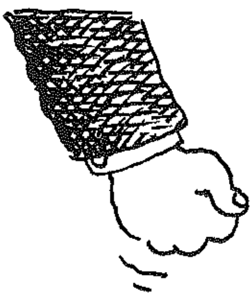
probabilidad condicionada

(¡UN CONCEPTO ESENCIAL DE LA ESTADÍSTICA!)



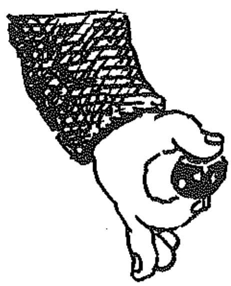


SUPONGAMOS QUE ALTERAMOS UN POCO NUESTRO EXPERIMENTO, Y AHORA LANZAMOS EL DADO BLANCO **ANTES** QUE EL NEGRO. ¿QUÉ PROBABILIDAD HAY DE QUE **LOS DOS SUMEN 3**?



ANTES DE TIRAR LOS DADOS, LA PROBABILIDAD ES

$$P(A) = \frac{2}{36}$$

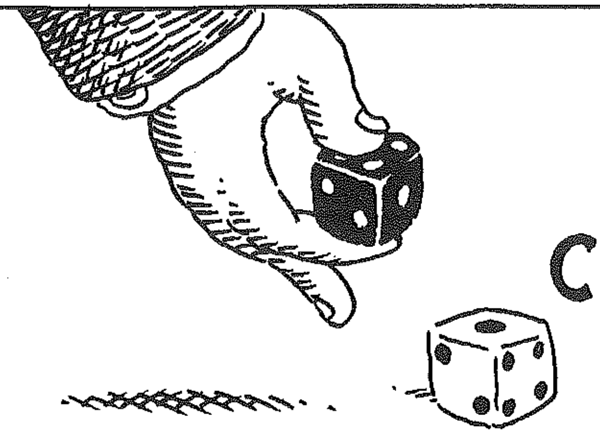


PERO SI SUPONEMOS QUE EN EL DADO BLANCO HA SALIDO UN 1 (SUCESO C), ¿CUÁL ES AHORA LA PROBABILIDAD DE **A**?

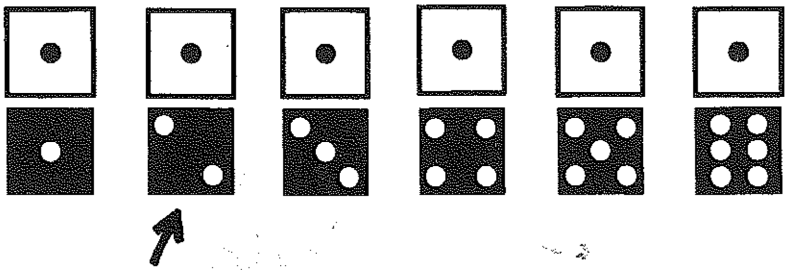
A ESTO SE LE LLAMA LA **PROBABILIDAD CONDICIONADA** DE QUE EL SUCESO A OCURRA, SABIENDO QUE EL SUCESO C YA SE HAYA DADO. ESCRIBIMOS

$$P(A|C)$$

Y DECIMOS «PROBABILIDAD DE A **CONDICIONADA** A C».



ANTES DE TIRAR LOS DADOS, EL ESPACIO MUESTRAL TENÍA 36 RESULTADOS, PERO AHORA QUE EL SUCESO C HA OCURRIDO, EL RESULTADO PERTENECE AL **ESPACIO MUESTRAL C REDUCIDO.**



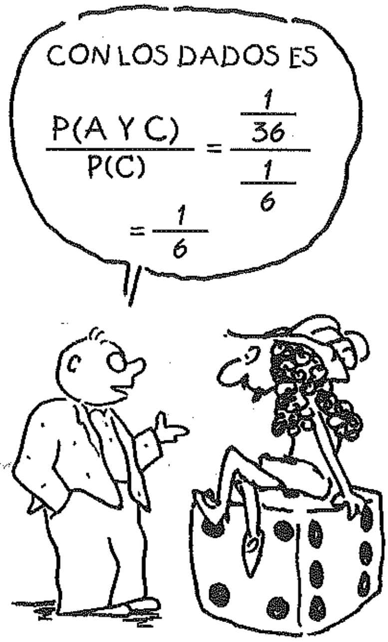
EN EL ESPACIO MUESTRAL REDUCIDO DE SEIS RESULTADOS ELEMENTALES, SÓLO UNO (1, 2) SUMA 3. ASÍ QUE LA PROBABILIDAD CONDICIONADA ES 1/6.

EN GENERAL, PARA ENCONTRAR LA PROBABILIDAD CONDICIONADA $P(E|F)$, CONTEMPLAMOS LOS SUCESOS E Y F COMO PARTE DEL ESPACIO MUESTRAL F REDUCIDO.

TRADUCIMOS ESTO A UNA DEFINICIÓN FORMAL:

LA PROBABILIDAD **DE E CONDICIONADA A F** ES

$$P(E|F) = \frac{P(E \text{ y } F)}{P(F)}$$



CON ESTA FÓRMULA SE PUEDEN VERIFICAR ALGUNOS HECHOS INTUITIVOS:

$$P(E|E) = 1$$ (UNA VEZ **E** HA OCURRIDO, YA ES SEGURO.)

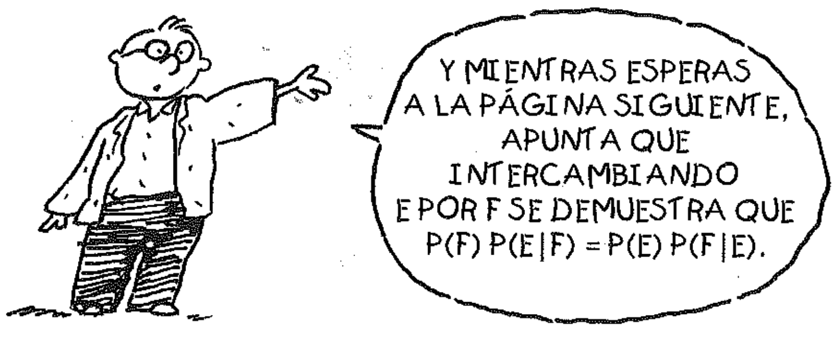
CUANDO E Y F SON EXCLUYENTES,

$$P(E|F) = 0$$ (UNA VEZ **F** HA OCURRIDO, **E** ES IMPOSIBLE.)

SI REESCRIBIMOS LA DEFINICIÓN, OBTENEMOS LA **REGLA DE LA MULTIPLICACIÓN:**

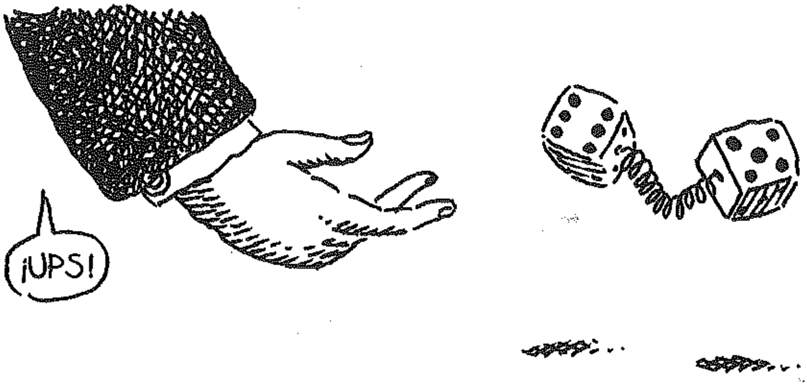
$$P(E \text{ Y } F) = P(E|F)P(F)$$

QUE NOS GUSTARÍA REDUCIR HASTA UNA REGLA «ESPECIAL» DE MULTIPLICACIÓN, BAJO LAS CONDICIONES FAVORABLES DE QUE $P(E|F) = P(E)$. ¡SERÍA FANTÁSTICO!



LA INDEPENDENCIA y la regla especial de multiplicación.

DOS SUCESOS E Y F SON **INDEPENDIENTES** SI LA APARICIÓN DE UNO **NO INFLUYE** EN LA PROBABILIDAD DEL OTRO. POR EJEMPLO, LA TIRADA DE UN DADO NO TIENE NINGÚN EFECTO SOBRE LA DEL OTRO (¡A NO SER QUE ESTÉN PEGADOS, UNIDOS POR UN IMÁN, ETC.!).



EN TÉRMINOS DE PROBABILIDAD CONDICIONADA, ESTO EQUIVALE A DECIR QUE **$P(E) = P(E|F)$** O, DE IGUAL FORMA, **$P(F) = P(F|E)$**. CUANDO E Y F SON INDEPENDIENTES, TENEMOS UNA **REGLA ESPECIAL DE MULTIPLICACIÓN:**

$$P(E \text{ Y } F) = P(E)P(F)$$

VAMOS A VERIFICAR LA INDEPENDENCIA DE LOS DADOS UTILIZANDO ESTAS FÓRMULAS. **C** ES EL SUCESO EL **DADO BLANCO DÉ *1***; **D** ES EL SUCESO EL **DADO NEGRO DÉ *1***, Y TENEMOS:

$$P(C|D) = \frac{P(C \text{ Y } D)}{P(D)} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6} = P(C)$$

¡PERO QUE EL DADO BLANCO DÉ *1* AFECTA CLARAMENTE A LAS POSIBILIDADES DE QUE LOS DOS DADOS SUMEN 3!

$$P(A|C) = \frac{P(A \text{ Y } C)}{P(C)} = \frac{P(1,2)}{P(C)} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6} \neq P(A) = \frac{1}{18}$$

ASÍ QUE ESTOS DOS SUCESOS **NO SON INDEPENDIENTES.**

ANTES DE CONTINUAR, VAMOS A RESUMIR LAS REGLAS QUE HEMOS ACUMULADO:

REGLA DE SUMA:

P(E O F) = P(E) + P(F) - P(E Y F)

REGLA ESPECIAL DE LA SUMA: CUANDO E Y F SON MÚTUAMENTE EXCLUYENTES,

P(E O F) = P(E) + P(F)

REGLA DE LA RESTA:

P(E) = 1 - P(NO E)

REGLA DE LA MULTIPLICACIÓN:

P(E Y F) = P(E|F)P(F)

REGLA ESPECIAL DE LA MULTIPLICACIÓN: CUANDO E Y F SON INDEPENDIENTES,

P(E Y F) = P(E)P(F)

Y, POR FIN, EL PROBLEMA DE DE MERE... VAMOS A SUPONER QUE EL SUCESO E ES CONSEGUIR AL MENOS UN SEIS EN CUATRO TIRADAS DE UN SOLO DADO. ¿CUÁNTO ES P(E)? ESTE ES UNO DE LOS CASOS EN LOS QUE ES MÁS SENCILLO DESCRIBIR EL NEGATIVO: **NO E** ES EL SUCESO **NO CONSEGUIR UN SEIS EN CUATRO TIRADAS.**

SI A_i ES EL SUCESO NO CONSEGUIR UN SEIS EN LA TIRADA NÚMERO i, SABEMOS QUE $P(A_i) = \frac{5}{6}$. TAMBIÉN SABEMOS QUE LAS TIRADAS SON INDEPENDIENTES, ASÍ QUE

$$P(\text{NO } E) =$$

$$P(A_1 \text{ Y } A_2 \text{ Y } A_3 \text{ Y } A_4)$$

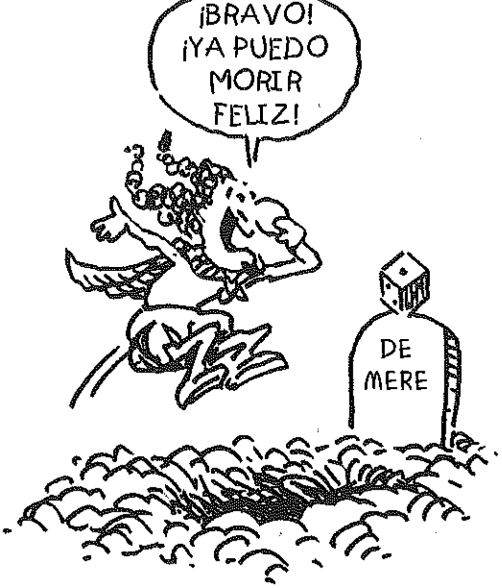
REGLA DE MULTIPLICACIÓN →

$$= \left(\frac{5}{6}\right)^4 = 0{,}482,$$

ENTONCES,

$$P(E) = 1 - P(\text{NO } E) = 0{,}518$$

AHORA, A POR LA SEGUNDA PARTE: F ES EL SUCESO CONSEGUIR AL MENOS UN DOBLE SEIS EN 24 TIRADAS. DE NUEVO, **NO** F ES MÁS SENCILLO DE DESCRIBIR: ES EL SUCESO NO CONSEGUIR NINGÚN DOBLE SEIS.



SI B_i ES EL SUCESO, NO CONSEGUIR NINGÚN DOBLE SEIS EN LA TIRADA NÚMERO i, ASÍ QUE NO F = B_1 Y B_2 Y ... B_{24}.
LA PROBABILIDAD DE CADA B ES

$$P(B_i) = \frac{35}{36}$$, ASÍ QUE

$$P(\text{NO } F) = \left(\frac{35}{36}\right)^{24} = 0{,}509$$

(SEGÚN LA REGLA DE LA MULTIPLICACIÓN), Y PODEMOS LLEGAR A LA CONCLUSIÓN DE QUE

$$P(F) = 1 - P(\text{NO } F) = 1 - 0{,}509$$

$$= 0{,}491$$

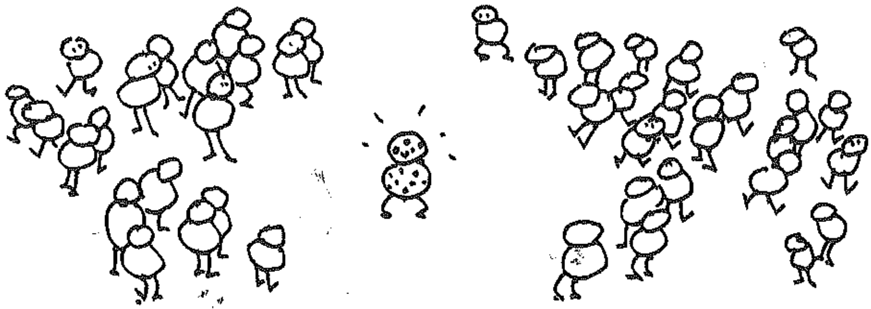
DE MERE LE DIJO A PASCAL QUE HABÍA OBSERVADO QUE EL SUCESO F APARECÍA CON MENOS FRECUENCIA QUE EL SUCESO E, PERO ERA INCAPAZ DE EXPLICAR POR QUÉ... DE LO QUE DEDUCIMOS QUE DE MERE JUGABA MUY A MENUDO Y **¡ANOTABA SUS JUGADAS!**

VAMOS A DEJARNOS DE CASINOS Y VOLVER AL MUNDO REAL...

