

## Fútbol y Estadística

Álvaro Valdés Menéndez, Profesor de Matemáticas del I.E.S. "Pérez de Ayala", de Oviedo (Asturias)



El 10 de Junio de 2006, durante la celebración del Mundial de Fútbol en Alemania, el seleccionador de Trinidad y Tobago Leo Beenhakker declaró en una rueda de prensa:

*Se ha probado que lo que está en el papel, pocas veces se traspasa al terreno de juego. Esto es fútbol, no matemáticas, y aquí dos más dos, rara vez suelen ser cuatro, suelen ser tres o cinco. No puedo estar más orgulloso de mis jugadores, lo dieron todo en el campo.*

¿Realmente están tan alejados el fútbol y la matemática? El objetivo de este artículo es plantear tres ejemplos de utilización del fútbol en clase.



### Problema 1

Analicemos en primer lugar la distribución normal. Esta distribución explica teóricamente la mayoría de los sistemas naturales, y se busca con este ejemplo que los alumnos lo descubran por sí mismos:

**“La tabla 1 muestra el porcentaje de variantes, equis y doses en las quinielas a lo largo de la historia. Representa gráficamente las tres variables y di qué información puedes extraer de las mismas”**

Introduciendo los datos en la calculadora:

	LiSt 1	LiSt 2	LiSt 3	LiSt 4
SUB		UTES	EQUIS	DOSES
1	1	0.21	5.61	16.28
2	2	1.09	12.79	24.01
3	3	3.56	21.27	23.8
4	4	7.73	23.6	15.18

GRPH CALC TEST INTR DISTR

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Variantes	0,0	0,21	1,09	3,56	7,73	13,27	18,33	19,41	16,69	11,70	6,02	1,64	0,21	0,07	0,07
Nº "X"	1,23	5,61	12,79	21,27	26,60	17,10	11,83	4,04	1,85	0,55	0,14				
Nº "2"	5,27	16,28	24,01	23,80	15,18	9,78	3,35	1,64	0,68						

Tabla 1.- Porcentaje de variantes en las quinielas

es inmediato obtener la representación gráfica de cada variable ajustando los parámetros asociados:

Pulsando "F1" (GRPH) y "F6" (SET); en la pantalla que aparece

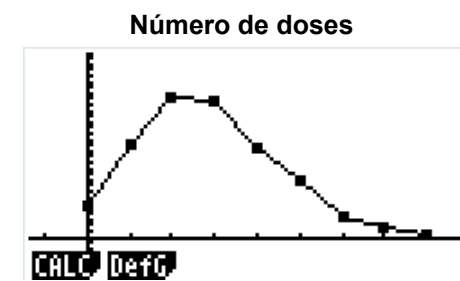
```
StatGraph1
Graph Type : Scatter
XList      : List1
YList      : List2
Frequency  : 1
Mark Type  : □

|GPH1| |GPH2| |GPH3|
```

se modifican:

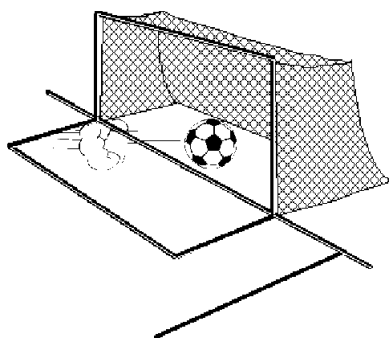
- "Graph Type": seleccionar "xyLine"
- "XList": dejar "List1"
- en "YList" cambiar a:
  - "List2" para el número de variantes (GPH1)
  - "List3" para el número de equis (GPH2)
  - "List4" para el número de doses (GPH3)

Con los ajustes hechos, se obtienen las gráficas:



El acuerdo cualitativo con la gráfica de la Distribución Normal es apreciable, incluso teniendo una variable discreta y con pocos valores.

