UNIVERSIDAD CENTROCCIDENTAL
“LISANDRO ALVARADO”

Decanato de Ciencias y Tecnología
Licenciatura en Ciencias Matemáticas

MODELOS ESTADÍSTICOS PARA PRONÓSTICOS DE PARTIDOS DEL FUTBOL VENEZOLANO

TRABAJO ESPECIAL DE GRADO PRESENTADO POR
Br: RONNY H. QUINTERO L.

COMO REQUISITO FINAL
PARA OBTENER EL TÍTULO DE LICENCIADO
EN CIENCIAS MATEMÁTICAS

ÁREA DE CONOCIMIENTO: OPTIMIZACIÓN-PROBABILIDAD Y ESTADÍSTICA

TUTOR: MSC. JHONNY ESCALONA

Barquisimeto -Venezuela

Enero 2015

Dedicado a
DIOS Y MI FAMILIA

Agradecimientos

A Dios mi padre celestial, mi guía por darme sabiduría y entendimiento, a mi sobrina Shofia por sacarme una sonrisa con solo pensar en ella, a Máximo amigo fiel.

A mis padres Maritza Leguisamo e Hipólito Quintero por darme la vida, jamás podré pagarle todo lo que me han brindado.

A mis hermanos Robert y Rocyeri, siempre me he sentido orgulloso de ustedes, gracias por todo, los amo.

A Claudia Rodríguez, una persona que supo traerme quietud en momentos difíciles y que hizo de mí una mejor persona.

A Erianny Mendoza, Francis Álvarez, María Esperanza Martínez, Dignora Pagés, Rafael Chavez y por último, pero no menos importante Rafael González compañero desde mis inicios en la carrera, a todos esos compañeros que de una forma u otra hicieron vida junto a mí en la U.C.L.A, a todos ustedes gracias.

A mi tutor Jhonny Escalona quien fue más que un profesor un amigo, un guía, mil gracias por todo sus consejos y por siempre brindarme su ayuda.

Al profesor Javier Hernández, por brindarme siempre una mano amiga.

A los profesores Miguel Vivas, Liliana Pérez, Eibar Hernández y Mario Rodríguez pilares importantes en mi formación, gracias por compartir sus conocimientos.

Al personal obrero y de biblioteca por su aporte, ayuda y dedicación en su trabajo.

RESUMEN

La estadística es una ciencia aplicada de utilidad en múltiples disciplinas. Este trabajo presenta una aplicación al Deporte, con el objetivo de modelar los resultados de los partidos de la Primera División del Futbol Venezolano, más específicamente el número de goles obtenidos por cada equipo en un partido, mediante el uso de la distribución de Poisson. En primer lugar se consideran los datos pertenecientes a las temporadas 2009-2010 hasta 2013-2014 y se muestra además que estos siguen una distribución de Poisson. Por otro lado después de presenta los modelos, se obtiene la estimación de los parámetros que nos permiten realizar simulaciones para obtener las predicciones de los partidos utilizando el paquete estadístico R.

Índice general

Introducción

El modelado estadístico de datos de futbol es un tema muy popular, y muchas investigaciones han sido desarrolladas es esta área, desde el punto de vista estadístico la tarea es estimulante, debido a que en su estudio se plantean algunos aspectos que llaman la atención, uno de estos esta relacionado con la forma de distribución asociada al número de goles; pero yendo mas allá, otro aspecto interesante que ha llevado a un auge importante en el desarrollo de modelos matemáticos para predecir resultados deportivos es el creciente mercado de apuestas deportivas, ya que no solo los corredores de apuestas están interesados en modelos que les ayuden a predecir los resultados, sino también, las empresas de apuestas, que tiene un mayor interés, el interés económico, ya que estas buscan estar un pie adelante sobre los corredores de apuesta, para así obtener mayores ganancia, conocer las predicciones les permite establecer el límite de ganancia para los apostadores según el juego y los posibles resultados a la mano.

En la gran mayoría de las ligas de futbol pertenecientes a las llamadas potencias del futbol, el desarrollo de este deporte va a la par con estudios estadísticos, actualmente en Venezuela no se cuenta con estudios en esta área, de acá surge nuestro interés para implementar modelos estadísticos utilizando datos de la primera división del futbol Venezolano, en primer lugar dándonos a la tarea de estudiar y analizar los modelos propuestos en [1], [3] y [4]. Todos estos modelos tienen la peculiaridad de que el número de goles en un partido de futbol se distribuye Poisson, aunque cada uno de los autores aporta herramientas que permiten mejorar los resultados en términos de predicción. Al identificar los modelos estadísticos de interés, se procedió a realizar una búsqueda de los datos que llevo a su consecuente estudio, en primer lugar, seleccionando los datos correspondientes a la temporada 2009-2010 hasta la temporada 2013-2014 del principal torneo de futbol venezolano. También, se realizo una serie de gráficas que nos permitió identicar visiblemente un efecto en los juegos en casa y más aún, se realizo una prueba de bondad de ajuste que nos mostró la forma en que se distribuyen los datos, todo estos resultados son presentados en el capítulo 2 de este trabajo.

