MODA mediana polígono de frecuencias

diagrama de sectores desviación típica

estadística distribución de frecuencias

Experiencias estadísticas

Autores:

Cristina de Vega Bigotes

Daniel Pasarón Suárez

Eva Rodríguez Martínez

Saúl Fernández Crespo

Tutor:

Arturo Llamedo Fernández

Paradojas juegos azar probabilidad intuición

Experimentos deterministas sucesos

media aritmética aleatorio cualitativo

diagrama de barras Teorema de Bayes porcentaje

Experiencias estadísticas

Saúl Fernández Crespo, Daniel Pasarón Suárez, Eva Rodríguez Martínez, Cristina de Vega Bigotes

I.E.S. Carmen y Severo Ochoa de Luarca C/ Catedrático Ramón Losada s/n, 33700 Luarca

Resumen: En este trabajo hemos realizado un estudio empírico de determinadas paradojas y juegos con los que hemos pretendido comprobar la influencia del azar, así como demostrar que la intuición muchas veces nos juega una mala pasada, si antes no reflexionamos detenidamente sobre los fenómenos y experimentos que realizamos.

Abstract: In this job we have made an empirical study of certain paradox and games through which we expect to check the influence of fate, as well as prove that intuition many times plays a dirty trick on us if we don't reflect carefully about the experiments and phenomenon we work on.

1. Introducción

¿Las coincidencias existen o todo tiene una explicación razonable? ¿Somos capaces de saber lo que nos va a suceder o es impredecible? ¿Debemos guiarnos por la intuición en nuestras investigaciones o ésta nos puede llevar a error? ¿El azar rige nuestro entorno o tal vez el destino está escrito para cada uno de nosotros? ¿Podemos seguir una estrategia fiable para ganar a las cartas, al dominó o a los dados, o el azar es más poderoso que nuestra mente? ¿Es mejor llevar una bomba cuando realizamos un viaje en avión para reducir la probabilidad de que haya un atentado terrorista? Estas son algunas de las preguntas que nos hemos planteado a lo largo de estos meses mientras realizábamos este trabajo de estadística.

Después de lo analizado con los juegos y paradojas que nos ha planteado nuestro tutor, creemos que habitualmente la mayoría de las cosas no suceden de forma imprevista. El problema estriba en que en la mayoría de las situaciones no conocemos las reglas, pero eso no quiere decir que éstas no existan. Por ejemplo, uno se puede plantear la siguiente cuestión: ¿Por qué el cuadro de *la Mona Lisa,* de Leonardo Da Vinci, es una obra maestra de la pintura? Pues parece ser, entre otros motivos, que la razón áurea aparece repetidas veces en el cuadro. A este tipo de cuestiones nos referimos cuando hacemos alusión al desconocimiento de las reglas.

A menudo, coincidencias que pueden parecernos raras, improbables o por lo menos muy sorprendentes, no lo son tales y tienen una explicación razonable y lógica gracias a las matemáticas. Existen fenómenos que desconciertan a nuestra intuición, ya que la mayoría pensaríamos en una respuesta que dista mucho de la realidad.

Pues eso es lo que hemos hecho en este trabajo: guiarnos de nuestra intuición en algunos casos, y demostrar empíricamente que esta nos puede llevar a error. Para ello, nos hemos servido de ciertas paradojas y juegos que nos ha ido presentando nuestro tutor, desde finales de octubre.

2. *Paradojas y juegos de azar*

Comenzaremos buscando en Internet lo que es una *paradoja* y los tipos de *paradojas* que hay:

Según la Real Academia de la Lengua, una *paradoja* “es una idea extraña u opuesta a la común opinión y al sentir de las personas, es decir, una aserción inverosímil o absurda, que se presenta con apariencias de verdadera”. El término *paradoja* viene del griego (para y doxos) y significa "más allá de lo creíble". En la actualidad la palabra "paradoja" tiene numerosos significados:

- Afirmación que parece falsa, aunque en realidad es verdadera.
- Afirmación que parece verdadera, pero en realidad es falsa.
- Cadena de razonamientos aparentemente impecables, que conducen sin embargo a contradicciones lógicas. (Las paradojas de esta clase suelen llamarse falacias.)
- Declaración cuya veracidad o falsedad es indecible.
- Verdad que se vuelve patas arriba para llamar la atención.

Roberto Muñoz Sánchez define la *paradoja* como un enunciado que aparentemente es verdadera pero que conlleva a una autocontradicción o a una situación que contradice el sentido común. Dicho de otro modo, una *paradoja* es “lo opuesto a lo que uno considera cierto”. Se encuentra compuesta por el prefijo *para*, que significa “contrario a” o “alterado”, en conjunción con el sufijo *doxa*, que significa “opinión”.

Hay paradojas sobre el infinito, en matemáticas, en estadísticas, en físicas o en economía y hasta incluso las hay bíblicas y religiosas

Según su veracidad hablamos de:

- *Paradojas verídicas*, que enuncian resultados que aparentan ser absurdos aunque se pueda demostrar su veracidad.
- *Paradojas antinomias,* que llegan a un resultado que se autocontradice, a pesar de aplicar correctamente el razonamiento.
- *Paradojas antinomias,* de definición que se basan en definiciones ambiguas, sin las cuales no se alcanza una contradicción.
- *Paradojas condicionales,* que solamente lo son si se hacen ciertas suposiciones. Algunas muestran que esas suposiciones son falsas o incompletas.

Respecto a la *Teoría de Juegos,* queremos destacar que estudia de manera formal y abstracta las decisiones óptimas que deben tomar diversos adversarios en conflicto, pudiendo definirse como el estudio de modelos matemáticos que describen el conflicto y la cooperación entre entes inteligentes que toman decisiones. Estas decisiones se consideran estratégicas, es decir, que los entes que participan en el juego actúan teniendo en cuanta las acciones que tomarían los demás.

Aunque en principio la palabra *juego* tiene connotaciones lúdicas y relativas al azar, la *teoría de juegos* no tiene como principal objetivo el estudio de los juegos de salón, aunque sí entran dentro de su dominio. Una terminología alternativa que ilustra más claramente el objeto de la Teoría de Juegos es el “análisis matemático de conflictos” y la “toma interactiva de decisiones”.

3. La paradoja del cumpleaños

La primera paradoja que nos planteó nuestro tutor fue **la paradoja del cumpleaños**. Un día en clase, en el mes de octubre, Arturo nos preguntó:

¿Cuál creéis que es la probabilidad de que en un grupo de 23 personas dos de ellas celebren su cumpleaños el mismo día? Y si en vez de 23 fueran 57 ¿Cuál sería la probabilidad de que 2 de ellas celebren su cumpleaños el mismo día?

En un principio nos parecía bastante complicado que dos de ellas coincidieran. Le propusimos un resultado siguiendo nuestra intuición y nuestro tutor nos comentó que distaba mucho del verdadero resultado. Así que nos propuso las siguientes actividades prácticas para dar respuesta a esta primera paradoja.

En primer lugar, nos pidió que cogiéramos a los 23 jugadores de cada una de las 32 selecciones que participaron en el último mundial de fútbol disputado en Brasil y viéramos en cuántas de ellas había al menos dos jugadores que nacieran el mismo día. Dichos datos los extrajimos de:

http://www.losmundialesdefutbol.com/mundiales/2014_mundial.php

El resultado fue curioso. De las 32 selecciones que participaron en el mundial de fútbol, en 13 hay 2 o más jugadores cuyo día y mes de nacimiento coincide. Este dato representa algo más del 41%.

Después nos pidió que repitiéramos este mismo estudio con los 23 últimos premios Nobel de medicina, de la paz, de física y de economía. En esta ocasión el resultado fue el siguiente:

De los últimos 23 premios Nobel de medicina hubo coincidencia en los de los años 2006 y 2014.

De los últimos 23 premios Nobel de la paz no hubo coincidencias.

De los últimos 23 premios Nobel de física hubo coincidencia en los años 2012 y 2014; y en 2011 y 2012. En el año 2012 el galardón fue compartido.

Y de los últimos 23 premios Nobel de economía hay coincidencia en los años 2007 y 2013.