EL TEOREMA DE BAYES y el caso de los falsos positivos

PARA CONTEMPLAR APLICACIONES MÁS SERIAS DE LA PROBABILIDAD CONDICIONADA, VAMOS A ADENTRARNOS EN UN TERRENO DE VIDA O MUERTE...

SUPONGAMOS QUE UNA EXTRAÑA ENFERMEDAD INFECCIOSA AFECTA A UNO DE CADA 1.000 HABITANTES DE UNA POBLACIÓN...



Y SUPONGAMOS QUE EXISTE UNA PRUEBA FIABLE, PERO NO INFALIBLE, PARA DETECTAR LA ENFERMEDAD: SI UNA PERSONA HA CONTRAÍDO LA ENFERMEDAD, LA PRUEBA RESULTA POSITIVA EN UN 99% DE LOS CASOS. POR OTRO LADO, LA PRUEBA TAMBIÉN PRODUCE **FALSOS** RESULTADOS **POSITIVOS**. UN 2% DE PACIENTES SANOS TAMBIÉN DAN UN RESULTADO POSITIVO, Y TÚ ACABAS DE RECIBIR EL TUYO: POSITIVO. ¿CUÁL ES LA PROBABILIDAD DE QUE TENGAS LA ENFERMEDAD?

TENEMOS DOS SUCESOS CON LOS QUE TRABAJAR:

A: EL PACIENTE PADECE LA ENFERMEDAD
B: EL PACIENTE DA UN RESULTADO POSITIVO.

LA INFORMACIÓN SOBRE LA EFICACIA DE LAS PRUEBAS SE PUEDE ESCRIBIR:

$P(A) = 0{,}001$ (UN PACIENTE DE CADA 1.000 PADECE LA ENFERMEDAD)

$P(B|A) = 0{,}99$ (LA PROBABILIDAD DE UN RESULTADO POSITIVO, DADA LA ENFERMEDAD, ES DE 0,99)

$P(B|\text{NO } A) = 0{,}02$ (LA PROBABILIDAD DE UN FALSO POSITIVO, DADO UN CASO DE NO INFECCIÓN, ES DE 0,02)

Y NOS PREGUNTAMOS:

$P(A|B) = $ ¿QUÉ? (LA PROBABILIDAD DE PADECER LA ENFERMEDAD, DADO UN RESULTADO POSITIVO)

Y A QUE EL TRATAMIENTO DE LA ENFERMEDAD PRODUCE GRAVES EFECTOS SECUNDARIOS, LA DOCTORA, SU ABOGADA Y EL ABOGADO DE SU ABOGADA LLAMAN A JOE BAYES, QUE TIENE UN CONSULTORIO DE PROBABILIDADES, PARA QUE LES DÉ UNA RESPUESTA. JOE APLICA EL TEOREMA QUE DESARROLLÓ UN ANTEPASADO SUYO, EL RDO. THOMAS **BAYES** (1744-1809).

JOE EMPIEZA CON UNA TABLA DE 2×2, QUE DIVIDE EL ESPACIO MUESTRAL EN CUATRO CASOS EXCLUYENTES. REPRESENTA TODAS LAS COMBINACIONES POSIBLES DEL ESTADO DE LA ENFERMEDAD Y EL RESULTADO DE LA PRUEBA.

	A	NO A
B	A Y B	NO A Y B
NO B	A Y NO B	NO A Y NO B

VAMOS A ENCONTRAR LA PROBABILIDAD DE CADA CASO EN LA TABLA:

	A	NO A	SUMA
B	P(A Y B)	P(NO A Y B)	P(B)
NO B	P(A Y NO B)	P(NO A Y NO B)	P(NO B)
	P(A)	P(NO A)	1

LAS PROBABILIDADES DE LOS MÁRGENES SE CALCULAN SUMANDO FILAS Y COLUMNAS.

AHORA VAMOS A CALCULAR:

$$P(A \text{ Y } B) = P(B|A)\,P(A) = (0{,}99)(0{,}001) = 0{,}00099$$
$$P(\text{NO } A \text{ Y } B) = (B|\text{NO } A)\,P(\text{NO } A) = (0{,}02)(0{,}999) = 0{,}01998$$

Y ASÍ PODEMOS RELLENAR ALGUNAS ENTRADAS:

	A	NO A	SUMA
B	0,00099	0,01998	0,02097
NO B	P(A Y NO B)	P(NO A Y NO B)	P(NO B)
	0,001	0,999	1

Y ENCONTRAMOS LAS OTRAS PROBABILIDADES RESTANDO EN CADA COLUMNA Y DESPUÉS SUMANDO CADA FILA.

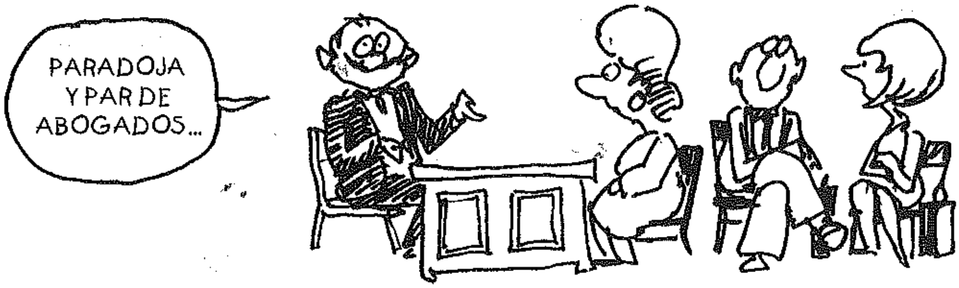
LA TABLA FINAL ES:

	A	NO A		
B	0,00099	0,01998	0,02097	P(B)
NO B	0,00001	0,97902	0,97903	P(NO B)
	0,001	0,999	1	
	P(A)	P(NO A)		

Y DE AHÍ DEDUCIMOS DIRECTAMENTE QUE

$$P(A|B) = \frac{P(A \text{ Y } B)}{P(B)} = \frac{0{,}0009}{0{,}0209} = 0{,}0472$$

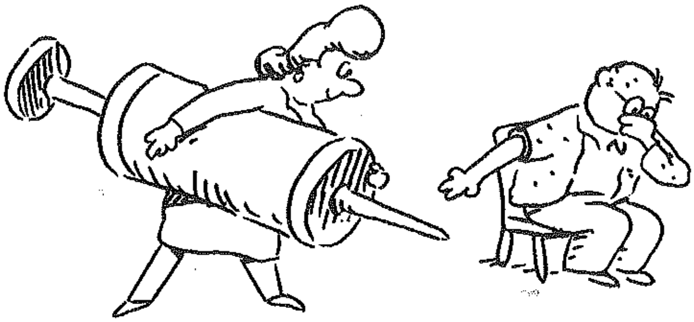
A PESAR DE LA GRAN PRECISIÓN DE LA PRUEBA, **MENOS DE UN 5%** DE LOS QUE DAN POSITIVO PADECEN LA ENFERMEDAD. A ESTO SE LE LLAMA LA **PARADOJA DEL FALSO POSITIVO.**



ESTA TABLA MUESTRA LO QUE PASA EN UN GRUPO DE MIL PACIENTES. COMO MEDIA, SÓLO 21 PACIENTES DARÁN POSITIVO (¡Y SÓLO **UNO** PADECERÁ LA ENFERMEDAD!), Y 20 FALSOS POSITIVOS SE ENCONTRARÁN EN EL **GRAN GRUPO DE NO AFECTADOS.**

	ENFERMEDAD	NO ENFERMEDAD	
RESULTADO POSITIVO	1	20	21
RESULTADO NEGATIVO	0	979	979
	1	999	1.000

¿QUÉ DEBE HACER LA DOCTORA? JOE BAYES LE ACONSEJA NO EMPEZAR EL TRATAMIENTO BASÁNDOSE SÓLO EN ESA ÚNICA PRUEBA. SIN EMBARGO, LA PRUEBA APORTA CIERTA INFORMACIÓN: CON UN RESULTADO POSITIVO, LAS PROBABILIDADES DE QUE EL PACIENTE PADEZCA LA ENFERMEDAD HAN AUMENTADO DE *1* ENTRE *1.000* A *1* ENTRE 23. LA DOCTORA CONTINÚA HACIENDO OTRAS PRUEBAS.



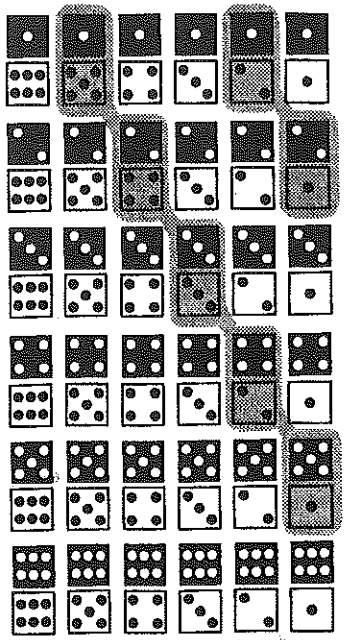
JOE BAYES COBRA EL CHEQUE POR LA CONSULTA ANTES DE CONFESAR QUE TODOS ESOS PASOS QUE HA DADO SE PUEDEN COMPRIMIR EN UNA SOLA FÓRMULA, EL TEOREMA DE BAYES:

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A)+P(\text{NO } A)P(B|\text{NO } A)}$$

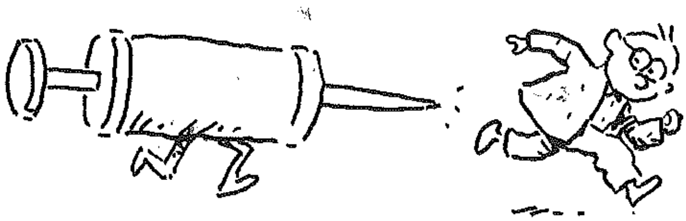
ESTA FÓRMULA CALCULA P(A|B) A PARTIR DE P(A) Y LAS DOS PROBABILIDADES CONDICIONADAS P(B|A) Y P(B|NO A). SE PUEDE CALCULAR TENIENDO EN CUENTA QUE ESA GRAN FRACCIÓN TAMBIÉN SE PUEDE EXPRESAR COMO

$$\frac{P(A \text{ y } B)}{P(A \text{ y } B)+P(\text{NO } A \text{ y } B)} = \frac{P(A \text{ y } B)}{P(B)} = P(A|B)$$

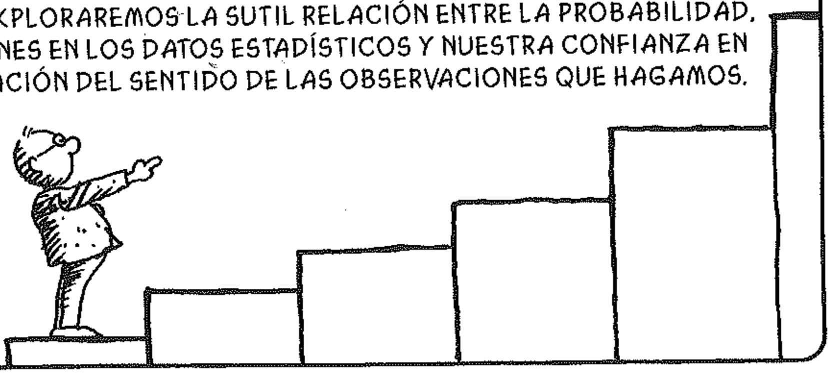
EN ESTE CAPÍTULO, HEMOS HABLADO DE LOS ASPECTOS ESENCIALES DE LA PROBABILIDAD: SU DEFINICIÓN, LOS ESPACIOS MUESTRALES Y LOS RESULTADOS ELEMENTALES, LA PROBABILIDAD CONDICIONADA Y ALGUNAS FÓRMULAS BÁSICAS PARA CALCULAR LAS PROBABILIDADES. HEMOS ILUSTRADO ESTAS IDEAS CON UN ESPACIO MUESTRAL DE DOS DADOS. PARA EL JUGADOR MODERNO, LA PROBABILIDAD ES UNA PODEROSA HERRAMIENTA DE ELECCIÓN.



POR ÚLTIMO, EN EL EJEMPLO MÉDICO, HEMOS ENSEÑADO CÓMO ESTAS IDEAS ABSTRACTAS PUEDEN AYUDAR A TOMAR BUENAS DECISIONES CUANDO SE TIENE **INFORMACIÓN IMPERFECTA** Y **RIESGOS REALES**; EL **OBJETIVO ÚLTIMO** DE LA ESTADÍSTICA.



PERO ESTO NO ES MÁS QUE EL PRINCIPIO. PARA NOSOTROS, LA PROBABILIDAD ES SÓLO UNA **HERRAMIENTA** (UNA HERRAMIENTA ESENCIAL, POR SUPUESTO) PARA EL ESTUDIO DE LA ESTADÍSTICA. EN LOS CAPÍTULOS SIGUIENTES, EXPLORAREMOS LA SUTIL RELACIÓN ENTRE LA PROBABILIDAD, LAS VARIACIONES EN LOS DATOS ESTADÍSTICOS Y NUESTRA CONFIANZA EN LA INTERPRETACIÓN DEL SENTIDO DE LAS OBSERVACIONES QUE HAGAMOS.



♦ Capítulo 4 ♦
VARIABLES ALEATORIAS

EN EL CAPÍTULO 2, VIMOS QUE LAS OBSERVACIONES BASADAS EN DATOS NUMÉRICOS, COMO LOS PESOS DE LOS ESTUDIANTES, SE PUEDEN REPRESENTAR MEDIANTE GRÁFICOS Y RESUMIR EN TÉRMINOS DE PUNTO MEDIO, DISPERSIÓN, OBSERVACIONES ATÍPICAS, ETC. EN EL CAPÍTULO 3, HEMOS VISTO CÓMO SE PUEDEN ASIGNAR PROBABILIDADES A LOS RESULTADOS DE UN EXPERIMENTO ALEATORIO.

SI IMAGINAMOS QUE UN EXPERIMENTO ALEATORIO SE REPITE MUCHAS VECES, ESPERAMOS QUE LOS RESULTADOS ACABEN OBEDECIENDO A SUS PROBABILIDADES. LA PROBABILIDAD CONFORMA UN **MODELO** PARA EXPERIMENTOS REALES... ASÍ QUE, ¿POR QUÉ NO HACEMOS CON EL MODELO LO QUE YA HEMOS HECHO CON LOS DATOS QUE DESCRIBE?

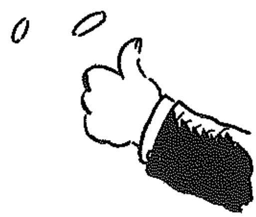
La idea principal es la **variable aleatoria**, que escribimos con una X mayúscula.