Posteriormente en el capítulo 3, se introducen el modelo de Poisson enfoque Karlis-Ntzoufras, el Modelo de Regresión de Poisson y el modelo de Poisson enfoque Dixon-Coles. Se realiza el desarrollo de cada modelo explicándolo en detalle, seguidamente al terminar el estudio teórico de cada modelo se muestran las estimaciones de los parámetros, así como también se obtiene una serie de predicciones de los partidos, y simulaciones que permiten tener una idea de la clasificación de los equipos en torneos futuros, todas las simulaciones y predicciones se llevaron a cabo en el programa estadístico R. Los códigos, tanto de los modelos, como de la prueba de bondad de ajuste Chi-cuadrado (χ^2) que se realizo, así como también los datos, se presentan en un CD junto con este trabajo.

Finalmente, en un apartado de conclusiones se mencionan los hallazgos más importantes de cada modelo, posibles líneas de investigación en el análisis de los resultados de futbol y posibles modificaciones que se pueden hacer en estos modelos para estudios futuros.

Capítulo 1

Preliminares

Este capítulo tiene el propósito de introducir los conceptos teóricos y la terminología utilizada.

1.1. Valor Esperado y Varianza

El valor esperado se puede calcular como la suma de la probabilidad de todos los posibles resultados multiplicada por la ganancia o la pérdida correspondiente. También se suele llamar media, expectativa, esperanza matemática o primer momento. Esto implica que el valor esperado es el valor que en promedio se puede esperar de un escenario dado, si el escenario se repite un número de veces

Definición 1.1 (Valor Esperado). Sea X una variable aleatoria discreta que puede tomar diferentes valores $x_1,...,x_i,...$ con $p_i{=}\mathbb{P}(X{=}x_i)$. El valor esperado de X, representado por E(X) o μ_X, se define como

$$E(X) = \sum_i p_i x_i$$

Ejemplo 1.1 . Consideremos un partido de futbol entre dos equipos, donde la probabilidad de los tres resultados posibles (Victoria en casa, Empate, Victoria visitante) son (0.30, 0.31, 0.39) y sus correspondientes valores de las apuestas son (3, 3.1, 2.65). Los valores esperados de las tres apuestas posibles son:

1. E(Apuesta=Victoria casa)=$\sum_i p_i x_i = 0{,}30 \times 3 + 0{,}31 \times 0 + 0{,}39 \times 0 = 0{,}9$

2. E(Apuesta=Empate)=$\sum_i p_i x_i = 0{,}30 \times 0 + 0{,}31 \times 3{,}1 + 0{,}39 \times 0 = 0{,}961$

3. E(Apuesta=Victoria visita)=$\sum_i p_i x_i = 0{,}30 \times 0 + 0{,}31 \times 0 + 0{,}39 \times 2{,}65 = 1{,}034$

De las tres posibles apuestas sobre el resultado, con la evaluación de la probabilidad mencionada, la única apuesta con un valor esperado favorable es la apuesta por la “victoria visitante”.

Si consideramos una variable aleatoria X, Intuitivamente la varianza de X, es una medida de lo dispersa que es la distribución de X, o cual aleatoria es X o cuanto varia X.

Definición 1.2 (Varianza). La *varianza* de una variable aleatoria X esta definida por

$$Var(X) = \sigma_X^2 = E[(X - \mu_X)]^2$$

donde $\mu_X = E(X)$ es la media de X.

De esto, se puede observar que la varianza es siempre positiva, una pequeña varianza indica que los datos tienden a ser muy cercanos al valor esperado, mientras que una alta varianza indica que los datos están muy extendidos alrededor de la media.Una medida equivalente es la raíz cuadrada de la varianza, llamada desviación estándar, La desviación estándar se define como $\sqrt{\sigma^2}$. Por lo tanto, la desviación estándar a menudo se denota por σ.

1.2. Función de Probabilidad Conjunta

En lo que sigue estaremos trabajando con el supuesto de que se tiene un *espacio de probabilidad*, el cual es denotado usualmente por $(\Omega, \mathcal{F}, \mathbb{P})$, donde Ω es un conjunto arbitrario, $\mathcal{F}$ es una σ-álgebra de subconjuntos de Ω, y $\mathbb{P}$ es una medida de probabilidad definida sobre $\mathcal{F}$.

Definición 1.3 (Vector Aleatorio). Se dice que $X = (X_1, ..., X_n)$ es un vector aleatorio, si cada coordenada X_i con $i \in \{1, ..., n\}$ es una variable aleatoria.