A continuación, solicitamos en Jefatura de Estudios las fechas de nacimiento de los alumnos del instituto matriculados en la E.S.O. y tomamos 3 muestras aleatorias de 23 alumnos de cada nivel educativo.

En las 3 muestras de los 23 alumnos de 1º de la E.S.O. hubo coincidencia en dos; en las 3 muestras de los 23 alumnos de 2º de la E.S.O. hubo coincidencia en 2; en las 3 muestras de los 23 alumnos de 3º de la E.S.O. hubo coincidencia en 2; y por último, en las 3 muestras de los 23 alumnos de 4º de la E.S.O. hubo coincidencia en 1. Es decir, en las 12 muestras de alumnos que habíamos elaborado en 7 de ellas hubo coincidencias.

En resumen, de las 48 muestras escogidas, hemos encontrado que en 23 hubo coincidencia lo que representa el 47.91%.

Al comparar la prueba empírica que habíamos hecho con la suposición que nosotros habíamos realizado los resultados eran muy distintos: hecho que nos hizo dudar del resultado que habíamos calculado.

Arturo nos pidió que realizáramos un nuevo experimento. En esta ocasión nos pidió tomar muestras de 57 individuos. Nosotros elaboramos 10 muestras aleatorias de 57 alumnos de nuestro Instituto. Y, curiosamente, en las 10 muestras aleatorias hubo coincidencias. Esta circunstancia aún nos sorprendió más.

Finalmente, hicimos una **simulación** con el ordenador y elaboramos con Excel una hoja de cálculo que construía muestras aleatorias de fechas de nacimiento, sin tener en cuenta el año de nacimiento para ver las posibles coincidencias. Estas muestras eran de tamaño de 23 y 57. Construimos primero 100 muestras aleatorias y después 200, utilizando las funciones *ENTERO* y *ALEATORIO*. Los resultados fueron los siguientes:

Muestras de tamaño 23

En las cien muestras de tamaño 23 hubo 45 coincidencias. Lo que representa el 45%

En las doscientas muestras de tamaño 23 hubo 111 coincidencias. Lo que representa el 55,5%

Muestras de tamaño 57

En las 100 muestras de tamaño 57 hubo 96 coincidencias.

En las 200 muestras de tamaño 57 hubo 200 coincidencias.

En resumen: De las 348 muestras de 23 individuos que elaboramos en 159 no hubo coincidencia, lo que supone un 45,68%. Y respecto a las 310 muestras de tamaño 57 que elaboramos en tan solo 4 no hubo coincidencia lo que representa un 1,29%. Ninguno de los 2 valores se acerca ni por asomo, a lo que nosotros suponíamos que podía pasar.

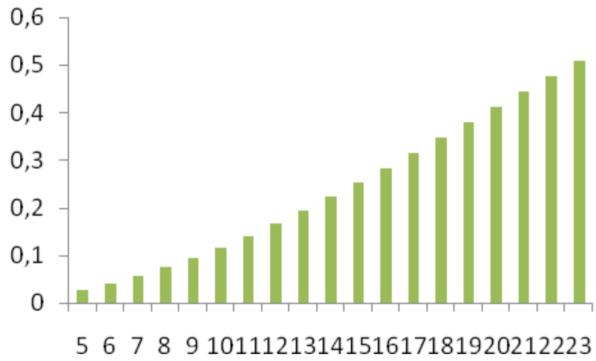
Después de estas experiencias, nuestro tutor no dijo que las preguntas que nos había planteado se corresponden con un fenómeno que se llama *la Paradoja del cumpleaños* y que se trataba de una teoría que se describió en 1938 en el marco de la teoría de estimación del total de población de peces en un lago, de **Zoe Emily Schnabel.**

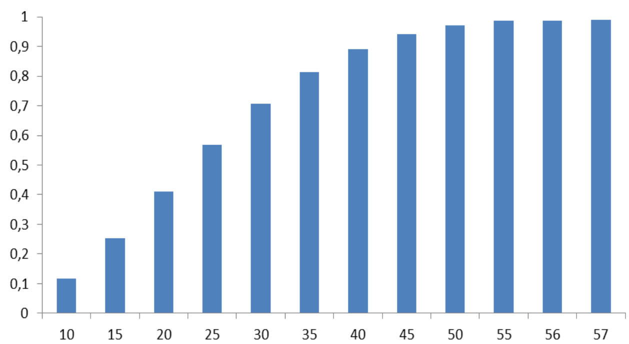
Nuestro tutor nos presentó la explicación de la paradoja. Como punto final a nuestra primera paradoja, construimos una tabla con Excel y calculamos la probabilidad de que en un grupo formado por 2, 3, 4, 5,...personas 2 de ellas hayan nacido en el mismo mes y día. Dichos datos están recogidos en la siguiente tabla:

En la primera columna aparece el número de personas de la muestra; en la segunda, la probabilidad de que 2 personas no cumplan años en el mismo día; y en la tercera, la probabilidad de que 2 personas cumplan el mismo día.

Se puede observar que el resultado de nuestro experimento es bastante próximo al valor real.

n	p	1-p
1	1	0
2	0,997	0,003
3	0,991	0,009
4	0,983	0,017
5	0,972	0,028
10	0,883	0,117
15	0,747	0,253
20	0,588	0,412
21	0,556	0,444
22	0,524	0,476
23	**0,492**	**0,508**
24	0,461	0,539
25	0,431	0,569
30	0,293	0,707
35	0,185	0,816
40	0,108	0,892
45	0,059	0,941
50	0,029	0,971
55	0,013	0,987
56	0,011	0,989
57	**0,009**	**0,991**





El error de nuestro cálculo venía dado porque nosotros nos basábamos en ambos casos, tanto para el de 23 como para el de 57 personas, en que hallábamos la probabilidad de que uno de los miembros del grupo cumpliera años el mismo día que otro que habíamos fijado previamente, y en ese caso las probabilidades son del 0,27 % y 6 %, respectivamente, que es bastante dispar al real.

Nuestro tutor, a modo de curiosidad, nos comentó que entre los 21 monarcas españoles desde los Reyes Católicos o entre los 45 Presidentes de los Estados Unidos existen coincidencias de fechas de nacimiento.

Una vez finalizado nuestro estudio, hemos concluido que la intuición en muchos casos nos lleva al error.

4. El *Juego del cerdo*

La siguiente actividad que nos propuso nuestro tutor fue jugar al *Juego del cerdo*.

El Juego del cerdo consiste en lanzar un dado hasta que el jugador decide detenerse y en cuyo caso se suman los puntos obtenidos o hasta que obtiene un "6", en cuyo caso, pierde el turno y no suma nada. Gana aquel jugador que alcanza primero los 100 puntos.

Nuestro tutor nos preguntó qué estrategia deberíamos seguir para ganar. En vista de que no sabíamos muy bien que método era el mejor para vencer, él nos pidió que realizáramos con Excel distintas **simulaciones** para ver cuál sería la mejor estrategia a seguir. Para la realización de dichas simulaciones hemos utilizado las funciones, ALEATORIO(), ENTERO y SI.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
2	**Tirada 1ª**	**3**	**5**	**3**	**6**	**3**	**1**	**6**	**3**	**3**	**1**	**5**	**4**	**4**	**3**	**6**	**5**
3		3	5	3	0	3	1	0	3	3	1	5	4	4	3	0	5
4	Sumas parciales		5	17	14	17	30	42	45	45	54	58	73	87	86	95	100
5																	16
6	**Tirada 2º**	**6**	**1**	**6**	**5**	**5**	**5**	**4**	**6**	**1**	**6**	**4**	**5**	**6**	**2**	**3**	**5**
7		0	1	0	0	5	5	0	0	1	0	4	5	0	2	0	5
8	Sumas parciales	0	6	0	0	22	35	0	0	46	0	62	78	0	88	0	105
9	Sumas acumuladas		6	14	14	22	35	42	42	46	53	62	78	83	88	95	105
10																	16
11	**Tirada 3º**	**2**	**5**	**3**	**6**	**3**	**3**	**2**	**4**	**2**	**1**	**3**	**1**	**4**	**2**	**4**	**1**
12		0	5	0	0	3	3	0	0	2	0	3	1	0	2	0	1
13	Sumas parciales	0	6	0	0	8	8	0	0	3	0	7	6	0	4	0	6
14	Sumas acumuladas		11	14	14	25	38	42	42	48	53	65	79	83	90	95	106
15																	16
16	**Tirada 4º**	**3**	**3**	**1**	**6**	**4**	**4**	**3**	**1**	**5**	**6**	**4**	**4**	**6**	**5**	**6**	**3**
17		0	3	0	0	4	4	0	0	5	0	4	4	0	5	0	3
18	Sumas parciales	0	9	14	14	12	10	42	42	9	53	13	13	83	10	95	13
19	Sumas acumuladas	0	14	14	14	29	42	42	42	53	53	69	83	83	95	95	109
20																	16
21	número de tiradas = 16																