Una variable aleatoria se define como el **resultado numérico de un experimento aleatorio.**

Por ejemplo, imaginemos que escogemos a un estudiante al azar de todo el grupo. Ése es el experimento aleatorio. **Altura, peso, ingresos familiares, nota de selectividad y nota media** serían las **variables** numéricas que describen a este estudiante. Todas ellas son **variables aleatorias.**

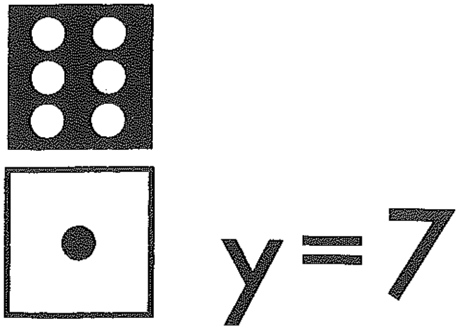
Otro ejemplo: lanza dos monedas (el experimento aleatorio) y anota el **número** de caras: *0, 1,* o *2.*

RESULTADO	XX	CX XC	CC
x	0	1	2



¡Cuidado con la notación! La variable se escribe con una X mayúscula. La fila inferior, x, representa un solo valor de X, por ejemplo $x = 2$, si salen dos caras.

OTRO EJEMPLO SERÍA EL DE LA FAMOSA TIRADA DE DADOS. Y REPRESENTA LA SUMA DE LOS PUNTOS DE LOS DOS DADOS. EN ESTA VARIABLE ALEATORIA, Y PUEDE SER CUALQUIER NÚMERO ENTRE *2* Y *12*.



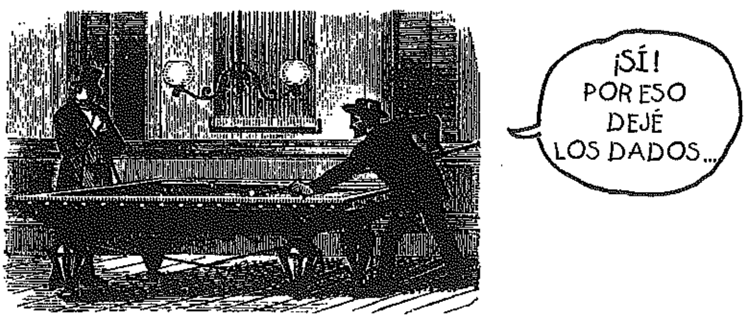
AHORA QUEREMOS SABER LAS **PROBABILIDADES** DE LOS RESULTADOS. PARA LA PROBABILIDAD DE QUE LA VARIABLE X TENGA EL VALOR *x*, ESCRIBIMOS PR(X = *x*), O SIMPLEMENTE *P(x)*. CON LA VARIABLE X DEL LANZAMIENTO DE LA MONEDA PODEMOS CONFECCIONAR UNA TABLA:

x	0	1	2
$Pr(X=x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

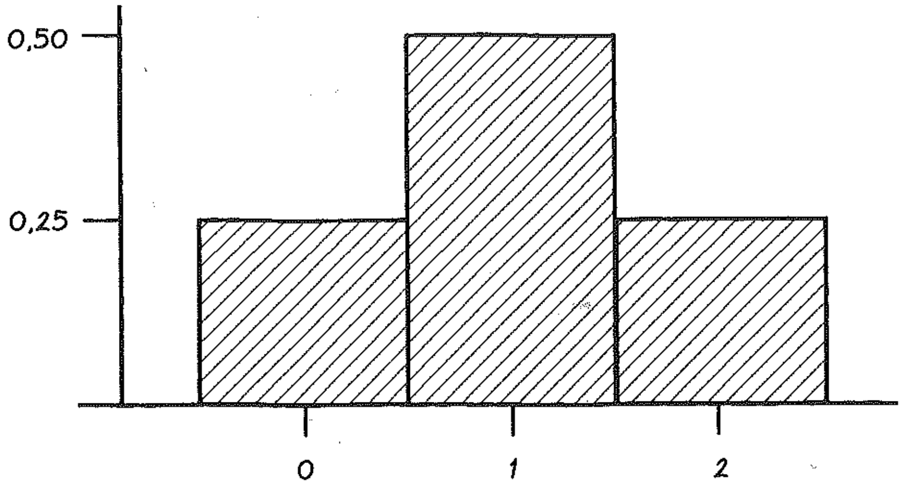
ESTA TABLA SE LLAMA LA **DISTRIBUCIÓN DE PROBABILIDAD** DE LA VARIABLE ALEATORIA X.

PARA LA VARIABLE ALEATORIA Y (LA SUMA DE LOS DOS DADOS), LA DISTRIBUCIÓN DE PROBABILIDAD ES ASÍ:

y	2	3	4	5	6	7	8	9	10	11	12
$Pr(Y=y)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

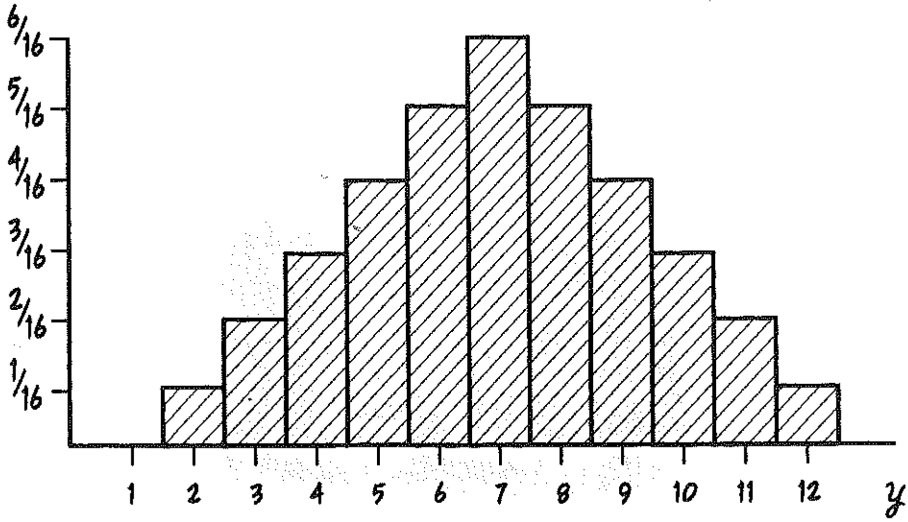


AHORA VAMOS A DIBUJAR GRÁFICOS, O **HISTOGRAMAS**, PARA REPRESENTAR ESTAS DISTRIBUCIONES DE PROBABILIDAD. POR CADA VALOR DE X, DIBUJAMOS UNA BARRA CON LA ALTURA IGUAL A $p(x)$.

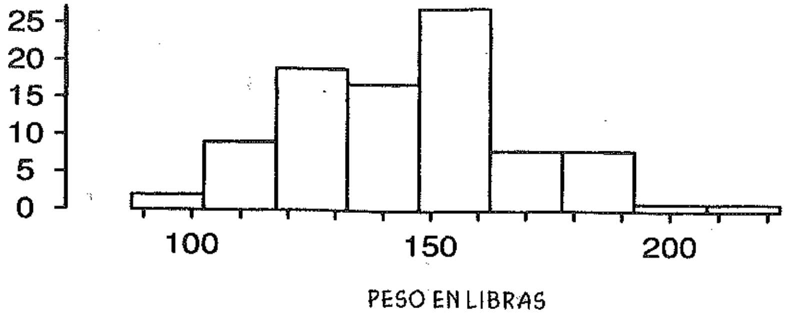


ES FÁCIL VER QUE EL ÁREA TOTAL DE LAS CAJAS ES *1*: TODAS LAS CAJAS TIENEN BASE *1* Y ALTURA $p(x)$, ASÍ QUE EL ÁREA TOTAL ES LA SUMA DE LAS PROBABILIDADES DE TODOS LOS RESULTADOS, ES DECIR, *1*.

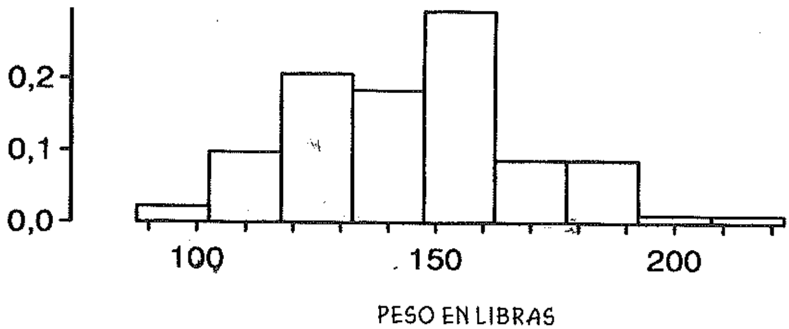
ÉSTE ES EL HISTOGRAMA DE LA PROBABILIDAD DE LA VARIABLE ALEATORIA Y, Y MUESTRA LA DISTRIBUCIÓN DE PROBABILIDAD DE LA SUMA DE DOS DADOS:



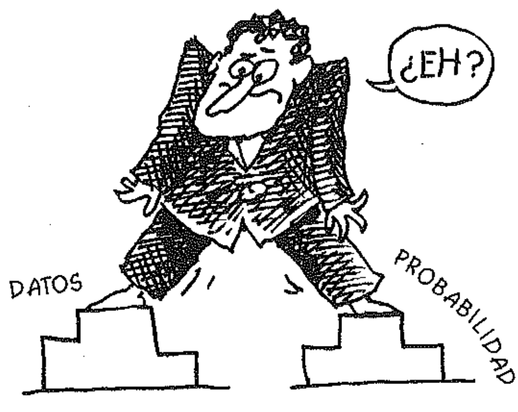
¿POR QUÉ LLAMAMOS A TODOS ESTOS GRÁFICOS HISTOGRAMAS? SEGURO QUE RECUERDAS QUE, EN EL CAPÍTULO 2, UN HISTOGRAMA ERA UN GRÁFICO EN EL QUE SE REPRESENTABA CUÁNTOS DATOS PERTENECÍAN A CADA INTERVALO:



A PARTIR DE ESTE HISTOGRAMA DE **FRECUENCIAS**, DESARROLLAMOS EL HISTOGRAMA DE LA FRECUENCIA **RELATIVA**, EN EL QUE SE VEÍA LA **PROPORCIÓN** DE DATOS QUE TENÍA CADA INTERVALO:



PERO TAMBIÉN RECORDARÁS QUE, SEGÚN UNA DE SUS DEFINICIONES, LA PROBABILIDAD ES LA **FRECUENCIA RELATIVA DE UN SUCESO «A LARGO PLAZO»**. SI REPETIMOS EL EXPERIMENTO ALEATORIO MUCHAS VECES, EL HISTOGRAMA DE LA **FRECUENCIA RELATIVA** DE LOS RESULTADOS DEBERÍA PARECERSE MUCHO AL HISTOGRAMA DE **PROBABILIDAD** DE LA VARIABLE ALEATORIA.

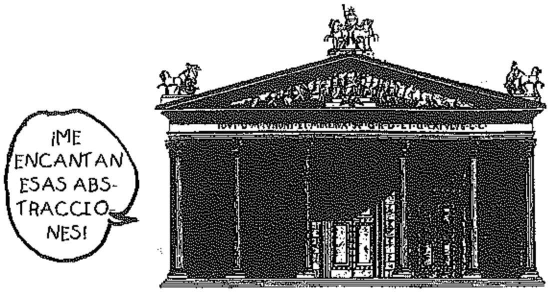


YA CONOCEMOS LA DISTRIBUCIÓN DE PROBABILIDAD DE X, Y TAMBIÉN QUE LAS TIRADAS DE LAS MONEDAS SE CORRESPONDERÁN MÁS O MENOS CON LAS PROBABILIDADES. DESPUÉS DE 1.000 TIRADAS, LA LANZADORA LOCA HACE UN RECUENTO DE LOS DATOS:

MODELO DE PROBABILIDAD	DATOS OBSERVADOS		
$p(x)$	x	n_x = NÚMERO DE OCURRENCIAS	$\frac{n_x}{n}$ = FRECUENCIA RELATIVA
0,25	0	260	0,260
0,5	1	517	0,517
0,25	2	223	0,223

Y ASÍ VEMOS QUE EL HISTOGRAMA DE PROBABILIDAD DE X ES COMO LA «FORMA PURA», O MODELO DEL HISTOGRAMA DE FRECUENCIA RELATIVA DE LOS DATOS.

PARA EXTENDER LA ANALOGÍA ENTRE LA FRECUENCIA RELATIVA Y LOS DATOS, DEBERÍAMOS HABLAR AHORA DE LA MEDIA Y LA VARIANZA (O DESVIACIÓN TÍPICA) DE UNA DISTRIBUCIÓN DE PROBABILIDAD...



Y PARA RECORDAR QUE NOS ENCONTRAMOS EN EL REINO DE LO ABSTRACTO, VAMOS A SOLTAR UNAS CUANTAS **LETRAS GRIEGAS**...

MEDIA Y VARIANZA DE LAS VARIABLES ALEATORIAS

UTILIZAMOS TERMINOLOGÍA Y SÍMBOLOS ESPECIALES PARA DISTINGUIR LAS PROPIEDADES DE LOS CONJUNTOS DE DATOS DE LAS DISTRIBUCIONES DE PROBABILIDAD:

LAS PROPIEDADES DE LOS DATOS SE LLAMAN PROPIEDADES MUESTRALES O ESTADÍSTICOS, MIENTRAS QUE LAS PROPIEDADES DE LA DISTRIBUCIÓN DE PROBABILIDAD SE LLAMAN PARÁMETROS DEL **MODELO** O **POBLACIONALES**. PARA LA MEDIA POBLACIONAL UTILIZAMOS LA LETRA GRIEGA μ (MU), Y σ (SIGMA MINÚSCULA) PARA LA DESVIACIÓN TÍPICA POBLACIONAL. (PARA LOS DATOS, UTILIZAMOS LOS SÍMBOLOS ROMANOS $\bar{x}$ Y s.)

LA MEDIA MUESTRAL SE DEFINÍA CON LA ECUACIÓN

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

PUEDE QUE ALGUNOS DE ESTOS DATOS x_i TENGAN VALORES IGUALES. ACUÉRDATE DE LA LANZADORA LOCA DE MONEDAS: LOS ÚNICOS VALORES POSIBLES ERAN 0, 1 Y 2, Y EFECTUÓ 1.000 TIRADAS. EL VALOR 0 RESULTÓ EN 260 OCASIONES, UNA CARA EN 517, Y DOS CARAS EN 223 OCASIONES.