Si se quisiera ir un poco más lejos, los alumnos de 1º o 2º de Bachillerato podrían obtener la media y la desviación típica para comparar las frecuencias relativas acumuladas obtenidas a partir de la tabla con las teóricas. Por su parte, los alumnos de cursos universitarios podrían realizar *test* de ajuste a la normal ( $\chi^2$ , Kolmogorov, Lilliefors, ...).



### Problema 2

Una vez vista la validez de la aproximación normal, se puede intentar responder a otras preguntas. ¿Quién tiene razón en la eterna polémica sobre la primacía del espectáculo o de los resultados en el fútbol? Averigüémoslo en un estudio de la relación entre el número de goles (la "salsa" del fútbol) y el total de puntos conseguidos:

**“Analiza la clasificación de la última temporada, y determina si existe alguna relación entre el número total de goles marcados y los puntos totales obtenidos”**

	GF	PT
1 Real Madrid C.F.	66	76
2 F.C. Barcelona	78	76
3 Sevilla F.C.	64	71
4 Valencia C.F.	57	66
5 Villarreal C.F.	48	62
6 R. Zaragoza C.D.	55	60
7 At. de Madrid	46	60
8 R.C.R. de Huelva	54	54
9 Getafe C.F.	39	52
10 Real Racing Club	42	50
11 R.C.D. Espanyol	46	49
12 R.C.D. Mallorca	41	49
13 R.C. Deportivo	32	47
14 C.At. de Osasuna	51	46
15 Levante U.D.	37	42
16 Real Betis B. S.	36	40
17 Ath. Club Bilbao	44	40
18 R.C. Celta de Vigo	40	39
19 Real Sociedad	32	35
20 C.G. Tarragona	34	28

Tabla 2.- Goles y Puntos en la temporada 2006-07

Introduciendo los datos:

	List 1	List 2	List 3	List 4
SUB	GF	PTOS		
1	66	76		
2	78	76		
3	64	71		
4	57	66		

GP1 GP2 GP3 SEL SET

se obtiene rápidamente la recta de regresión:

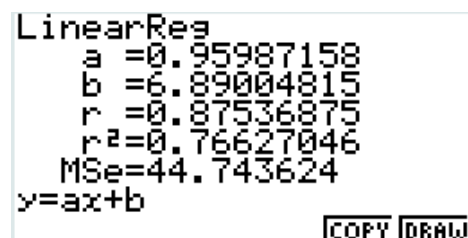
Pulsando “F1” (GRPH) y “F6” (SET); en la misma pantalla descrita antes:

- “Graph Type”: seleccionar “Scatter”
- en “XList” mantener “List1”
- en “YList” cambiar a: “List2”

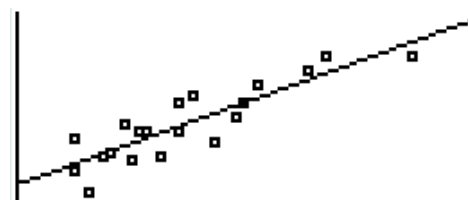
pulsando “EXE” y “GP1” (F1).



Para mostrar la recta de regresión, se debe seleccionar “CALC” (F1), “X” (F2), que muestra la pantalla con los parámetros del ajuste lineal:



y seleccionando “DRAW” (F6) añade la recta de regresión sobre la nube de puntos



Puede concluirse, finalmente, que el número de goles **sí** está relacionado con el total de puntos obtenidos, y que merece la pena buscar con ahínco el gol.

Por supuesto, siempre habrá alguien que piense que este razonamiento puede hacerse a la inversa: **“... existe alguna relación entre el número total de goles encajados y los partidos perdidos ...”**

Para este planteamiento se obtiene:

$$y = 0.3682 \cdot x - 3.24$$

$$R^2 = 0.6366$$

$$r = 0.7979$$

quedando como ejercicio la discusión en clase de qué ajuste es mejor, y qué tipo de juego (ofensivo o defensivo) resulta más rentable.