Elaboramos distinta hojas de cálculo en las que simulamos:

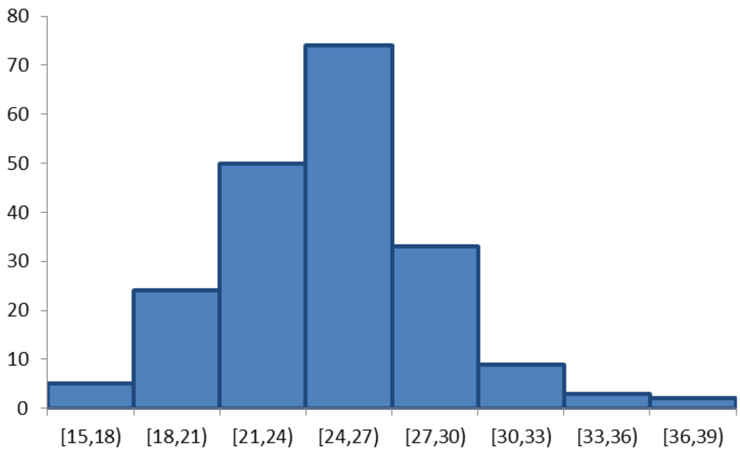
- Detenernos después de realizar un lanzamiento.

- Detenernos después de realizar 2 lanzamientos.
- Detenernos después de realizar 3 lanzamientos.
- Detenernos después de realizar 4 lanzamientos.
- Detenernos después de realizar 5 lanzamientos.
- Detenernos después de realizar 6 lanzamientos.

Repetimos las simulaciones 200 veces en cada modelo y los resultados están recogidos en las tablas y gráficos adjuntos:

Nos detenernos después de realizar un lanzamiento.

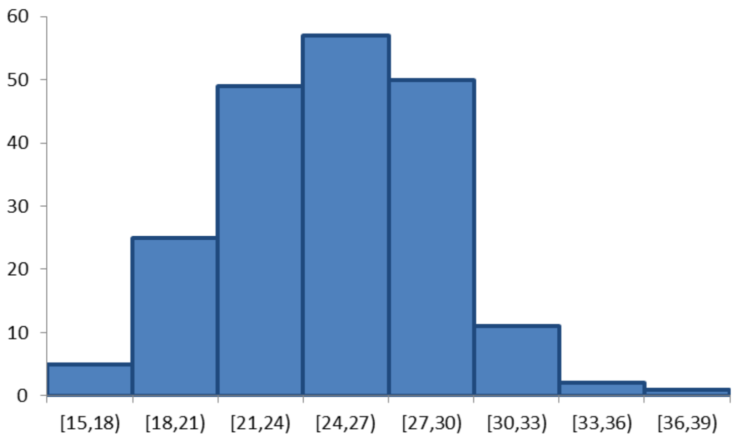
Nº. de turnos	n_i
[15,18)	5
[18,21)	24
[21,24)	50
[24,27)	74
[27,30)	33
[30,33)	9
[33,36)	3
[36,39)	2



El número medio de turnos necesarios para alcanzar los 100 puntos fue de 24,4 y la desviación típica de 3,73. El número de turnos que más se repitió fue 26, siendo 16 el número mínimo de turnos necesarios para alcanzar los 100 puntos y 38 el número máximo de turnos utilizados para alcanzar la victoria.

Nos detenernos después de realizar dos lanzamientos.

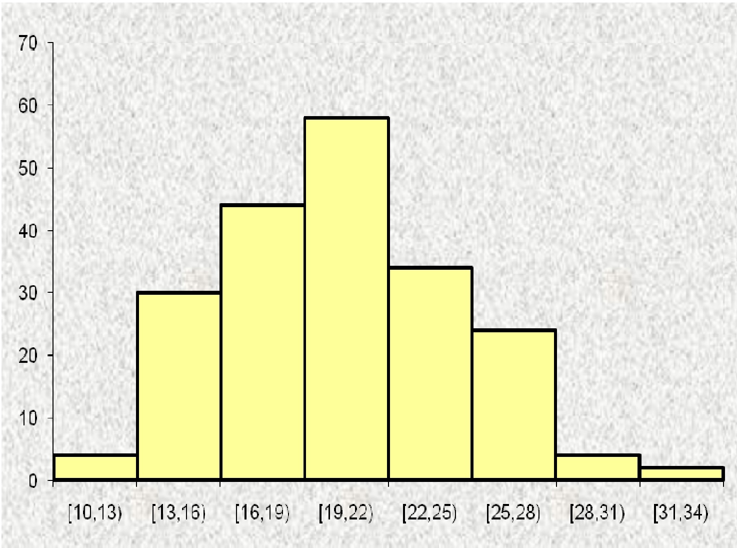
Nº. de turnos	n_i
[15,18)	5
[18,21)	25
[21,24)	49
[24,27)	57
[27,30)	50
[30,33)	11
[33,36)	2
[36,39)	1



El número medio de turnos necesarios para alcanzar los 100 puntos fue de 24,5 y la desviación típica de 3,68. El número de turnos que más se repitió fue 28, siendo 16 el número mínimo de turnos necesarios para alcanzar los 100 puntos y 36 el número máximo de turnos utilizados para alcanzar la victoria.

Analizamos ahora qué ocurre si lanzamos 3 veces el dado en cada turno.

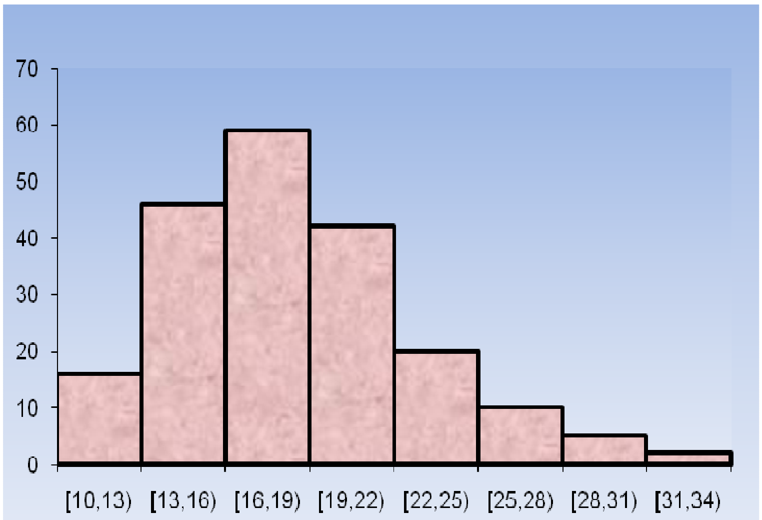
Nº. de turnos	n_i
[10,13)	4
[13,16)	30
[16,19)	44
[19,22)	58
[22,25)	34
[25,28)	24
[28,31)	4
[31,34)	2



El número medio de turnos necesarios para alcanzar los 100 puntos fue de 19,7 y la desviación típica de 4,06. El número de turnos que más se repitió fue 20, siendo 11 el número mínimo de turnos necesarios para alcanzar los 100 puntos y 34 el número máximo de turnos utilizados para alcanzar la victoria.

Analizamos qué ocurre si lanzamos 4 veces el dado en cada turno.