YA QUE x PUEDE TENER TODOS LOS VALORES DE X, n_x ES EL NÚMERO DE DATOS CON VALOR x. ENTONCES PODEMOS REESCRIBIR LA FÓRMULA COMO

$$\bar{x} = \frac{1}{n}\sum_{\text{TODAS LAS } x} n_x x$$

O COMO

$$\bar{x} = \sum_{\text{TODAS LAS } x} x\frac{n_x}{n}$$

¡AH! PERO AHORA $\frac{n_x}{n}$ ES LA FRECUENCIA RELATIVA... LA «PROBABILIDAD APROXIMADA»... EL NÚMERO QUE SE ACERCA A $p(x)$... LUEGO, POR ANALOGÍA, FORMAMOS LA EXPRESIÓN

$$\sum_{\text{TODAS LAS } x} x\,p(x)$$

Y LA DEFINIMOS COMO **MEDIA DE LA DISTRIBUCIÓN DE PROBABILIDAD.**

DEFINICIÓN:
LA **media** DE LA VARIABLE ALEATORIA X SE DEFINE COMO

$$\mu = \sum_{\text{TODAS LAS } x} x\,p(x)$$

A ESTO SE LE LLAMA EL **VALOR ESPERADO** DE X, O E[X]. IMAGINA QUE ES LA SUMA DE TODOS LOS VALORES POSIBLES, CADA UNO PONDERADO POR SU PROBABILIDAD.

EL EXPERIMENTO DE LA LANZADORA LOCA DE MONEDAS NOS PERMITE COMPARAR SU MEDIA MUESTRAL $\bar{x}$ CON NUESTRA MEDIA POBLACIONAL μ:

MUESTRA

x	$\frac{n_x}{n}$	$x\frac{n_x}{n}$
0	0,26	0
1	0,517	0,517
2	0,223	0,446
		$0{,}963 = \bar{x}$

MODELO

x	$p(x)$	$xp(x)$
0	0,25	0
1	0,5	0,5
2	0,25	0,5
		$1 = \mu$

AHORA VAMOS A HACER LO MISMO CON LA **VARIANZA**. A LO MEJOR RECUERDAS LA FÓRMULA

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

MIDE (CASI) LA DISTANCIA CUADRÁTICA MEDIA ENTRE LOS DATOS Y LA MEDIA. IGUAL QUE ANTES, LA PODEMOS REFORMULAR:

$$s^2 = \sum_{\text{TODAS LAS } x}(x-\bar{x})^2\frac{n_x}{n-1}$$

SALVO ESE MOLESTO DENOMINADOR $n-1$ EN LUGAR DE n, ESTA FÓRMULA TAMBIÉN PARECE UNA SUMA PONDERADA DE DISTANCIAS AL CUADRADO... ASÍ QUE FORMULAMOS OTRA DEFINICIÓN:

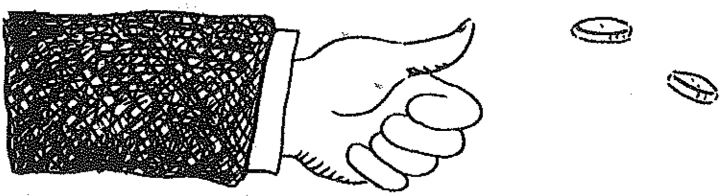
LA **varianza** DE UNA VARIABLE ALEATORIA X ES LA ESPERADA DEL CUADRADO DE LA DISTANCIA ENTRE LOS POSIBLES VALORES DE X Y LA MEDIA POBLACIONAL:

$$\sigma^2 = \sum_{\text{TODAS LAS } x} (x-\mu)^2 p(x)$$

LA **desviación típica** σ ES LA RAÍZ CUADRADA DE LA VARIANZA.

AHORA UTILIZAMOS LA TABLA DE LA PÁGINA ANTERIOR PARA ENCONTRAR LA VARIANZA DE UNA TIRADA DE DOS MONEDAS (EN LA QUE $\mu = 1$).

x	$p(x)$	$(x-\mu)^2 p(x)$
0	0,25	$(0-1)^2\, 0{,}25 = 0{,}25$
1	0,5	$(1-1)^2\, 0{,}50 = 0$
2	0,25	$(2-1)^2\, 0{,}25 = 0{,}25$
TOTAL		$0{,}50 = \sigma^2$



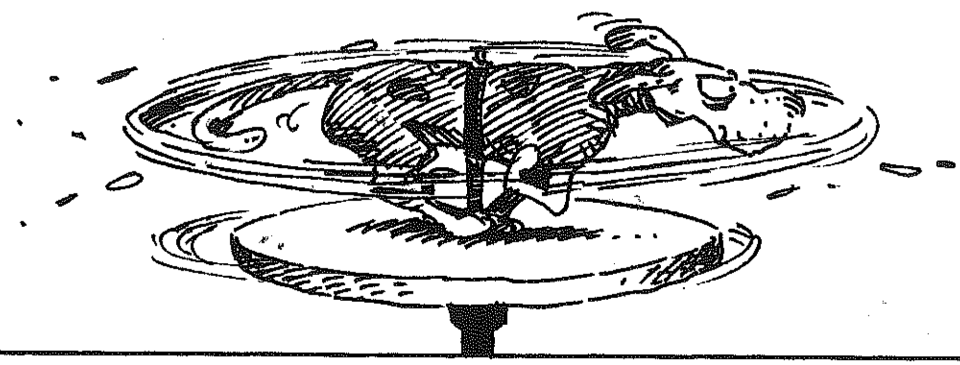
EN RESUMEN: μ Y σ, LA **MEDIA** Y LA **DESVIACIÓN TÍPICA** POBLACIONALES, SON PARÁMETROS QUE PODEMOS CALCULAR A PARTIR DE LAS **DISTRIBUCIONES DE PROBABILIDAD**. SON TOTALMENTE ANÁLOGAS A LA MEDIA MUESTRAL $\bar{x}$ Y A LA DESVIACIÓN TÍPICA s DE LOS DATOS MUESTRALES.

HASTA AHORA, NUESTROS EJEMPLOS HAN CONSISTIDO EN VARIABLES ALEATORIAS **DISCRETAS**. SUS RESULTADOS SON UN CONJUNTO DE VALORES AISLADOS («DISCRETOS»), COMO LOS QUE VIMOS EN EL CAPÍTULO 3, PERO TAMBIÉN HAY

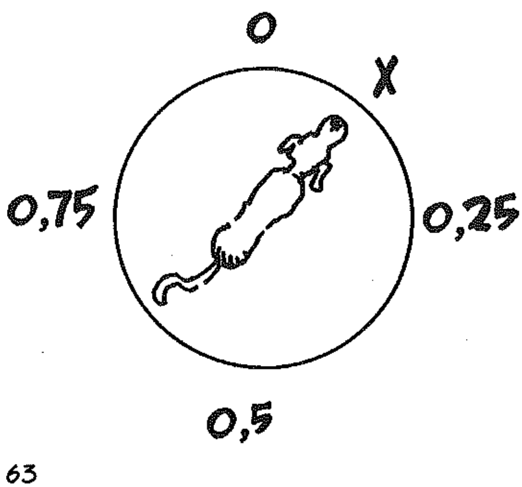
variables aleatorias continuas

IMAGINEMOS UN EXPERIMENTO ALEATORIO EN EL QUE **TODOS LOS RESULTADOS TENGAN PROBABILIDAD CERO**. ESO ES, $P(x) = 0$ PARA CUALQUIER x.

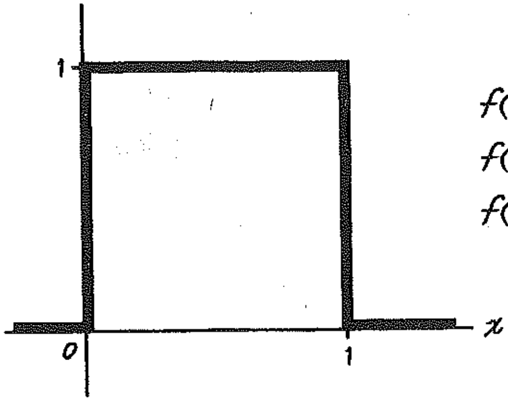
UN EJEMPLO MUY SIMPLE ES EL DE UN PERRO DE CAZA SOBRE UNA SUPERFICIE CIRCULAR QUE GIRA EN EQUILIBRIO. PUEDE PARAR EN CUALQUIER PUNTO DEL CÍRCULO. SI **X** REPRESENTA LA PROPORCIÓN DE TODA LA CIRCUNFERENCIA EN LA QUE SE ENCUENTRA, LA VARIABLE ALEATORIA **X** PUEDE TENER CUALQUIER VALOR ENTRE *0* Y *1*; UNA SERIE **INFINITA** DE VALORES.



ALGUNAS PROBABILIDADES SON FÁCILES DE ENCONTRAR, COMO LA PROBABILIDAD DE QUE X ESTÉ DENTRO DE UNA REGIÓN: POR EJEMPLO, $Pr(0{,}25 \leq X \leq 0{,}75) = 0{,}5$, PORQUE ES LA MITAD DEL CÍRCULO. SIN EMBARGO, ¿QUÉ PASA CON $Pr(X = 0{,}5)$? YA QUE X PUEDE REPRESENTAR UN NÚMERO INFINITO DE VALORES, Y TODOS SON IGUAL DE POSIBLES, LA PROBABILIDAD DE QUE X SEA EXACTAMENTE 0,5 (O CUALQUIER OTRO VALOR EXACTAMENTE) ES PRECISAMENTE 0.



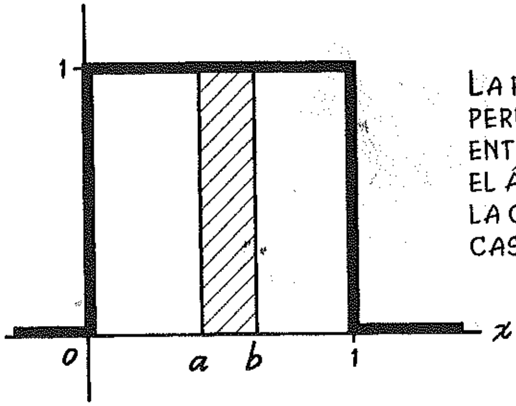
¿CÓMO PODEMOS REPRESENTARLO EN UN DIBUJO? POR ANALOGÍA CON EL CASO DE LAS PROBABILIDADES DISCRETAS, INTENTAMOS OBSERVAR LAS PROBABILIDADES CONTINUAS COMO **ÁREAS BAJO ALGO**. EN EL CASO DEL PERRO GIRATORIO, ESE «ALGO» TIENE ESTE ASPECTO:



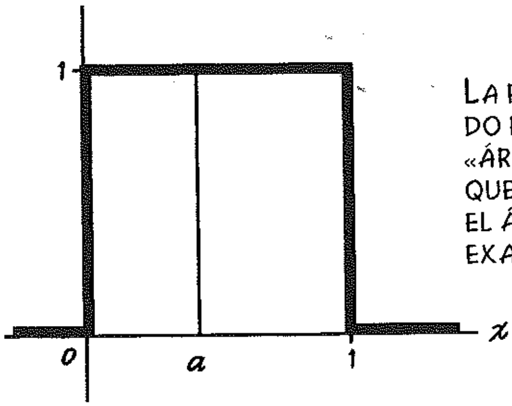
$f(x) = 0$ CUANDO $x < 0$

$f(x) = 1$ CUANDO $0 \leqslant x \leqslant 1$

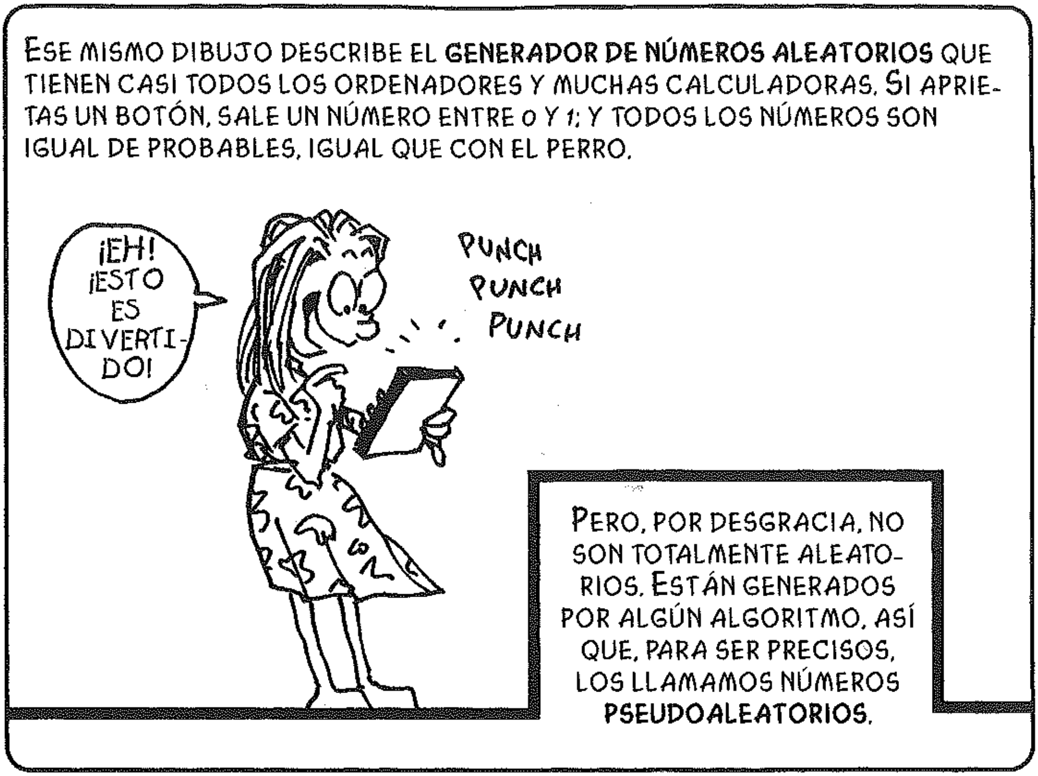
$f(x) = 0$ CUANDO $x > 1$



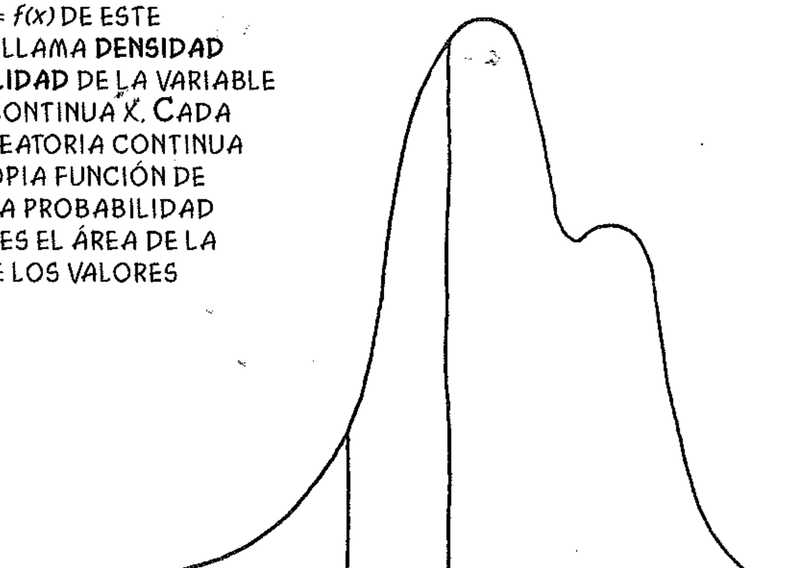
LA PROBABILIDAD DE QUE EL PERRO SEÑALE CUALQUIER LUGAR ENTRE a Y b ES PRECISAMENTE EL ÁREA SOMBREADA BAJO LA CURVA ENTRE a Y b (EN ESTE CASO, $b - a$).