Definición 1.4 (Vector Discreto). Se dice que $X = (X_1, ..., X_n)$ es vector discreto, si cada coordenada X_i con $i \in \{1, ..., n\}$ es una variable aleatoria discreta.

Como en el caso unidimensional, los vectores aleatorios tienen asociada una función llamada de probabilidad o de densidad, la cual se define a continuación para el caso de un vector aleatorio discreto bidimensional.

Definición 1.5 (Función de Probabilidad Conjunta). La función de probabilidad de un vector discreto (X,Y) es la función $f : \mathbb{R}^2 \longrightarrow [0, 1]$ dada por:

$$f(x,y) = \mathbb{P}(X = x, Y = y).$$

A esta función también se le llama **función de probabilidad conjunta o función de densidad** de las variables X y Y.

Además, es claro ver que por ser $\mathbb{P}$ una medida de probabilidad, la función de probabilidad de un vector discreto cumple las Siguientes propiedades.

1. $f(x,y) \geq 0, \quad \forall\, (x,y) \in (X,Y)$
2. $\sum_x \sum_y f(x,y) = 1$

Recíprocamente, toda función no negativa : $\mathbb{R}^2 \longrightarrow [0,1]$ que sea estrictamente positiva únicamente en un subconjunto discreto de $\mathbb{R}^2$ y que sume uno, se llama **función de probabilidad conjunta.**

1.3. Independencia de variables aleatorias

La independencia es un tema central en la teoría de la probabilidad y en la estadística, y su negación, la dependencia es estudiada en diferentes áreas, intuitivamente podemos decir que n-variables aleatorias son independientes si los valores que toma cada una de ellas no afectan los valores de cada una de las otras ni a sus probabilidades.

Definición 1.6 (Independencia de variables aleatorias). Se dice que las variables $X_1,...,X_n$ son independientes, si para cualesquiera conjuntos de Borel $A_1,...,A_n$ de $\mathbb{R}$, se cumple que

$$\mathbb{P}(X_1 \in A_1,...,X_n \in A_n) = \mathbb{P}(X_1 \in A_1)...\mathbb{P}(X_n \in A_n).$$

Más aún, una colección infinita de variables aleatorias es independiente si cualquier subconjunto finito de ella lo es.

Definición 1.7 (Independencia de dos variables aleatorias). Se dice que las variables X y Y son independientes, si para cada par de conjuntos de Borel A,B de $\mathbb{R}$, se cumple que

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(X \in B).$$

El concepto de independencia de variables aleatorias es una extensión de la misma propiedad para eventos. Cuando la función de densidad conjunta existe, la condición de independencia de X y Y es equivalente a solicitar que para cualesquiera números reales x y, se cumpla la identidad

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) \tag{1.1}$$

En el caso discreto, la afirmación anterior es completamente correcta. Para el caso continuo hay una observación técnica que es necesario mencionar. Como en este caso las funciones de densidad pueden ser modificadas sin que cambie la función de distribución asociada, la igualdad 1.1 puede no cumplirse para cada $(x,y) \in \mathbb{R}^2$, entonces se permite que la igualdad no se cumpla en un conjunto de medida de Lebesgue cero, por ejemplo, un conjunto numerable de pares (x,y) en $\mathbb{R}^2$, entonces habrá independencia en el caso continuo si se cumple 1.1, salvo conjuntos de medida de Lebesgue cero.

Definición 1.8 (Conteo o Recuento). Se define como el número de sucesos o eventos que ocurren en una misma unidad de observación en un intervalo espacial o temporal definido, también se considera como una realización no negativa de valor entero de una variable aleatoria.

1.4. Distribución de Poisson

La distribución de Poisson es una distribución de probabilidad discreta, muy conocida en la teoría de probabilidad y la estadística. Expresa la probabilidad de un número de eventos que ocurren dentro de un período fijo de tiempo, fue descubierta por Simeón Deniss Poisson en 1873 como límite de la distribución Binomial.

Definición 1.9 (Distribución de Poisson). Sea x el número de veces que ocurre el evento, y $\lambda \in \mathbb{N}$ la media, entonces la distribución Poisson indica que la probabilidad para x es:

$$f(x|\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & \text{para } x \in \mathbb{N},\ x \geq 0 \\ 0, & \text{en otro caso} \end{cases} \tag{1.2}$$

observe que $f(x|\lambda) \geq 0$, gracias a las restricciones $e^{-\lambda} > 0$, $\lambda^x \geq 0$ y $x! \geq 1$.