Nº. de turnos	n_i
[10,13)	16
[13,16)	46
[16,19)	59
[19,22)	42
[22,25)	20
[25,28)	10
[28,31)	5
[31,34)	2

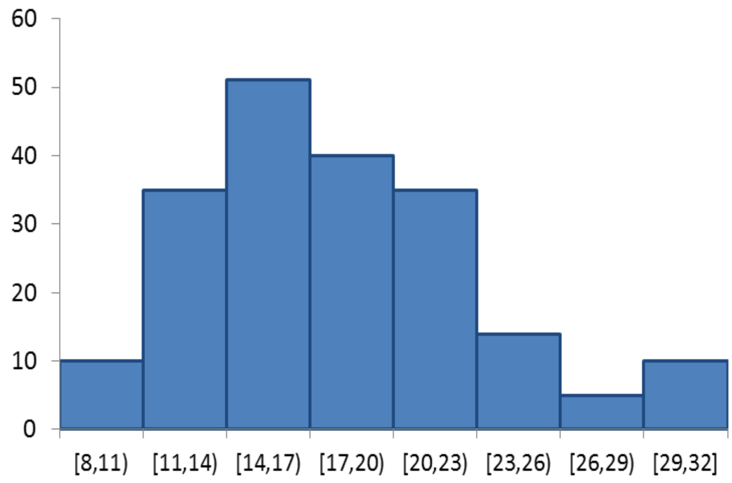


El número medio de turnos necesarios para alcanzar los 100 puntos fue de 17,9 y la desviación típica de 4,38. El número de turnos que más se repitió fue 17, siendo

10 el número mínimo de turnos necesarios para alcanzar los 100 puntos y 34 el número máximo de turnos utilizados para alcanzar la victoria.

Analizamos qué ocurre si lanzamos 5 veces el dado en cada turno.

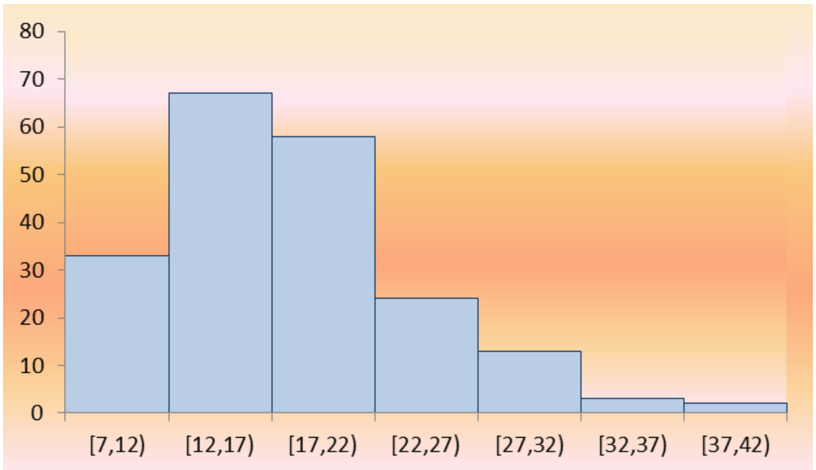
Nº. de turnos	n_i
[8,11)	10
[11,14)	35
[14,17)	51
[17,20)	40
[20,23)	35
[23,26)	14
[26,29)	5
[29,32]	10



El número medio de turnos necesarios para alcanzar los 100 puntos fue de 17,59 y la desviación típica de 5,15. El número de turnos que más se repitió fue 16, siendo 8 el número mínimo de turnos necesarios para alcanzar los 100 puntos y 32 el número máximo de turnos utilizados para alcanzar la victoria.

Analizamos finalmente qué ocurre si lanzamos 6 veces el dado en cada turno.

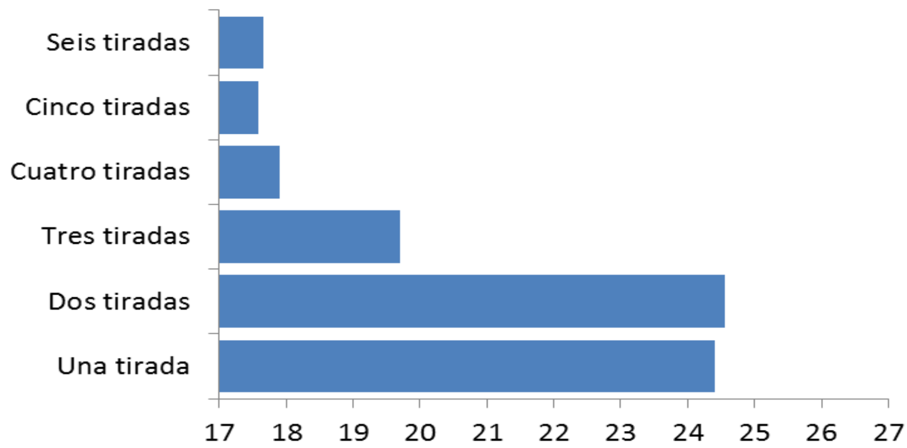
Nº. de turnos	n_i
[7,12)	33
[12,17)	67
[17,22)	58
[22,27)	24
[27,32)	13
[32,37)	3
[37,42)	2



El número medio de turnos necesarios para alcanzar los 100 puntos fue de 17,65 y la desviación típica de 6,04. El número de turnos que más se repitió fue 16, siendo 7 el número mínimo de turnos necesarios para alcanzar los 100 puntos y 42 el número máximo de turnos utilizados para alcanzar la victoria.

También, hemos calculado el número medio de tiradas antes de alcanzar los 100 puntos necesarios para ganar, siguiendo las distintas estrategias que nos hemos planteado. La estrategia con menor número medio de tiradas necesarias es la de detenerse después de realizar 5 lanzamientos y la estrategia con mayor número de tiradas es la de detenerse después de realizar 2 lanzamientos.

En el siguiente diagrama de barras podemos ver el número medio de tiradas necesarias para alcanzar los 100 puntos siguiendo las 6 estrategias que hemos presentado anteriormente.



El número mínimo y máximo de turnos necesarios para alcanzar los 100 puntos fue siguiendo la estrategia de detenerse después de realizar 6 lanzamientos. El mayor rango y la mayor desviación típica se encuentra siguiendo la anterior estrategia.

De forma conjunta hemos simulado 1200 partidas: 200 siguiendo cada estrategia, y el número medio de turnos necesarios para alcanzar los 100 puntos fue de 20,26. El valor que más se repite en estas 1200 partidas es 22. El mayor número de lanzamientos hasta llegar a los 100 puntos fue 42 y el menor 7. La desviación típica es 5,507.

Después de todo lo analizado, creemos que el mejor método para alcanzar la victoria sería realizar 5 lanzamientos y la peor realizar 2 lanzamientos.

Aunque no lo hemos realizado, hubiera sido interesante analizar que en vez de tener en cuenta el número de lanzamientos, sería haber tenido en cuenta si la suma de los valores obtenidos daba un determinado valor.

La estadística nos permite establecer estrategias para ganar en un juego en el que aparentemente solo influye el azar.

5. Ley de Benford

Una vez que habíamos resuelto la *Paradoja* del cumpleaños y diseñado un método más o menos fiable para ganar al *Juego del cerdo,* nuestro tutor nos preguntó:

¿La cantidad de números que empiezan por 1 ó 2 ó 3……en vuestros libros de historia o de biología, de física o de química son más o menos la misma? ¿O tal vez hay muchos más números que comienzan por una determinada cifra? Y en los folletos publicitarios de un supermercado o en de una tienda de deportes ¿ocurre lo mismo?