LA PROBABILIDAD DE UN RESULTADO EXACTO, SIN EMBARGO, ES EL «ÁREA» QUE HAY SOBRE UN PUNTO, QUE ES **CERO**. (TEN EN CUENTA QUE EL ÁREA TOTAL DE LA CURVA ES EXACTAMENTE *1*.)

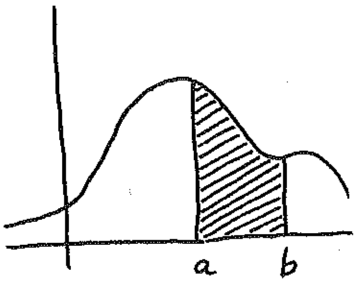


LA CURVA $y = f(x)$ DE ESTE EJEMPLO SE LLAMA **DENSIDAD DE PROBABILIDAD** DE LA VARIABLE ALEATORIA CONTINUA X. CADA VARIABLE ALEATORIA CONTINUA TIENE SU PROPIA FUNCIÓN DE DENSIDAD. LA PROBABILIDAD $\Pr(a \leq X \leq b)$ ES EL ÁREA DE LA CURVA ENTRE LOS VALORES DE x, a Y b.

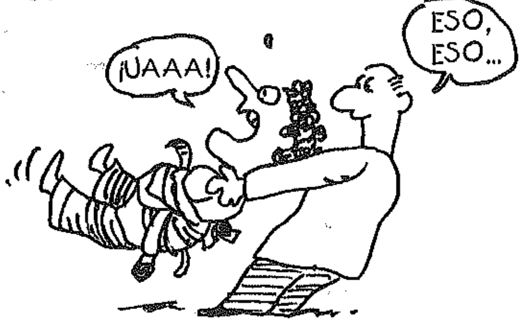


EN GENERAL, LA DENSIDAD DE PROBABILIDAD NO ES TAN SIMPLE, Y A VECES, CALCULAR EL ÁREA NO TIENE NADA DE TRIVIAL.

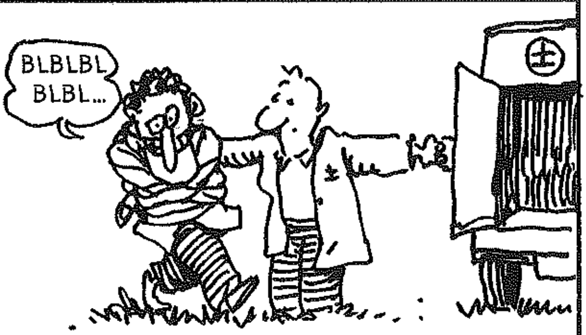
$$\int_a^b f(x)\,dx$$



NOS VEMOS OBLIGADOS A UTILIZAR NOTACIÓN DE CÁLCULO PARA DESCRIBIR EL ÁREA DE LA CURVA *f(x)*. ESTE SÍMBOLO SE LEE «INTEGRAL DE *f* DESDE *a* HASTA *b*».



AL IGUAL QUE LAS PROBABILIDADES DISCRETAS, LAS DENSIDADES CONTINUAS TIENEN DOS PROPIEDADES QUE YA CONOCEMOS:

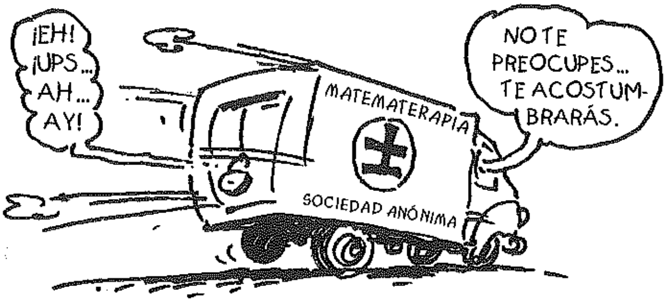


$$f(x) \geq 0$$

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

(INTENTA QUE NO TE ASUSTEN ESOS INFINITOS... SÓLO QUIEREN DECIR QUE OBSERVAMOS TODA EL ÁREA DE LA CURVA, DEL PRINCIPIO AL FINAL, ¡SÓLO QUE NO HAY NI PRINCIPIO NI FINAL!)

A PESAR DE QUE LA NOTACIÓN TE RESULTE EXTRAÑA, NO REPRESENTA MÁS QUE UN **ÁREA**... EL SIGNO DE LA INTEGRAL ES UNA «S» ALARGADA, DE «SUMA», QUE ES MÁS O MENOS LA FUNCIÓN QUE DESEMPEÑA LA INTEGRAL.



COMO ES ALGO PARECIDO A UNA SUMA, LA INTEGRAL SIRVE PARA DEFINIR LA **MEDIA Y LA VARIANZA de una variable aleatoria continua.**

$$\mu = \int_{-\infty}^{\infty} x f(x)\,dx$$

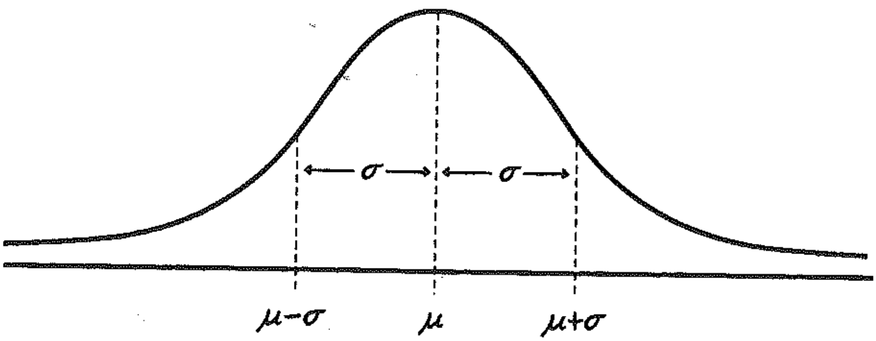
$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

POR ANALOGÍA CON LAS FÓRMULAS DISCRETAS:

$$\mu = \sum_{\text{TODAS LAS } x} x p(x)$$

$$\sigma^2 = \sum_{\text{TODAS LAS } x} (x-\mu)^2 p(x)$$

AUNQUE NO RESULTE OBVIO AL OBSERVAR LAS FÓRMULAS, ESTAS DEFINICIONES DE MEDIA Y VARIANZA SON TOTALMENTE COHERENTES CON SU PAPEL DE CENTRO Y DISPERSIÓN MEDIA DE LAS PROBABILIDADES DADAS POR LA DENSIDAD $f(x)$. ÉSTE ES EL GRÁFICO QUE HAY QUE RECORDAR:



SUMA
de variables aleatorias

UNA VEZ CONOCIDA LA MEDIA Y LA VARIANZA DE UNA VARIABLE ALEATORIA, ¿QUÉ PODEMOS HACER CON ELLAS? BUENO, PARA EMPEZAR, SE PUEDEN BUSCAR LA MEDIA Y LA VARIANZA DE **OTRAS** VARIABLES ALEATORIAS...

POR EJEMPLO, VAMOS A TOMAR EL CASO DEL LANZAMIENTO DE UNA MONEDA. SI SALE CARA, $x = 1$, Y $x = 0$ SI SALE CRUZ.

x	0	1
$p(x)$	0,5	0,5

AHORA DEBERÍAS SER CAPAZ DE ENCONTRAR LA MEDIA

$$E[X] = 0 \cdot p(0) + 1 \cdot p(1)$$
$$= 0 + 0{,}5$$
$$= 0{,}5$$

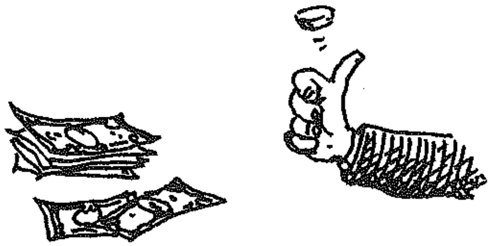
Y LA VARIANZA

$$\sigma^2 = (0-0{,}5)^2 p(0) + (1-0{,}5)^2 p(1)$$
$$= 0{,}25$$

VAMOS A HACER UNA APUESTA: TE JUEGAS 6 DÓLARES Y YO LANZO UNA MONEDA: SI SALE CARA, GANAS 10 DÓLARES, Y CERO SI SALE CRUZ. ENTONCES, TUS GANANCIAS G SON

$$G = 10X - 6$$

¡UNA NUEVA VARIABLE ALEATORIA! ¿CUÁLES SON SU MEDIA Y SU VARIANZA?



SI LO PIENSAS UN POCO TE CONVENCERÁS DE QUE E[G] VIENE DADO POR

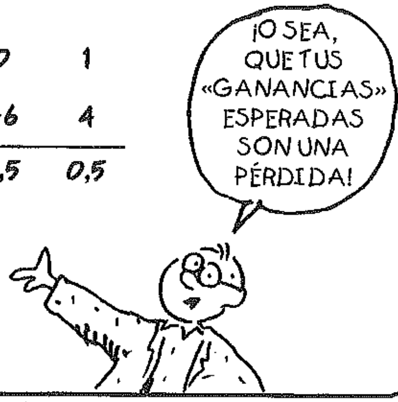
$$E[G] = E[10X - 6]$$
$$= 10E[X] - 6$$

QUE RESULTA EN

$$10(0{,}5) - 6 = -1$$

PUEDES COMPROBARLO CON ESTA TABLA:

x	0	1
g	−6	4
$p(g)$	0,5	0,5

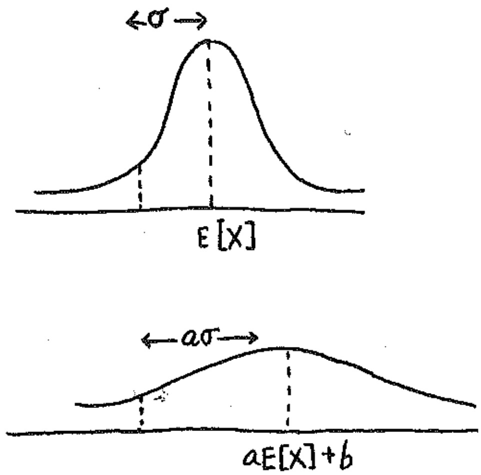


EN GENERAL, NO ES DIFÍCIL DEMOSTRAR QUE

$$E[aX+b] = aE[X]+b$$

CUANDO a Y b SON CUALQUIER NÚMERO Y X ES CUALQUIER VARIABLE ALEATORIA. EN CUANTO A LA VARIANZA, TAMBIÉN EXISTE UN RESULTADO GENERAL:

$$\sigma^2(aX+b) = a^2\sigma^2(X)$$



EN LA APUESTA ANTERIOR, LOS POSIBLES RESULTADOS SON − 6 Y 4, ASÍ QUE ESTÁ CLARO QUE LA VARIANZA DE G TIENE QUE SER MAYOR QUE LA VARIANZA DE X. DE HECHO,

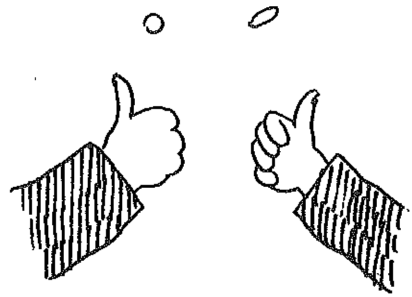
$$\sigma^2(G) = \sigma^2(10X-6)$$
$$= 100\sigma^2(X)$$
$$= 25$$

Y

$$\sigma(G) = 5$$

TAMBIÉN PUEDES **SUMAR** DOS VARIABLES ALEATORIAS. POR EJEMPLO, SUPÓN QUE LANZAMOS UNA MONEDA **DOS VECES**. EL NÚMERO DE CARAS DE LOS DOS LANZAMIENTOS ES $X_1 + X_2$, DONDE X_1 Y X_2 SON LAS VARIABLES ALEATORIAS DE LOS RESULTADOS DEL PRIMER Y SEGUNDO LANZAMIENTO.

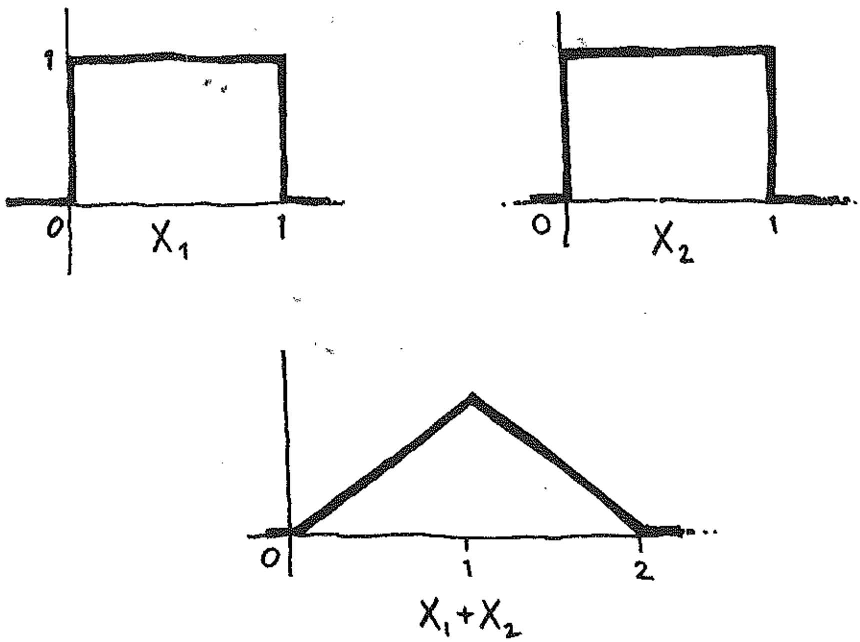
x_1+x_2	0	1	2
$p(x_1+x_2)$	0,25	0,5	0,25



DE NUEVO, ES MUY SENCILLO VER QUE

$$E[X_1+X_2] = E[X_1] + E[X_2]$$

(NO PREGUNTES POR LA DISTRIBUCIÓN DE PROBABILIDAD DE $X_1 + X_2$ PORQUE DEPENDE DE FORMA MUY COMPLICADA DE LAS DOS DISTRIBUCIONES ORIGINALES. POR EJEMPLO, SI TANTO X_1 COMO X_2 SON LA DISTRIBUCIÓN DEL PERRO GIRATORIO, LOS HISTOGRAMAS SE COMPORTARÍAN ASÍ:)



LA VARIANZA DE LA SUMA DE VARIABLES ALEATORIAS TIENE UNA FORMA MUY SIMPLE EN EL CASO ESPECIAL DE QUE X E Y SEAN **INDEPENDIENTES**. LA DEFINICIÓN TÉCNICA DE INDEPENDENCIA SE BASA EN LA PROPIEDAD DE LA PROBABILIDAD P(A Y B) = P(A)P(B). PERO, PARA NOSOTROS, LA INDEPENDENCIA SÓLO SIGNIFICA QUE X E Y ESTÁN GENERADAS POR **MECANISMOS INDEPENDIENTES** COMO EL LANZAMIENTO DE UNA MONEDA, UNA TIRADA DE DADOS, ETC.