### Problema 3

El tercer enunciado trata sobre la inferencia estadística. Existen numerosas posibilidades que estudiar, teniendo en cuenta la cantidad de tópicos que circulan sobre el fútbol:

- el equipo local tiene mayor probabilidad de ganar que el visitante
- un jugador juega mejor o peor que en temporadas anteriores
- ...

Centrémonos en el siguiente caso:

**“En la Temporada 1995-96, se introdujo el sistema de puntuación actual, que proporciona tres puntos por cada victoria, frente a los dos del sistema anterior. El objetivo era hacer más atractivo el juego de ataque, aumentando el número de victorias y, por tanto, buscar más goles. Analiza los resultados anteriores a la temporada 94-95 y los posteriores a la 1995-96 para decidir si:**

1. Se logran más victorias que antes
2. Se marcan más goles que antes

**con nivel de significación  $\alpha = 0.05$ ”**

Este es un excelente problema para poner en práctica todos los pasos necesarios en cualquier análisis inferencial:

1. Selección del estadístico a contrastar
2. Muestreo
3. Análisis de los datos obtenidos
4. Obtención del estadístico
5. Determinar la función de distribución asociada al contraste
6. Resolución del contraste de hipótesis

Para este caso se debe invitar a los alumnos a realizar una búsqueda de información, siendo la fuente de los datos utilizados en este problema la página Web de la Liga de Fútbol Profesional:

<http://www.lfp.es>

cuyo histórico permite conocer los resultados y las clasificaciones desde la primera Temporada de Liga en España (1929-30).

En primer lugar se deben definir los conceptos: **“más victorias que antes”** y **“más goles que antes”**. ¿Comparar el número total de victorias logradas y goles marcados a lo largo de la Temporada?

Un segundo problema aparece al determinar qué temporadas son significativas para nuestro estudio. Es un hecho conocido que el fútbol moderno apenas tiene que ver con el que se jugaba a mediados del siglo pasado,<sup>1</sup> así que hay que elegir cuáles se considerarán en el estudio.

Como tercer “contratiempo” aparecen los continuos cambios en el número de equipos que conforman la “Primera División” (desde los 10 equipos en 1929 hasta los 20 actuales, pasando por los 22 de las temporadas 95/96 y 96/97), sin olvidar el “experimento” de la temporada 86/87 de introducir *play-offs* en el fútbol y que llevó a 18 equipos a disputar 44 jornadas.

Finalmente, y más importante, hay que determinar qué contraste se realiza. ¿Igualdad de medias o proporciones en dos muestras? ¿Realizar el *test* de igualdad para TODAS las muestras?

Viendo el temario de *Matemáticas Aplicadas a las CCSS II* de 2º de Bachiller sólo es posible la primera opción, pero alumnos de *Estadística* o Universitarios podrían realizar un análisis **ANOVA** de varias temporadas.

Dicho lo anterior, se decide:

1. Analizar:
  - a) la proporción de partidos que terminan en victoria
  - b) el número medio de goles por partido
2. Comparar las temporadas 1994/95 y 95/96

Los datos son:

Temporada	1994-95	1995-96
Nº Equipos	20	22
Partidos disputados	380	462
Media goles/partido	2.54	2.70
Desviación Típica	1.70	1.80
% Victorias	70.3	71.9

**Tabla 3.- Comparación temporadas 94-95 y 95-96**

Con ellos, el primer contraste de hipótesis: **“Ha aumentado el número de goles por partido”** frente a **“el número de goles por partido es el mismo”**:

$$H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 < \mu_2$$

Teniendo en cuenta que  $n$  es grande puede suponerse que  $\hat{s}_1 = \sigma_1$  y  $\hat{s}_2 = \sigma_2$ . El estadístico es, por tanto:

<sup>1</sup> ¿Sería este motivo de otro contraste?