En un principio nos pareció que no existía ningún motivo que llevara a pensar que hay más números que empiezan por una determinada cifra. Es posible que en el libro de Historia hubiera más cifras que empezaran por el 3 y en el de Biología por 2 o en el de Física por 6, y que en los folletos publicitarios, pues dependería si se trata de un folleto de un supermercado de alimentación o de un folleto de una tienda de deportes, e incluso también dependerá de si se trata de una época de rebajas. En principio creíamos que no existía nada que hiciera pensar que los números que aparecen de forma más cotidiana en nuestras vidas empiezan con más frecuencia por una cifra que por otra y así se lo expusimos a nuestro profesor. Él nos dijo que comprobáramos si nuestra suposición era o no cierta. Para ello realizamos el siguiente trabajo:

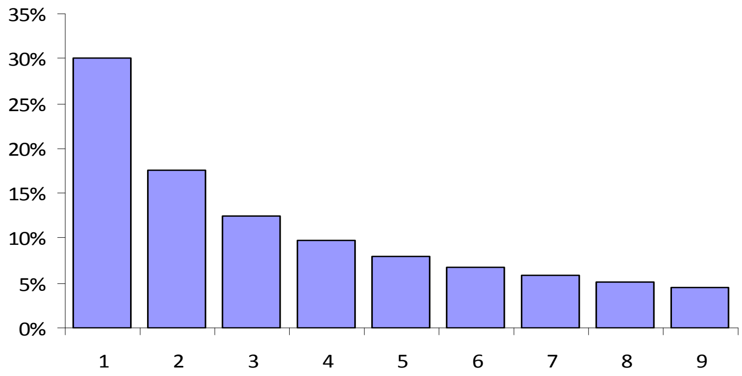
Cogimos nuestro libro de Química y metimos en una hoja de cálculo los 100 primeros datos numéricos que aparecen (fechas de nacimiento, constantes químicas,…) y vimos el número de valores que empiezan por “1” por “2”, por “3”,…Para ello utilizamos las funciones *CONTAR.SI*, IZQUIERDA y VALOR. Con los datos recogidos elaboramos la correspondiente tabla de frecuencias y el diagrama de barras. Repetimos este mismo proceso con el libro de Historia de España, el de Matemáticas y el de Ciencias de la Tierra y el Medio Ambiente.

Observamos que, salvo en el libro de Historia de España en el resto se mantenían parecidos los porcentajes de números que empezaban por “1”, por “2”, por “3”,… en todos los libros y que además el número de valores que empezaban por 1 era mayor que el resto. Este dato era especialmente significativo en el de historia de España.

Nuestro tutor nos mandó buscar en que consistía la ley Benford. Investigamos en internet y encontramos que dicha ley se trata de una curiosa teoría matemática que predice que en un conjunto determinado de números, aquellos, cuyo primer dígito es 1 aparecerán, de forma más frecuente que los números que empiezan por otros dígitos. La distribución de los primeros dígitos es bastante asimétrica**.** La frecuencia esperada para números que empiezan por uno es casi del 30%**;** para los que empiezan por 2 se aproxima al 17%**;** los que empiezan por 3 son algo más del 12 % y así sucesivamente.

En la siguiente tabla y gráfico ponemos la ley Benford extraída de la web http://www.estadisticaparatodos.es/taller/benford/benford.html

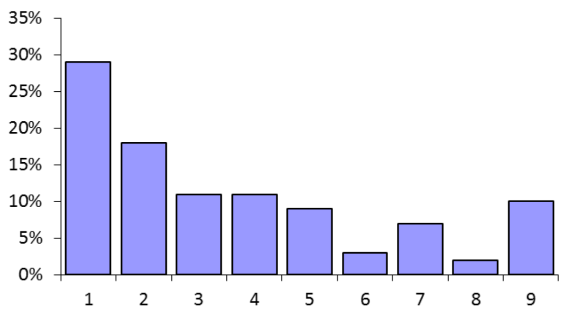
x_i	%
1	30,10%
2	17,61%
3	12,46%
4	9,69%
5	7,92%
6	6,69%
7	5,80%
8	5,12%
9	4,58%

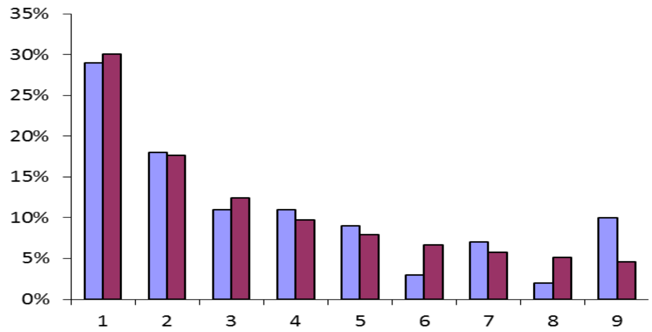


En las tablas y gráficos siguientes presentamos nuestros resultados y su comparación con la ley Benford.

x_i	f_i	%	Benford	Diferencia
1	29	29%	30,10%	**-1,10%**
2	18	18%	17,61%	**0,39%**
3	11	11%	12,46%	**-1,46%**
4	11	11%	9,69%	**1,31%**
5	9	9%	7,92%	**1,08%**
6	3	3%	6,69%	**-3,69%**
7	7	7%	5,80%	**1,20%**
8	2	2%	5,12%	**-3,12%**
9	10	10%	4,58%	**5,42%**

[1] Datos recogidos en el libro de Química



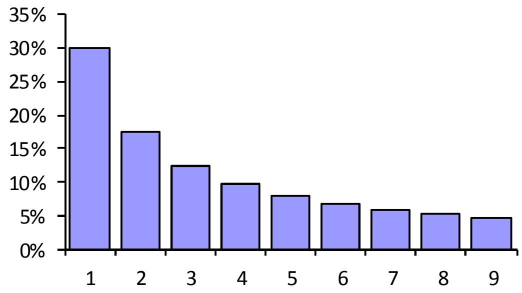


[2] Diagrama de Barras de Química

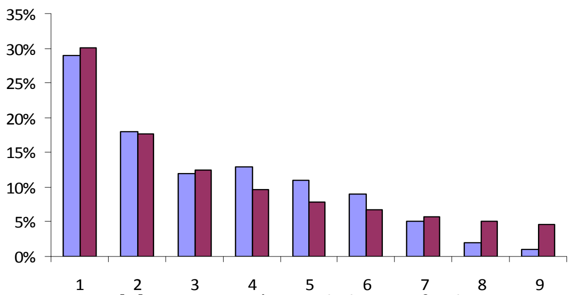
[3] Comparación con la ley Benford

Xi	Fi	%	Benford	Diferencia
1	29	29%	30,10%	**-1,10%**
2	18	18%	17,61%	**0,39%**
3	12	12%	12,46%	**-0,46%**
4	13	13%	9,69%	**3,31%**
5	11	11%	7,92%	**3,08%**
6	9	9%	6,69%	**2,31%**
7	5	5%	5,80%	**-0,80%**
8	2	2%	5,12%	**-3,12%**
9	1	1%	4,58%	**-3,58%**

[4] Datos recogidos en el libro de Matemáticas



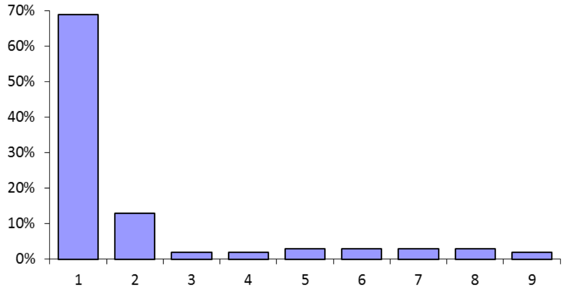
[5] Diagrama de barras de Matemáticas



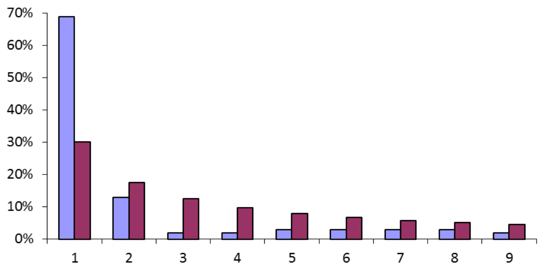
[6] Comparación con la ley Benford

x_i	f_i	%	Benford	Diferencia
1	69	69%	30,10%	**38,90%**
2	13	13%	17,61%	**-4,61%**
3	2	2%	12,46%	**-10,46%**
4	2	2%	9,69%	**-7,69%**
5	3	3%	7,92%	**-4,92%**
6	3	3%	6,69%	**-3,69%**
7	3	3%	5,80%	**-2,80%**
8	3	3%	5,12%	**-2,12%**
9	2	2%	4,58%	**-2,58%**