CUANDO X E Y SON INDEPENDIENTES, **SUS VARIANZAS SE SUMAN:**

$$\sigma^2(X+Y) = \sigma^2(X)+\sigma^2(Y)$$

EN EL CASO DEL LANZAMIENTO DE DOS MONEDAS,

$$\begin{aligned}\sigma^2(X_1+X_2) &= \sigma^2(X_1)+\sigma^2(X_2)\\ &= 0{,}25 + 0{,}25\\ &= 0{,}5\end{aligned}$$

TODO ESTO SE PUEDE GENERALIZAR A LA SUMA DE MUCHAS VARIABLES ALEATORIAS:

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Y CUANDO TODAS LAS X_i SON INDEPENDIENTES,

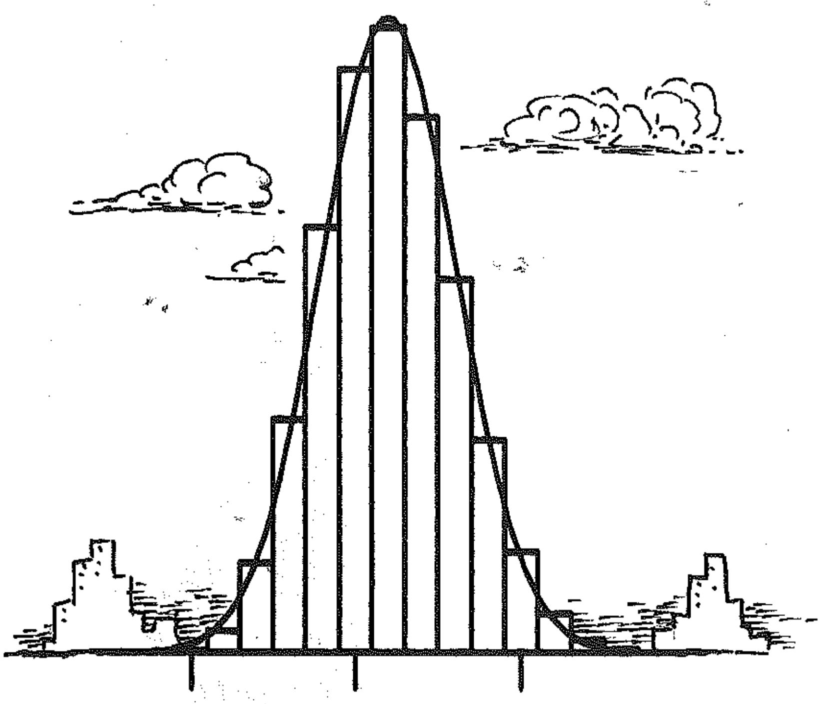
$$\sigma^2\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \sigma^2(X_i)$$

EN EL CAPÍTULO SIGUIENTE VEREMOS DOS IMPORTANTES EJEMPLOS DE VARIABLES ALEATORIAS: UNA, LA **BINOMIAL**, ES LA SUMA DE VARIAS VARIABLES ALEATORIAS INDEPENDIENTES. LA OTRA, LA **NORMAL**, ES UNA VARIABLE ALEATORIA CONTINUA QUE TIENE UNA SORPRENDENTE RELACIÓN CON LA BINOMIAL, Y TAMBIÉN CON CUALQUIER OTRA SUMA DE VARIABLES ALEATORIAS INDEPENDIENTES.

♦ Capítulo 5 ♦

HISTORIA DE DOS DISTRIBUCIONES

AHORA VEREMOS DOS IMPORTANTES EJEMPLOS DE VARIABLES ALEATORIAS, UNA DISCRETA Y OTRA CONTINUA.



EMPEZAREMOS POR LA DISCRETA, LA VARIABLE ALEATORIA **BINOMIAL**. IMAGINEMOS QUE TENEMOS UN PROCESO ALEATORIO CON TAN SÓLO **DOS POSIBLES RESULTADOS**: CARA O CRUZ, VICTORIA O DERROTA EN UN PARTIDO DE FÚTBOL, PASAR O NO PASAR LA INSPECCIÓN DE LA ITV. DE FORMA ARBITRARIA, A UNO DE ESTOS RESULTADOS LO LLAMAMOS **ÉXITO** Y AL OTRO, **FRACASO**.

LO QUE HACEMOS ES REPETIR EL EXPERIMENTO... EN FIN, REPETIDAS VECES. UN EXPERIMENTO DE ESTE TIPO SE LLAMA

Variable aleatoria de Bernoulli,

SIEMPRE QUE PRESENTE ESTAS PROPIEDADES CRÍTICAS:

1) EL RESULTADO DE CADA PRUEBA PUEDE SER ÉXITO O FRACASO.

2) LA PROBABILIDAD p DE ÉXITO ES LA MISMA EN TODAS LAS PRUEBAS.

3) LAS PRUEBAS SON **INDEPENDIENTES**: EL RESULTADO DE UNA NO AFECTA A LOS RESULTADOS POSTERIORES.

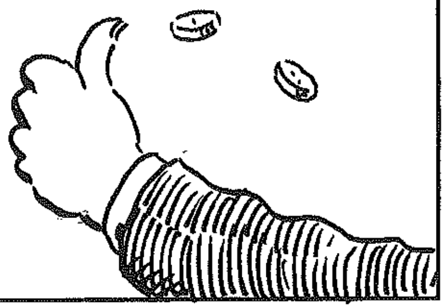
COMENZAREMOS POR UNA VARIABLE ALEATORIA DE BERNOULLI CON UNA PROBABILIDAD p DE ÉXITO. VAMOS A CONSTRUIR UNA NUEVA VARIABLE ALEATORIA REPITIENDO LA PRUEBA.

La variable aleatoria binomial

x ES EL **NÚMERO DE ÉXITOS** DE LAS PRUEBAS DE BERNOULLI REPETIDAS n VECES, CON UNA PROBABILIDAD p DE ÉXITO.

UN EJEMPLO DE VARIABLE ALEATORIA BINOMIAL ES EL NÚMERO DE CARAS (ÉXITOS) DE DOS LANZAMIENTOS DE UNA SOLA MONEDA. EN ESTE CASO $n = 2$ Y $p = 0{,}5$.

k = NÚMERO DE ÉXITOS	0	1	2
PR($x = k$)	0,25	0,5	0,25



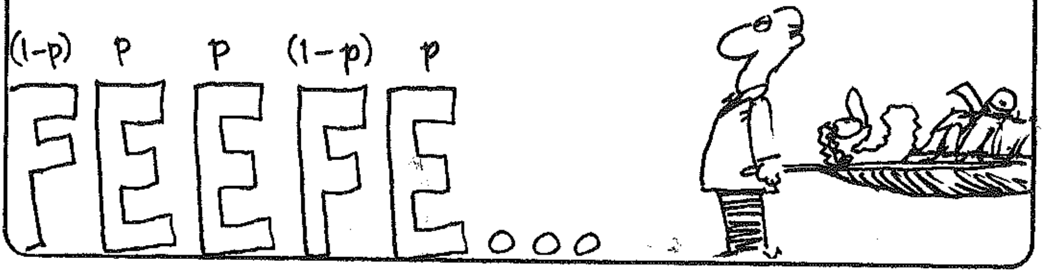
OTRO EJEMPLO ES LA PRIMERA PARTIDA DE DE MERE: TIRAR UN SOLO DADO CUATRO VECES SEGUIDAS. EL ÉXITO ES CONSEGUIR UN 6. LA DISTRIBUCIÓN ES:

¿QUÉ PROBABILIDAD HAY DE CONSEGUIR k VECES 6 EN CUATRO TIRADAS?

En general, ¿cuál es la distribución de probabilidad de una binomial de **cualquier** probabilidad p y número n de pruebas? Un simple cálculo de la probabilidad nos da la respuesta: la probabilidad de obtener k éxitos en n pruebas, $Pr(X = k)$, es

$$Pr(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

En este caso $\binom{n}{k}$, que se lee «combinaciones de n elementos tomados de k en k», es el **coeficiente binomial**. Éste cuenta todas las maneras posibles de obtener k éxitos en n pruebas. Cada secuencia individual de k éxitos y $n-k$ fracasos tiene una probabilidad $p^k(1-p)^{n-k}$, según la regla de multiplicación. El número de secuencias es $\binom{n}{k}$.



La fórmula de $\binom{n}{k}$ es

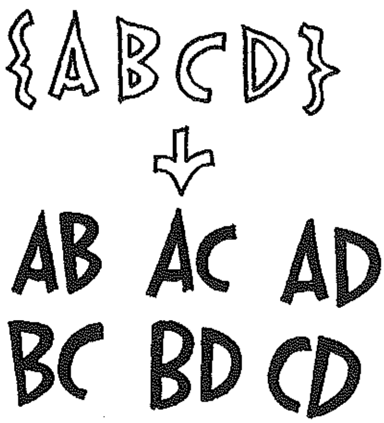
$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

En la que

$$n! = n \times (n-1) \times (n-2) \times \dots \times 1$$

Y $0!$ se considera 1. Por ejemplo, $\binom{4}{2}$, el número de combinaciones posibles de elegir dos letras de un conjunto de cuatro, es

$$\binom{4}{2} = \frac{4!}{2!2!} = \frac{24}{4} = 6$$



OTRO PUNTO DE VISTA DE LOS COEFICIENTES BINOMIALES ES EL **TRIÁNGULO DE PASCAL**. CADA ENTRADA ES LA SUMA DE LOS DOS NÚMEROS QUE TIENE ENCIMA.

$\binom{3}{1}$ 1

1 1

$\binom{7}{2}$ 1 2 1 $\binom{5}{3}$

1 3 3 1

1 4 6 4 1

1 5 10 10 5 1 $\binom{10}{7}$

$\binom{8}{0}$ 1 6 15 20 15 6 1

1 7 21 35 35 21 7 1

1 8 28 56 70 56 28 8 1

1 9 36 84 126 126 84 36 9 1 $\binom{10}{10}$

1 10 45 120 210 252 210 120 45 10 1

1 11 55 165 330 462 462 330 165 55 11 1

1 12 66 220 495 792 924 792 495 220 66 12 1

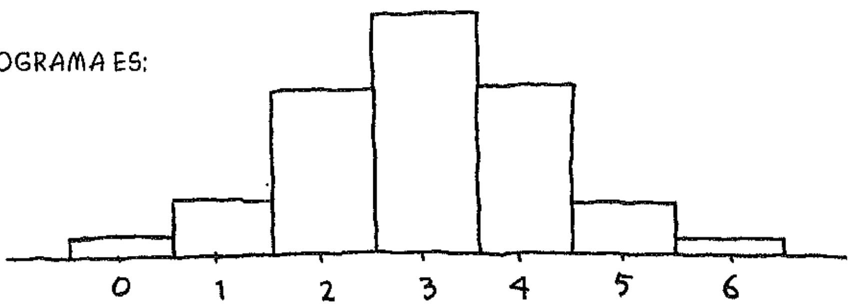
ETC.

PARA ENCONTRAR $\binom{n}{k}$ SÓLO HACE FALTA CONTAR HASTA LA FILA n Y HASTA LA ENTRADA k (SIN OLVIDAR QUE HAY QUE EMPEZAR CONTANDO EL CERO).

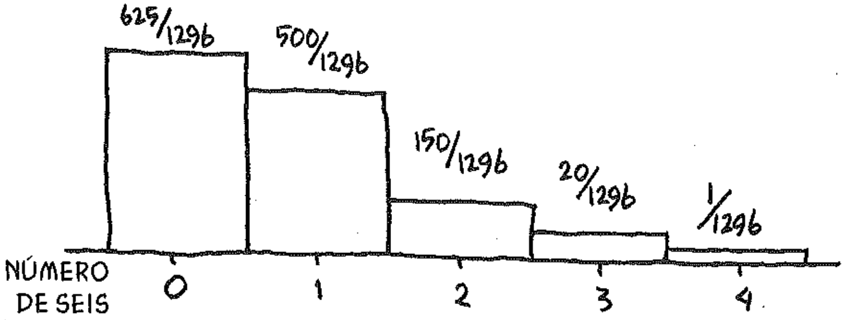
CUANDO $p = 0{,}5$, LA DISTRIBUCIÓN DE PROBABILIDAD DE LA BINOMIAL ES PERFECTAMENTE SIMÉTRICA. EN 6 LANZAMIENTOS DE UNA MONEDA, POR EJEMPLO, ES

k = # CARAS	0	1	2	3	4	5	6
$Pr(X=k)$	$\left(\frac{1}{2}\right)^6$	$\left(\frac{1}{2}\right)^6 \cdot 6$	$\left(\frac{1}{2}\right)^6 \cdot 15$	$\left(\frac{1}{2}\right)^6 \cdot 20$	$\left(\frac{1}{2}\right)^6 \cdot 15$	$\left(\frac{1}{2}\right)^6 \cdot 6$	$\left(\frac{1}{2}\right)^6$

Y EL HISTOGRAMA ES:



EN LA TIRADA DE CUATRO DADOS DE DE MERE, LA DISTRIBUCIÓN ES MÁS DESPROPORCIONADA:



LA **MEDIA** Y LA **VARIANZA** DE LA DISTRIBUCIÓN BINOMIAL SON

$$\mu = np$$

$$\sigma^2 = np(1-p)$$

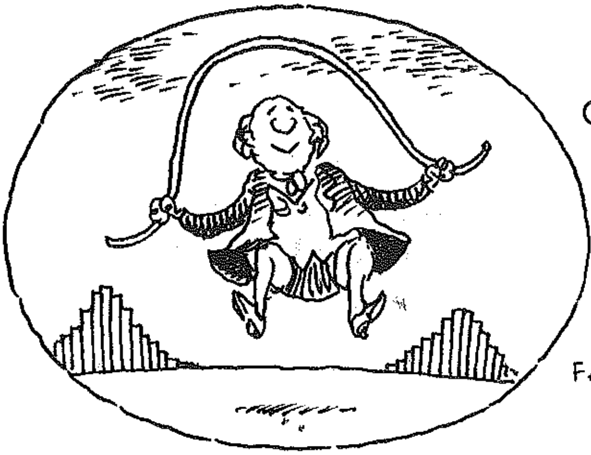
OBSERVA QUE LA MEDIA, CON UN POCO DE INTUICIÓN, TIENE MUCHO SENTIDO: EN n PRUEBAS DE BERNOULLI, EL NÚMERO DE ÉXITOS QUE SE ESPERA DEBERÍA SER np. LA VARIANZA SE DERIVA DEL HECHO DE QUE LA BINOMIAL ES LA SUMA DE n PRUEBAS DE BERNOULLI INDEPENDIENTES CON UNA VARIANZA $p(1-p)$.