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0; 1)$$

con:

$$\begin{aligned} \bar{x}_1 &= 2.54 & \hat{s}_1 &= 1.70 & n_1 &= 380 \\ \bar{x}_2 &= 2.70 & \hat{s}_2 &= 1.80 & n_2 &= 462 \end{aligned}$$

Sustituyendo:

$$T = \frac{(2.70 - 2.54) - 0}{\sqrt{\frac{1.70^2}{380} + \frac{1.80^2}{462}}} = 1.32 \sim N(0; 1)$$

cuyo  $p$  - valor es:

$$p - \text{valor} = p(z > T) = 0.0834$$

lo que lleva a rechazar la hipótesis alternativa, ya que  $\alpha = 0,05 < p$  - valor; luego:

**“La medida de dar tres puntos por victoria NO ha aumentado el número de goles por partido”**

Del mismo modo se puede proceder para el contraste acerca de la proporción de partidos que acaban en victoria.

El contraste:

**“Ha aumentado la proporción de partidos que acaban en victoria” frente a “la proporción es la misma”:**

$$H_0 : p_1 = p_2 \quad H_1 : p_1 < p_2$$

se resuelve con el estadístico:

$$T = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}} \sim N(0; 1)$$

que operando y hallando su  $p$  - valor:

$$T = 0.51 \Rightarrow p - \text{valor} = p(z > T) = 0.3085$$

Siendo  $\alpha < p$  - valor, se rechaza la hipótesis alternativa y se puede afirmar que:

**“Dar tres puntos por ganar no ha aumentado la proporción de victorias”**

Podrían plantearse dos preguntas:

1. Por qué es necesario realizar un contraste de hipótesis si se han considerado todos los partidos de ambas temporadas.
2. Cómo es posible que habiendo hecho los cálculos con todos los partidos y viendo que  $\bar{x}_2 > \bar{x}_1$  y  $\hat{p}_2 > \hat{p}_1$  haya salido que no pueden aceptarse un aumento en el número de goles ni de la proporción de victorias.

En ambos casos pueden mostrarse las tablas 4 y 5, que muestran la variabilidad de la media de goles por partido y de la proporción de victorias



Temporada	71	72	73	74	75	76	77	78	79	80	81	82	83	74	85	87	88	89	90	91	92	93	94
Nº Equipos	18	18	18	18	18	18	18	18	18	18	18	18	18	18	20	20	20	20	20	20	20	20	20
Media goles	2,21	2,14	2,32	2,43	2,50	2,71	2,75	2,70	2,51	2,71	2,79	2,55	2,56	2,16	2,61	2,39	2,28	2,42	2,16	2,40	2,51	2,60	2,54
% victorias	73,2	70,6	76,8	72,9	76,8	75,5	77,1	75,5	72,2	77,1	81,7	74,2	73,5	64,4	74,5	73,2	70,5	71,8	69,5	75,0	71,1	72,1	70,3

**Tabla 4.-** Número de goles y porcentaje de victorias en las temporadas en las que cada victoria daba 2 puntos. Se omite la 86/87 ya que se celebró con un sistema de *play-offs*

Temporada	95-96	96-97	97-98	98-99	99-00	00-01	01-02	02-03	03-04	04-05	05-06	06-07
Nº Equipos	22	22	20	20	20	20	20	20	20	20	20	20
Media goles	2,70	2,75	2,66	2,64	2,63	2,88	2,53	2,67	2,67	2,58	2,46	2,48
% victorias	71,9	74,2	71,1	74,5	70,5	73,9	73,4	72,4	75,0	73,7	72,4	74,2

**Tabla 5.-** Número de goles y porcentaje de victorias con el sistema de puntuación actual

### Conclusión

Se han planteado y resuelto tres ejercicios, en orden creciente de dificultad en los que se ha hecho uso de un tema cotidiano como el fútbol para acercar el estudio de las Matemáticas a los alumnos.

Se ha intentado que la parte “operativa” fuera secundaria en todo el proceso, para lo que se ha hecho uso de la calculadora **CASIO fx-9860G SD** y de la hoja de cálculo “**Calc**” de la *suite* OpenOffice.org. Con ello se pretende evitar que la realización de los cálculos distraiga la atención del alumno, buscando un mejor entendimiento de la estadística.