[7] Datos recogidos en el libro de Historia de España



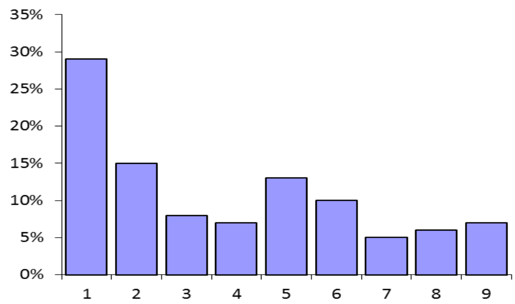
[8] Diagrama de barras de Historia de España

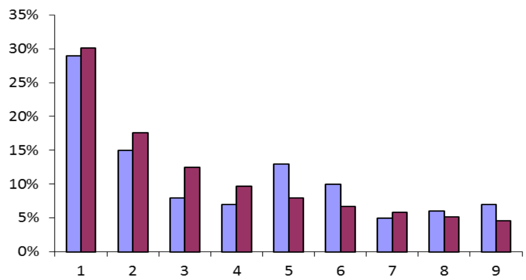


[9] Comparación con la ley Benford

x_i	f_i	%	Benford	Diferencia
1	29	29%	30,10%	**-1,10%**
2	15	15%	17,61%	**-2,61%**
3	8	8%	12,46%	**-4,46%**
4	7	7%	9,69%	**-2,69%**
5	13	13%	7,92%	**5,08%**
6	10	10%	6,69%	**3,31%**
7	5	5%	5,80%	**-0,80%**
8	6	6%	5,12%	**0,88%**
9	7	7%	4,58%	**2,42%**

[10] Datos recogidos en el libro de C.T.M.A.





Observamos que, salvo en el libro de Historia de España, en el resto se mantenían parecidos los porcentajes de números que empezaban por 1, por 2, por 3,… en todos los libros, y que además el número de valores que empezaban por 1 era mayor que el resto. Este dato era especialmente significativo en el de Historia de España.

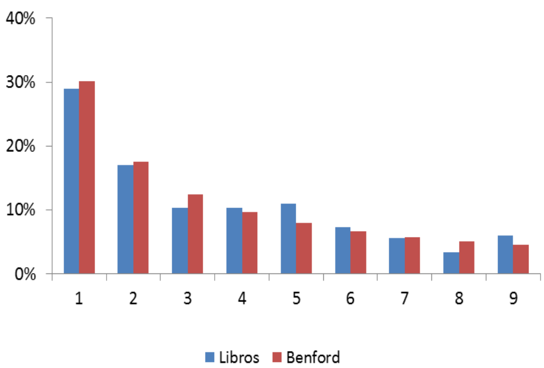
También analizamos conjuntamente lo que pasaba con todos los libros excepto el de Historia de España.

x_i	f_i	Libros (%)	Benford	Diferencia
1	156	39%	30,10%	**8,90%**
2	64	16%	17,61%	**-1,61%**
3	33	8%	12,46%	**-4,21%**
4	33	8%	9,69%	**-1,44%**
5	36	9%	7,92%	**1,08%**
6	25	6%	6,69%	**-0,44%**
7	20	5%	5,80%	**-0,80%**
8	13	3%	5,12%	**-1,87%**
9	20	5%	4,58%	**0,42%**

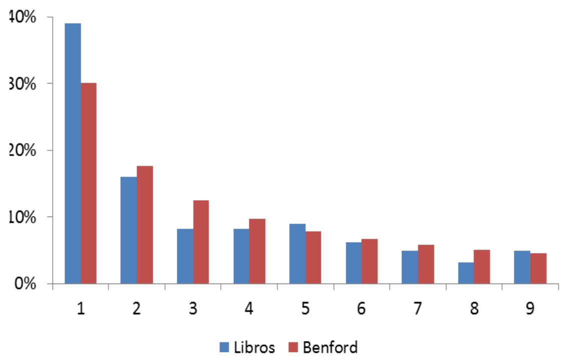
[11] Datos recogidos de los cuatro textos

x_i	f_i	Libros(%)	Benford	Diferencia
1	87	29,0%	30,10%	**-1,10%**
2	51	17,0%	17,61%	**-0,61%**
3	31	10,3%	12,46%	**-2,13%**
4	31	10,3%	9,69%	**0,64%**
5	33	11,0%	7,92%	**3,08%**
6	22	7,3%	6,69%	**0,64%**
7	17	5,7%	5,80%	**-0,13%**
8	10	3,3%	5,12%	**-1,79%**
9	18	6,0%	4,58%	**1,42%**

[12] Datos recogidos sin tener en cuenta el libro de Historia



[13] Diagrama correspondiente a la tabla [12]

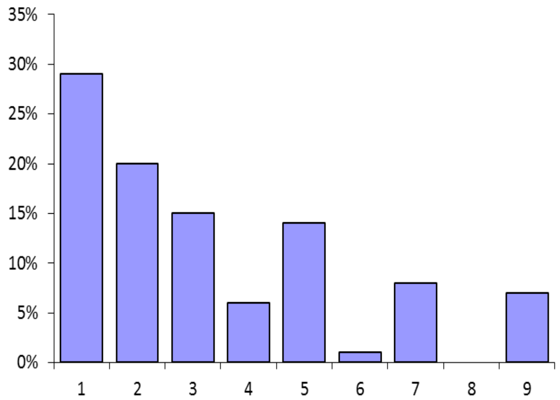


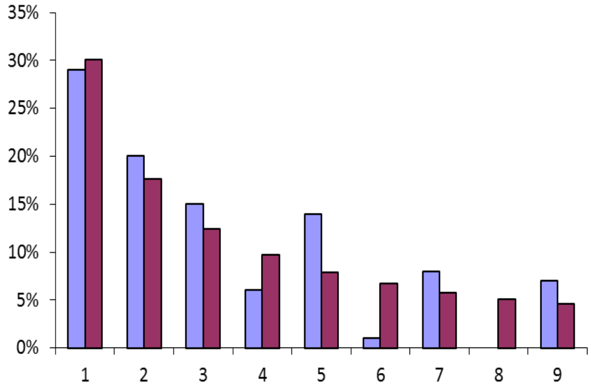
[14] Diagrama correspondiente a la tabla [11]

Además de analizar algunos de nuestros libros, analizamos los catálogos de las tiendas Warmart, Game, Decathlón Fnac y H&M, y realizamos el mismo proceso que habíamos hecho con nuestros libros de texto. Los resultados aparecen en las siguientes tablas y gráficos:

x_i	f_i	Fnac (%)	Benford	Diferencia
1	29	29,0%	30,10%	**-1,10%**
2	20	20,0%	17,61%	**2,39%**
3	15	15,0%	12,46%	**2,54%**
4	6	6,0%	9,69%	**-3,69%**
5	14	14,0%	7,92%	**6,08%**
6	1	1,0%	6,69%	**-5,69%**
7	8	8,0%	5,80%	**2,20%**
8	0	0,0%	5,12%	**-5,12%**
9	7	7,0%	4,58%	**2,42%**

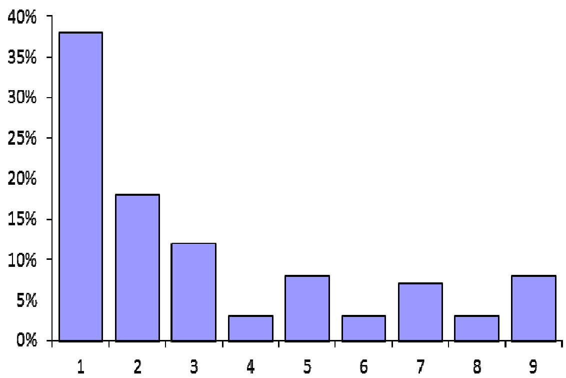
[1] Datos recogidos del catálogo de Fnac y H&M