LOS PARÁMETROS DE LA DISTRIBUCIÓN BINOMIAL SON n Y p. TANTO LA DISTRIBUCIÓN COMO LA MEDIA Y LA VARIANZA DEPENDEN **SÓLO** DE ESOS DOS NÚMEROS. EN LA MAYORÍA DE LIBROS Y PROGRAMAS INFORMÁTICOS APARECEN TABLAS DE DISTRIBUCIÓN BINOMIAL. ÉSTA ES LA TABLA DE $n = 10$.

VALORES DE $\Pr(X = k)$

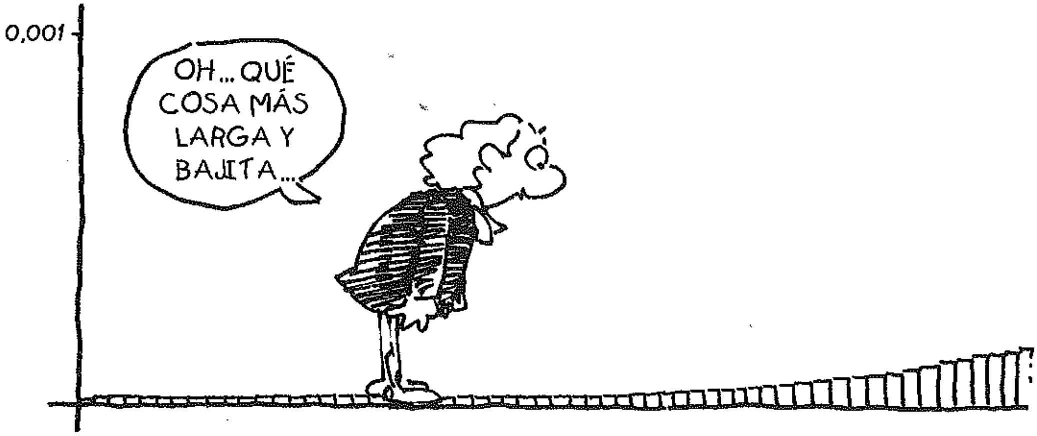
		k										
		0	1	2	3	4	5	6	7	8	9	10
	0,1	0,349	0,387	0,194	0,057	0,011	0,001	0,000	0,000	0,000	0,000	0,000
	0,25	0,056	0,188	0,282	0,250	0,146	0,058	0,016	0,003	0,000	0,000	0,000
p	0,50	0,001	0,010	0,044	0,117	0,205	0,246	0.205	0,117	0,044	0,010	0,001
	0,75	0,000	0,000	0,000	0,003	0,016	0,058	0,146	0,250	0,282	0,188	0,056
	0,9	0,000	0,000	0,000	0,000	0,000	0,001	0,011	0,057	0,194	0,387	0,349

SIN EMBARGO, HACER ESTOS CÁLCULOS CON VALORES GRANDES DE n PUEDE CONVERTIRSE EN UNA TORTURA... O, AL MENOS LO ERA EN EL SIGLO XVIII, CUANDO **JAMES BERNOULLI** Y **ABRAHAM DE MOIVRE** INTENTABAN HACERLO SIN LA AYUDA DE UN ORDENADOR.



CON UN ARMA DE RECIENTE INVENCIÓN, EL **CÁLCULO**, DE MOIVRE DEMOSTRÓ QUE CUANDO $p = 0{,}5$, LA DISTRIBUCIÓN BINOMIAL SE PODÍA OBTENER APROXIMADAMENTE MEDIANTE UNA **FUNCIÓN DE DENSIDAD CONTINUA**, MUY FÁCIL DE DESCRIBIR.

PARA VER SU FUNCIONAMIENTO, IMAGINEMOS UNA DISTRIBUCIÓN BINOMIAL CON $p = 0{,}5$ Y UN NÚMERO n MUY ELEVADO, POR EJEMPLO, UN MILLÓN...



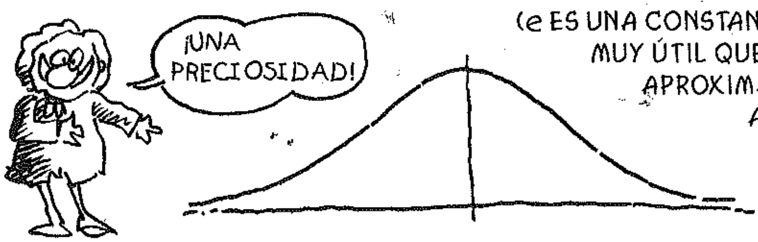
EL RESULTADO SE PARECE MUCHO A UNA **CURVA SUAVIZADA, EN FORMA DE CAMPANA, SIMÉTRICA**, Y DE MOIVRE DEMOSTRÓ QUE VIENE DADA POR UNA FÓRMULA MUY SIMPLE:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

ESTA FUNCIÓN RECIBE EL NOMBRE DE

distribución normal tipificada.

(*e* ES UNA CONSTANTE MATEMÁTICA MUY ÚTIL QUE EQUIVALE APROXIMADAMENTE A *2,718*.)



(CONVÉNCETE DE QUE ESTA FUNCIÓN TIENE UN GRÁFICO EN FORMA DE CAMPANA. PARA VALORES DE z ALEJADOS DE CERO, $f(z)$ ES PRÁCTICAMENTE CERO, TIENE UN DENOMINADOR MUY ELEVADO; Y ES SIMÉTRICO, YA QUE $f(z) = f(-z)$, Y TIENE UN MÁXIMO DE $z = 0$.)

ESTA DISTRIBUCIÓN SE LLAMA NORMAL **TIPIFICADO*** PORQUE TODA ESA COMPRESIÓN Y EXTENSIÓN A LO LARGO DE LOS EJES ESTÁ PENSADA PARA DARLES ESTAS SIMPLES PROPIEDADES, QUE AHORA NOSOTROS PRESENTAMOS SIN PRUEBA ALGUNA:

$$\mu = 0$$

$$\sigma = 1$$

* TAMBIÉN SE LLAMA DISTRIBUCIÓN NORMAL CENTRADA Y REDUCIDA. [N. T.]

PARA RESUMIR LA TEORÍA DE DE MOIVRE, SI **«NORMALIZAMOS» LA DISTRIBUCIÓN BINOMIAL** CON $p = 1/2$ (O SEA, HACIENDO QUE SU CENTRO SEA CERO Y SU DESVIACIÓN TÍPICA = 1) ENTONCES SE APROXIMA MUCHO A LA **DISTRIBUCIÓN NORMAL TIPIFICADA.**

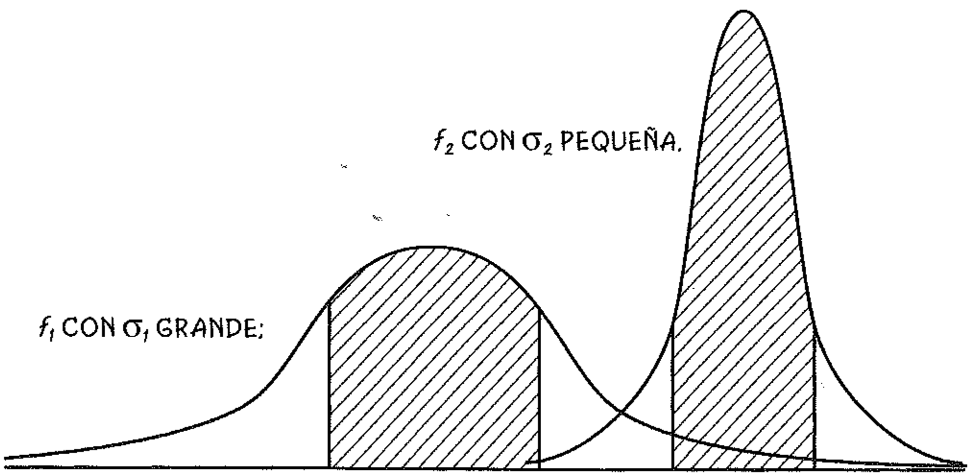
$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

OTRAS NORMALES, CON DISTINTAS MEDIAS Y VARIANZAS, SE OBTIENEN EXTENDIENDO Y DESPLAZANDO LA NORMAL TIPIFICADA. EN GENERAL, PODEMOS ESCRIBIR LA FÓRMULA

$$f(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

ESTO NOS DA UNA DISTRIBUCIÓN SIMÉTRICA Y CAMPANIFORME CON EL CENTRO EN LA MEDIA μ Y LA DESVIACIÓN TÍPICA σ.

AQUÍ TIENES DOS NORMALES DIFERENTES CON LA ZONA DE LA DESVIACIÓN TÍPICA SOMBREADA...

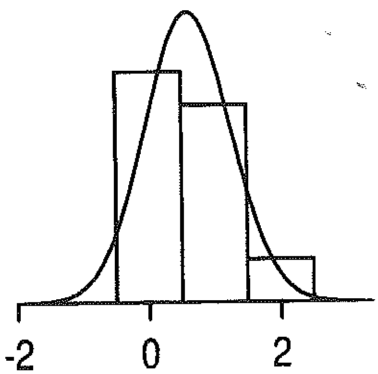


DE MOIVRE DEMOSTRÓ QUE LA NORMAL TIPIFICADA SE CORRESPONDE CON LA BINOMIAL (NORMALIZADA) DE $p = 0{,}5$, PERO LO CIERTO ES QUE FUNCIONA CON **CUALQUIER VALOR DE p.**

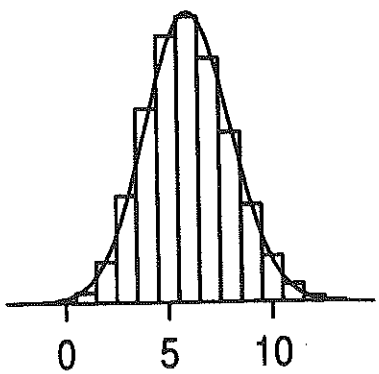
EN GENERAL: PARA **CUALQUIER** VALOR DE p, LA DISTRIBUCIÓN BINOMIAL DE n PRUEBAS CON PROBABILIDAD p SE APROXIMA A LA CURVA NORMAL CON $\mu = np$ Y $\sigma = np(1-p)$.

LA VERDAD ES QUE ESTO ES UN POCO EXTRAÑO. TODAS LAS NORMALES SON SIMÉTRICAS Y CAMPANIFORMES, PERO —COMO YA HEMOS VISTO— LAS DISTRIBUCIONES BINOMIALES **NO SON SIMÉTRICAS** CUANDO $p \neq 0{,}5$.

SIN EMBARGO, RESULTA QUE A MEDIDA QUE n CRECE, LA ASIMETRÍA DE LA BINOMIAL SE COMPENSA, COMO PUEDES VER EN ESTE EJEMPLO:



Binomial: $n = 2$ y $p = 0{,}3$



Binomial: $n = 20$ y $p = 0{,}3$

DE HECHO, EL DESCUBRIMIENTO DE DE MOIVRE SOBRE LA BINOMIAL ES UN CASO ESPECIAL DE UN RESULTADO AÚN **MÁS** GENERAL, QUE NOS AYUDA A EXPLICAR POR QUÉ LA NORMAL ES TAN IMPORTANTE Y DE NATURALEZA TAN EXTENDIDA. SE TRATA DEL SIGUIENTE:

«Teorema central del límite»:

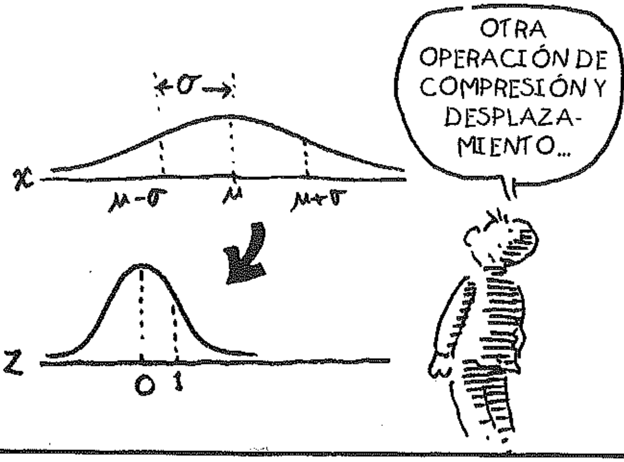
LOS DATOS INFLUIDOS POR **MUCHOS PEQUEÑOS EFECTOS ALEATORIOS INDEPENDIENTES** TIENEN, MÁS O MENOS, UNA **DISTRIBUCIÓN NORMAL**.

ASÍ SE EXPLICA QUE LA NORMAL **ESTÉ EN TODAS PARTES**: LAS FLUCTUACIONES DE LA BOLSA, LOS PESOS DE LOS ESTUDIANTES, LA MEDIA ANUAL DE TEMPERATURAS, LAS NOTAS DE SELECTIVIDAD: TODOS SON RESULTADOS DE MÚLTIPLES EFECTOS DIFERENTES. POR EJEMPLO, EL PESO DE UN ESTUDIANTE ES EL RESULTADO DE LA GENÉTICA, LA NUTRICIÓN, LAS ENFERMEDADES Y LA CERVEZA DE LA FIESTA DE LA NOCHE ANTERIOR. CUANDO LOS JUNTAMOS TODOS, ¡OBTENEMOS LA **NORMAL**! (RECUERDA QUE LA BINOMIAL ES EL RESULTADO DE n PRUEBAS DE BERNOULLI INDEPENDIENTES.)

LA TRANSFORMACIÓN *z*

$$z = \frac{x-\mu}{\sigma}$$

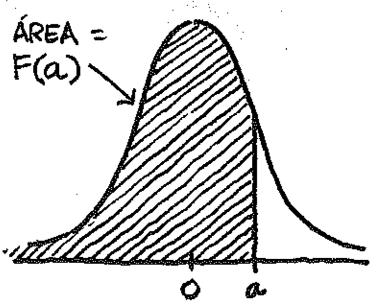
CONVIERTE UNA VARIABLE ALEATORIA NORMAL DE MEDIA μ Y DESVIACIÓN TÍPICA σ EN UNA VARIABLE ALEATORIA NORMAL **TIPIFICADA** CON MEDIA *0* Y DESVIACIÓN TÍPICA *1*.



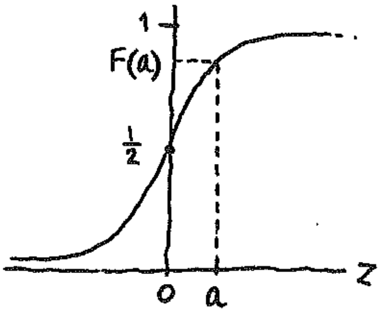
ENTONCES, TODO LO QUE NECESITAMOS PARA ENCONTRAR **CUALQUIER** DISTRIBUCIÓ *N* NORMAL ES UNA SOLA TABLA DE LA **NORMAL TIFIFICADA** ***f(z)***.

z	-2,5	-2,4	-2,3	-2,2	-2,1	-2,0	-1,9	-1,8	-1,7	-1,6
F(z)	**0,006**	**0,008**	**0,011**	**0,014**	**0,018**	**0,023**	**0,029**	**0,036**	**0,045**	**0,055**
z	-1,5	-1,4	-1,3	-1,2	-1,1	-1,0	-0,9	-0,8	-0,7	-0,6
F(z)	**0,067**	**0,081**	**0,097**	**0,115**	**0,136**	**0,159**	**0,184**	**0,212**	**0,242**	**0,274**
z	-0,5	-0,4	-0,3	-0,2	- 0,1	0,0	0,1	0,2	0,3	0,4
F(z)	**0,309**	**0,345**	**0,382**	**0,421**	**0,460**	**0,500**	**0,540**	**0,579**	**0,618**	**0,655**
z	0,5	0,6	0,7	0,8	0,9	1,0	1,1	1,2	1,3	1,4
F(z)	**0,691**	**0,726**	**0,758**	**0,788**	**0,816**	**0,841**	**0,864**	**0,885**	**0,903**	**0,919**
z	1,5	1,6	1,7	1,8	1,9	2,0	2,1	2,2	2,3	2,4
F(z)	**0,933**	**0,945**	**0,955**	**0,964**	**0,971**	**0,977**	**0,982**	**0,986**	**0,989**	**0,992**
z	2,5									
F(z)	**0,994**									

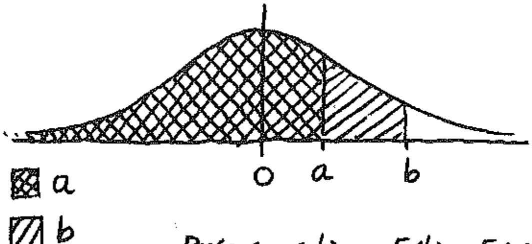
AQUÍ F(*a*) = *Pr*(*z* ≤ *a*). EL ÁREA DE LA CURVA DE DENSIDAD A LA IZQUIERDA DE *z* = *a*.



(TAMBIÉN PODEMOS CONFECCIONAR UN GRÁFICO DE *y* = F(*z*), LA **PROBABILIDAD ACUMULADA**. TIENE ESTE ASPECTO.)



LA TABLA NOS PERMITE ENCONTRAR LA PROBABILIDAD DE QUE Z ESTÉ EN UN INTERVALO $a \leq z \leq b$. TAN SÓLO ES LA DIFERENCIA ENTRE LAS ÁREAS F(b) Y F(a).

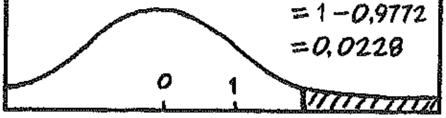


$$Pr(a \leqslant z \leqslant b) = F(b) - F(a)$$

DE ESTE MODO, POR EJEMPLO,

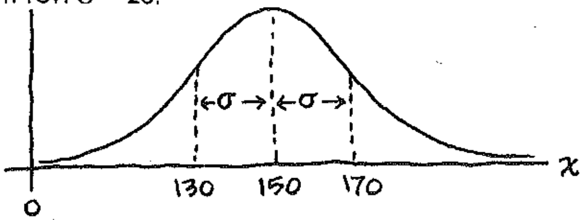
$$Pr(-1 < z < 1) = F(1) - F(-1)$$
$$= 0{,}8413 - 0{,}1587$$
$$= 0{,}6826$$

$$Pr(z \geqslant 2) = 1 - F(2)$$
$$= 1 - 0{,}9772$$
$$= 0{,}0228$$



SI UTILIZAMOS LA SUSTITUCIÓN $z = \frac{x-\mu}{\sigma}$, TAMBIÉN PODEMOS USAR LA MISMA TABLA PARA ENCONTRAR LAS PROBABILIDADES DE OTRAS DISTRIBUCIONES NORMALES.

POR EJEMPLO, SUPONGAMOS QUE LOS PESOS DE LOS ESTUDIANTES TIENEN UNA DISTRIBUCIÓN NORMAL CON MEDIA $\mu = 150$ LIBRAS Y UNA DESVIACIÓN TÍPICA $\sigma = 20$:

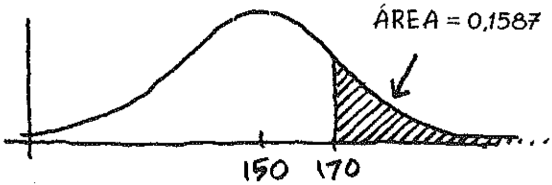


ENTONCES, ¿CUÁL ES LA PROBABILIDAD DE PESAR MÁS DE 170 LIBRAS?

AHORA SE TRATA «SÓLO» DE ÁLGEBRA.

$$Pr(X > 170) =$$
$$Pr\left(\frac{X-\mu}{\sigma} > \frac{170-150}{20}\right) =$$
$$Pr\left(Z > \frac{20}{20}\right) =$$
$$Pr(Z > 1)$$

ESO ES $1 - F(1)$, QUE COMO PODEMOS VER EN LA TABLA ES $1 - 0{,}8413 = \mathbf{0{,}1587}$

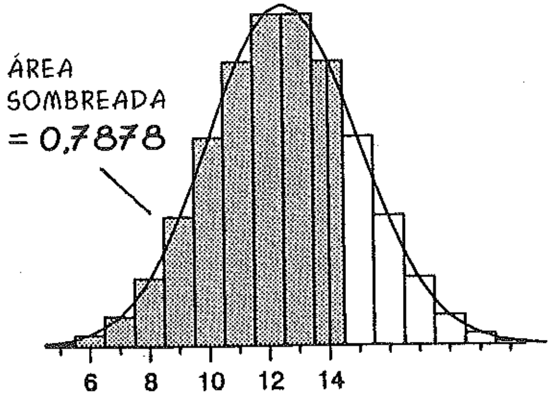


ALGO MENOS DE UN ESTUDIANTE DE CADA SEIS PESA MÁS DE 170 LIBRAS.

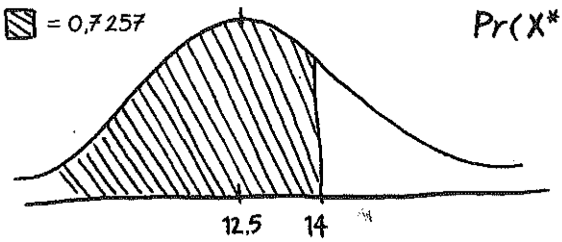
ENTONCES, LA REGLA GENERAL PARA CALCULAR LAS PROBABILIDADES ASOCIADAS A LA DISTRIBUCIÓN NORMAL ES:

$$Pr(a \leqslant X \leqslant b) = F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right)$$

Y AHORA, VOLVIENDO A DE MOIVRE Y SU APROXIMACIÓN BINOMIAL... VAMOS A VER UNA DISTRIBUCIÓN BINOMIAL CON $n = 25$ PRUEBAS Y $p = 0{,}5$ (25 LANZAMIENTOS DE UNA MONEDA, POR EJEMPLO). PODEMOS CALCULAR (O CONSULTAR EN LA TABLA) CUALQUIER PROBABILIDAD, POR EJEMPLO $\Pr(x \leq 14)$. Y ES EXACTAMENTE **0,7878**.



AHORA CALCULAMOS UNA VARIABLE ALEATORIA NORMAL X^* CON LA MISMA MEDIA $\mu = np = (25)(0{,}5) = 12{,}5$ Y DESVIACIÓN TÍPICA $\sigma = np(1 - p) = 2{,}5$.



$$\Pr(X^* \leq 14) = \Pr\left(Z \leq \frac{14 - 12{,}5}{2{,}5}\right)$$
$$= \Pr(Z \leq 0{,}6)$$
$$= \mathbf{0{,}7257}$$

¡AH! ¡PERO AÚN PODEMOS MEJORARLO! SI OBSERVAS EL HISTOGRAMA CON ATENCIÓN, VERÁS QUE LAS BARRAS TIENEN UN NÚMERO EN EL **CENTRO**. ESTO SIGNIFICA QUE $\Pr(X^* \leq 14)$ ES EN REALIDAD EL ÁREA DE LAS BARRAS MENORES A $x = 14{,}5$. DEBEMOS TENER EN CUENTA ESE 0,5, Y DE HECHO,

$$\Pr(X^* \leq 14{,}5) = \Pr(z \leq 0{,}8)$$
$$= \mathbf{0{,}7881}$$

¡UNA APROXIMACIÓN MAGNÍFICA A 0,7878!

ESE OTRO 0,5 QUE HEMOS AÑADIDO SE LLAMA

corrección de continuidad.

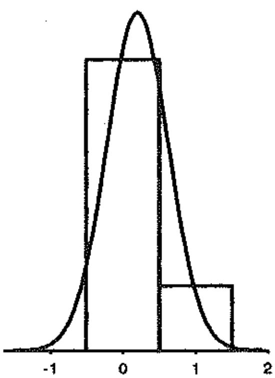
TENEMOS QUE INCLUIRLO PARA OBTENER UNA BUENA APROXIMACIÓN CONTINUA A NUESTRA VARIABLE ALEATORIA BINOMIAL DISCRETA X. TODO SE RESUME EN ESTA HORRIBLE FÓRMULA:

$$\Pr(a \le X \le b) \simeq \Pr\left(\frac{a-\frac{1}{2}-np}{\sqrt{np(1-p)}} \le Z \le \frac{b+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)$$

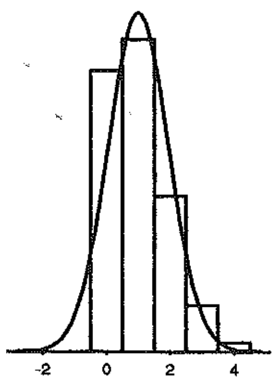
¿CUÁNDO ES LA APROXIMACIÓN «LO SUFICIENTEMENTE BUENA»? PARA LOS ESTADÍSTICOS, LA REGLA EMPÍRICA ES LA SIGUIENTE: SIEMPRE QUE n SEA LO BASTANTE GRANDE PARA QUE TANTO EL NÚMERO DE ÉXITOS COMO EL DE FRACASOS SEA MAYOR QUE **CINCO**:

$$np \geq 5 \quad \text{Y} \quad n(1-p) \geq 5$$

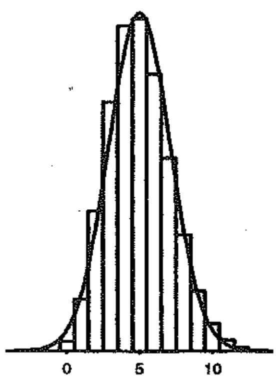
EN ESTOS HISTOGRAMAS PUEDES VER QUE CUANDO $p = 0{,}1$ LA EQUIVALENCIA ES BASTANTE MEDIOCRE, O INCLUSO MUY MALA, HASTA QUE n LLEGA A 50 Y HACE QUE $np = 5$.



$n=2,\ p=0{,}1$

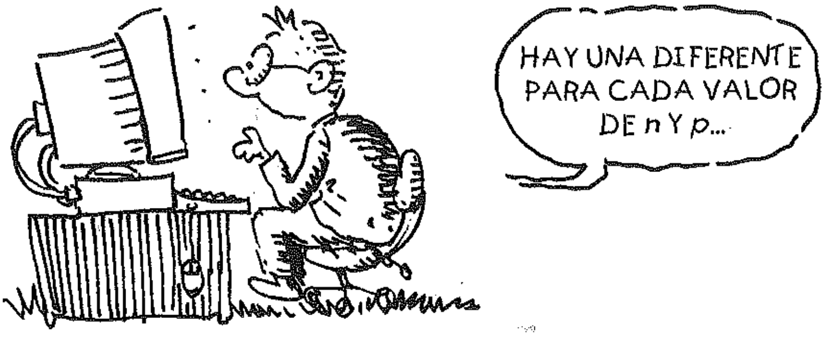


$n=10,\ p=0{,}1$

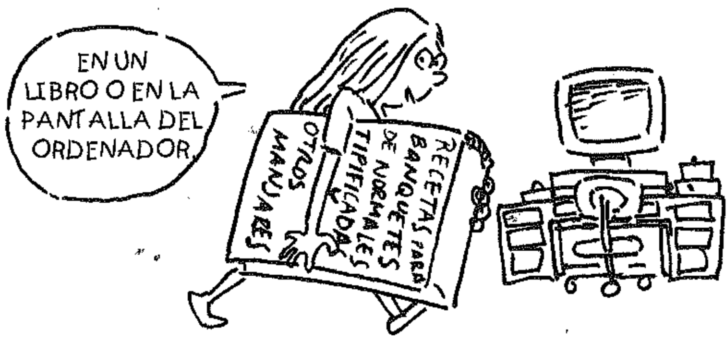


$n=50,\ p=0{,}1$

¿QUÉ TIENE DE MARAVILLOSO ESTA APROXIMACIÓN NORMAL? LA DISTRIBUCIÓN BINOMIAL SE DA MUY A MENUDO EN LA NATURALEZA, Y NO ES DIFÍCIL DE COMPRENDER, PERO CALCULARLA PUEDE SER AGOTADOR.



LA NORMAL QUE SE LE APROXIMA ES QUIZÁ MENOS INTUITIVA, PERO MUY FÁCIL DE USAR. LA TRANSFORMACIÓN z CONVIERTE **CUALQUIER** NORMAL A LA **NORMAL TIPIFICADA**. Y ESO NOS PERMITE LEER LAS PROBABILIDADES DIRECTAMENTE DE UNA SIMPLE TABLA NUMÉRICA.



Y ADEMÁS, ¡LA NORMAL ES LA **MADRE DE TODAS LAS DISTRIBUCIONES**!