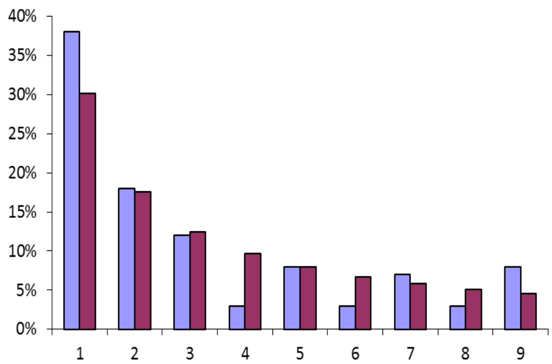




Xi	f_i	Walmart (%)	Benford	Diferencia
1	38	38%	30,10%	**7,90%**
2	18	18%	17,61%	**0,39%**
3	12	12%	12,46%	**-0,46%**
4	3	3%	9,69%	**-6,69%**
5	8	8%	7,92%	**0,08%**
6	3	3%	6,69%	**-3,69%**
7	7	7%	5,80%	**1,20%**
8	3	3%	5,12%	**-2,12%**
9	8	8%	4,58%	**3,42%**

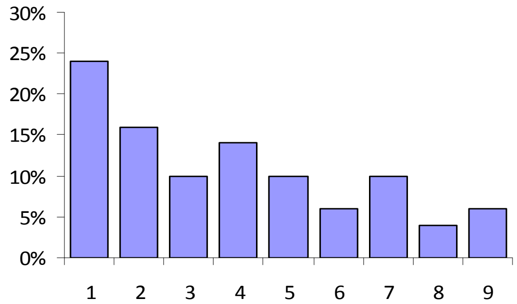
[2] Datos recogidos del catálogo de Walmart

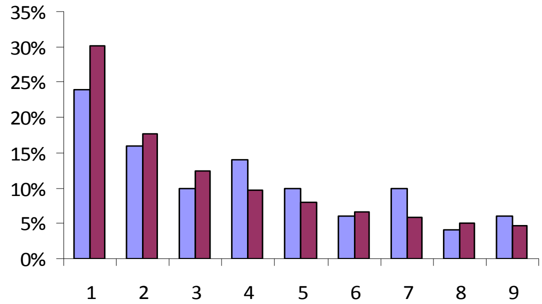




x_i	f_i	Decathlon (%)	Benford	Diferencia
1	24	24%	30,10%	**-6,10%**
2	16	16%	17,61%	**-1,61%**
3	10	10%	12,46%	**-2,46%**
4	14	14%	9,69%	**4,31%**
5	10	10%	7,92%	**2,08%**
6	6	6%	6,69%	**-0,69%**
7	10	10%	5,80%	**4,20%**
8	4	4%	5,12%	**-1,12%**
9	6	6%	4,58%	**1,42%**

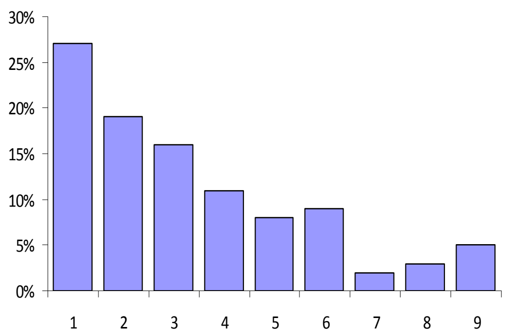
[3] Datos recogidos del catálogo de Decathlón

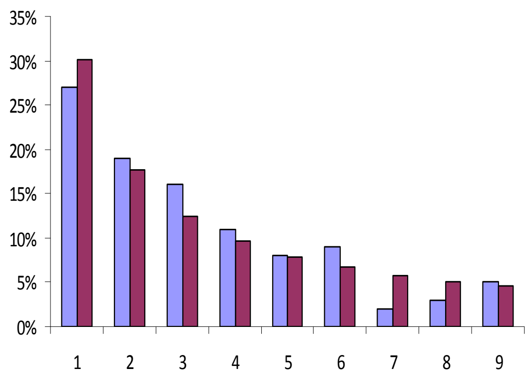




x_i	f_i	Game (%)	Benford	Diferencia
1	27	27%	30,10%	**-3,10%**
2	19	19%	17,61%	**1,39%**
3	16	16%	12,46%	**3,54%**
4	11	11%	9,69%	**1,31%**
5	8	8%	7,92%	**0,08%**
6	9	9%	6,69%	**2,31%**
7	2	2%	5,80%	**-3,80%**
8	3	3%	5,12%	**-2,12%**
9	5	5%	4,58%	**0,42%**

[4] Datos recogidos del catálogo de Game



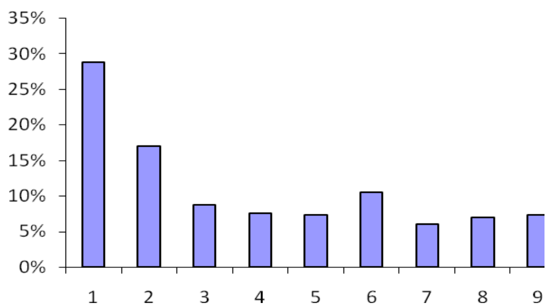


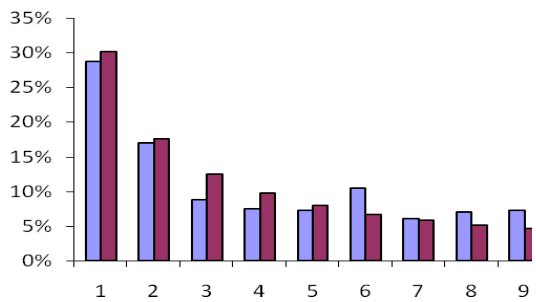
En la tabla y gráfico que aparecen a continuación están recogidos los 400 precios de los productos que hemos analizado y su comparación con la ley Benford.

Finalmente, solicitamos a familiares, amigos y profesores que nos dieran cuatrocientos números y repetimos el mismo proceso que habíamos seguido con los números que aparecían en los textos y en los folletos publicitarios y el resultado fue:

x_i	f_i	Azar (%)	Benford	Diferencia
1	115	28,8%	30,10%	**-1,35%**
2	68	17,0%	17,61%	**-0,61%**
3	35	8,8%	12,46%	**-3,71%**
4	30	7,5%	9,69%	**-2,19%**
5	29	7,3%	7,92%	**-0,67%**
6	42	10,5%	6,69%	**3,81%**
7	24	6,0%	5,80%	**0,20%**
8	28	7,0%	5,12%	**1,88%**
9	29	7,3%	4,58%	**2,67%**

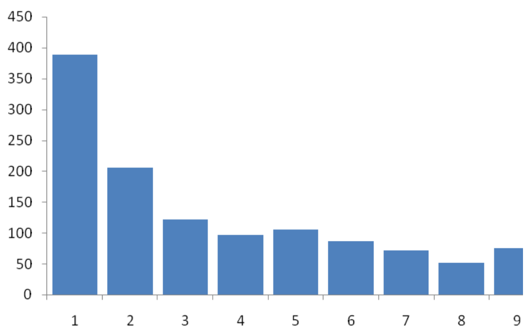
[5] Datos recogidos de los catálogos

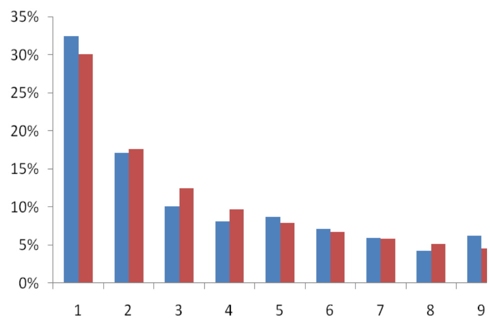




En la tabla que presentamos a continuación, queremos reflejar de forma conjunta todos los datos recogidos de libros, de folletos publicitarios y del azar.

x_i	f_i	Datos (%)	Benford	Diferencia
1	389	32,42%	30,10%	**2,32%**
2	205	17,08%	17,61%	**-0,53%**
3	121	10,08%	12,46%	**-2,38%**
4	97	8,08%	9,69%	**-1,61%**
5	105	8,75%	7,92%	**0,83%**
6	86	7,17%	6,69%	**0,48%**
7	71	5,92%	5,80%	**0,12%**
8	51	4,25%	5,12%	**-0,87%**
9	75	6,25%	4,58%	**1,67%**





Y como se puede observar casi se cumple la Ley Benford

Finalmente hemos hecho un resumen del mecanismo de la Ley Benford, que a continuación presentamos como parte del trabajo.

Quien primero se dio cuenta de este fenómeno fue, en 1881, el matemático y astrónomo Simon Newcomb. Un día Newcomb estaba usando un libro de logaritmos y se dio cuenta de que las páginas del libro estaban más viejas y usadas cuanto más cercanas estaban del principio. En aquella época, se utilizaban las tablas de logaritmos, entre otras cosas para multiplicaciones entre grandes números. Actualmente, equivaldría a examinar el desgaste de la tecla "1" en cajas registradoras o calculadoras. ¿A qué se debía? Sólo podía tener una explicación: a lo largo de los años había consultado mucho más el logaritmo de los números que comenzaban por 1 que de los que comenzaban por números más altos.

Simon Newcomb, astrónomo, dedujo que los dígitos iniciales de los números (al menos los utilizados en su trabajo que provenían de la observación de los astros principalmente) no son equiprobables sino que el 1 aparece como dígito inicial más frecuente seguido del 2 etc. hasta el 9 que es el menos frecuente. Mediante un breve e ingenioso razonamiento, aunque sin presentar realmente un argumento formal ni fórmula matemática, Newcomb enunció verbalmente una Ley de probabilidad. En 1938, Frank Benford, un físico de la compañía General Electric, se dio cuenta del mismo patrón. Estudió más de 20000 datos provenientes de constantes, magnitudes físicas, longitudes de ríos, estadísticas de deportes...y constató que la probabilidad de que un número en una serie de datos comience por el dígito d es de $P[d] = \log(1 + 1/d)$. Según la anterior ley, la probabilidad de que en una serie de datos el primer digito de un número sea 1 es del 30%, del 17,6% para un 2, 12,5% para el 3, y así va decreciendo hasta el 9 que es del 4,6%...

6. El problema de Monty Hall

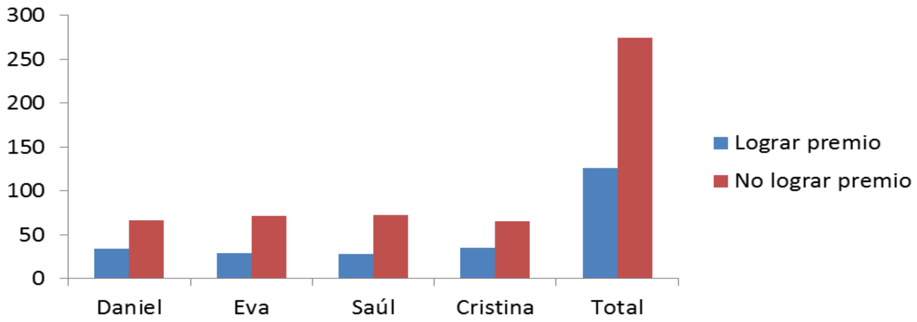
La última actividad que nos presentó nuestro tutor estaba relacionada con un concurso de televisión.

Imaginar que estáis en un concurso de televisión y os encontráis con el presentador frente a 3 puertas cerradas. Detrás de una de estas puertas hay un coche, y detrás de las otras hay 2 cabras. Vuestro objetivo es adivinar en cuál de ellas está el coche escondido, sin tener ninguna pista disponible. Después de elegir una, el presentador abre una de las otras 2 puertas donde esta una cabra. Los interrogantes de nuestro tutor fueron: ¿Debemos mantener nuestra primera opinión o cambiamos a la otra puerta restante? ¿Hay alguna diferencia? ¿O da lo mismo que hagamos una cosa u otra?

A nosotros nos pareció que el hecho de cambiar o no de puerta no tenía importancia y que únicamente el azar era el que influía en la posibilidad de lograr o no el premio.

Nuestro tutor nos pidió que cada uno de los 4 componentes realizáramos una simulación de la actividad utilizando un applet que había en la www.estadisticaparatodos.com Nos pidió que escogiéramos una puerta y que después de que se abriera automáticamente una de las 2 no elegidas y en las que había un mal premio que conserváramos nuestra puerta. Anotamos los casos en los que obteníamos el premio y el resultado fue:

Conservando la puerta elegida						
	Daniel	Eva	Saúl	Cristina	Total	%
Lograr premio	34	29	28	35	126	31,5%
No lograr premio	66	71	72	65	274	68,5%



Como en las situaciones anteriores, investigamos en internet para conocer el origen de este curioso problema y descubrimos que está inspirado en un concurso televisivo estadounidense Let's Make a Deal (Hagamos un trato). Este programa fue famoso entre los años 1963 y 1986. Los concursantes escogían una puerta de entre tres y el presentador, Monty Hall, abría una de las no escogidas por el concursante detrás de la cual había una cabra y volvía a preguntar al concursante. Si deseaba cambiar o continuar con la puerta elegida.

A continuación, vamos a explicar por qué es mejor cambiar la puerta una vez que lo pregunta el presentador:

Imagen extraída de www.estadisticaparatodos.es

Como se puede ver en el anterior diagrama la probabilidad de ganar cambiando de prueba es doble que si no cambiamos de puerta.

7. Agradecimientos

Queremos agradecer a nuestro tutor el tiempo que le ha dedicado a ayudarnos en la elaboración de este trabajo. También damos las gracias a la profesora del Departamento de Inglés Gema Hernández López, a José Antonio Pérez Sánchez, profesor de Lengua Castellana y literatura de nuestro instituto y a Pablo Rodríguez Alonso, profesor del Departamento de Geografía e Historia por su colaboración.

También agradecemos a Maria Cristina Acuña Cuervo, Jefa de Estudios del Centro, que nos facilitó las fechas de nacimiento de los alumnos del instituto.

Y por último, queremos dar las gracias a la Consejería de Educación del Principado de Asturias y al Departamento de Estadística e Investigación Operativa y Didáctica de las Matemáticas de la Universidad de Oviedo por brindarnos la posibilidad de participar en este concurso.

8. Propuestas de mejora

Respecto al juego del cerdo, podíamos haber estudiado que pasaba cuando con la suma de las puntuaciones obteníamos un determinado valor y no simplemente nos deteníamos después de realizar un número concreto de tiradas.

En relación con la Ley Benford, creemos que deberíamos haber utilizado un libro distinto al de Historia de España ya que en éste hay muchos valores que empiezan por uno.

Debido a la falta de tiempo creemos que no hemos profundizado lo suficiente con el problema de Montty Hall. Podíamos haber preguntado a nuestros profesores, padres y amigos en relación a este problema, con lo cual sabríamos lo que pensaba la gente en relación a este experimento y si creían que influía únicamente el azar o se podía seguir una determinada estrategia para que fuera más fácil lograr el premio gordo.

Creemos que quizás hubiera sido interesante para darle un carácter divulgativo a nuestro trabajo haber puesto murales en los pasillos del instituto con los juegos y paradojas que nos planteaba Arturo para fomentar un espíritu reflexivo entre la comunidad escolar.

9. Referencias bibliográficas

[1] http://www.rae.es/

[2] http://www-fa.upc.es/websfa/fluids/TJM/pdf/Paradojas_acertijos_y_demostraciones_invalidas.pdf

[3] http://www.losmundialesdefutbol.com/mundiales/2014_mundial.php

[4] El libro de las matemáticas. Autor Clifford A. Pickover. Ed. Librero. Pag.386

[5] http://www.wikipedia.org

[6] http://www.estadisticaparatodos.es/taller/benford/benford.html

[7] Webs de distintas tiendas de donde sacamos los catálogos de precios.

[8] http://www.abc.es/espana/20130206/abci-papeles-barcenas-falsos-segun-201302060805.html

[9] http://www.estadisticaparatodos.es/taller/montyhall/montyhall.html