Además al ser una distribución de probabilidad discreta debe cumplir,

$$\sum_{x=0}^{\infty} f(x|\lambda) = 1.$$

En efecto,

en primer lugar recordemos que para $\lambda \in \mathbb{R}$ se tiene que

$$e^{\lambda} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}.$$

Por otro lado,

$$\sum_{x=0}^{\infty} f(x|\lambda) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = e^{-\lambda+\lambda} = 1.$$

Ahora bien, demostremos que λ dada en la definición 1.5 es la media de la distribución de Poisson, para esto debemos probar que,

$$E(X) = \sum_{x=0}^{\infty} f(x|\lambda)x = \lambda,$$

en efecto,

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{\infty} f(x|\lambda)x \\
&= \sum_{x=1}^{\infty} f(x|\lambda)x \\
&= \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x}}{x!}x \\
&= \lambda \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} \\
&= \lambda \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^{y}}{y!} = \lambda.
\end{aligned}
$$

Por lo tanto,

$$E(X) = \sum_{x=0}^{\infty} f(x|\lambda)x = \lambda.$$

Además, si λ crece, la masa de la distribución de Poisson se desplaza hacia la derecha, y se aproxima a una distribución Normal por el Teorema del Limite Central.

Por otro lado, $Var(X) = \lambda$, en primer lugar observe que

$$E(X(X-1)) = \sum_{x=0}^{\infty} x(x-1)\frac{e^{-\lambda}\lambda^{x}}{x!} = \lambda^2 e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^{x-2}}{(x-2)!} = \lambda^2 e^{-\lambda}e^{\lambda} = \lambda^2$$

$$Var(X) = E(X^2) - [E(X)]^2 = E(X(X-1)+X) - [E(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

1.4.1. Aplicaciones de la variable de Poisson

1. Conteos en el Tiempo
   a) Número de anotaciones durante un partido de futbol.
   b) Número de accidentes de tráfico en un tramo de cierta carretera en un mes.
   c) Número de mutaciones en una población de animales durante 5 años.
2. Conteos en el Espacio
   a) Número de accidentes de tráfico que se originan en el cruce de 2 carreteras.
   b) Número de árboles infectados por hectárea en un bosque.

1.5. Modelo de Regresión de Poisson

El modelo de regresión de Poisson es un tipo de Modelo Lineal Generalizado, considerado como el modelo de referencia en estudios de variables de recuento, además es un modelo que resulta especialmente adecuado para modelar valores enteros no negativos, especialmente cuando la frecuencia de ocurrencia es baja. Los Modelos Lineales Generalizados permiten incluir distintas relaciones entre las medias condicionales de las variables respuesta y las explicativas. El modelo de regresión de Poisson se utiliza para datos de conteo. Este modelo es adecuado cuando la varianza muestral es igual a la media. En el modelo de regresión de Poisson la media (λ) se explica en términos de las variables explicativas mediante el uso de un enlace.

El modelo de regresión de Poisson surge cuando la variable respuesta es una cantidad discreta que se puede modelar con una Poisson y se quiere estudiar si ciertas variables explicativas afectan a la variable respuesta y cómo lo hacen. Este tipo de variable respuesta suele presentarse en el recuento de sucesos o hechos (por ejemplo, el número de goles marcados en un partido). Por tanto, se considera una variable respuesta Y que toma los valores en $\{0,1,2,3,...\}$, y se va estudiar su relación con otras variables explicativas X mediante un análisis de regresión. Se pretende construir un modelo para $\lambda(x) = E(Y \mid X = x)$, es decir, para la media de Y condicionada a cada valor de la variable explicativa.

Como Y nunca toma valores negativos, no procede utilizar un modelo lineal directo, y por tanto se necesita una función de enlace previa a cualquier modelo lineal. Además, la función de regresión está en el intervalo $(0,\infty)$, por lo que el logaritmo parece la función de enlace más adecuada. Esto se expresaría de la siguiente forma: $g(\lambda(x,\beta)) = \widehat{x}\beta$ donde lo más habitual como ya se dijo es tomar $g(r) = \ln(r)$, $r \in (0,\infty)$, Mediante $\widehat{x}\beta$ se representa el producto escalar del vector de variables explicativas $\widehat{x}$, por el vector de parámetros β. Para incluir un intercepto se considera una primera variable explicativa igual a 1. La función de regresión del modelo de Poisson se expresaría por: $\lambda(x,\beta) = e^{\widehat{x}\beta}$.

Los parámetros de este modelo, que se puede denominar modelo log-lineal, se interpretan de la siguiente forma: la exponencial del intercepto es el valor esperado de la respuesta en la categoría de referencia o cuando las variables explicativas numéricas valen cero, y las exponenciales de los coeficientes de cada variable representan tasas de incremento de la respuesta esperada al aumentar una unidad la variable si es numérica, o al pasar a la categoría correspondiente si es cualitativa. Es decir, el modelo supone efectos multiplicativos: si la componente explicativa unidimensional X_j aumenta n unidades, la media para la variable de Poisson se multiplica por la potencia n-ésima de e^{β_j}.

Si se tiene una muestra aleatoria simple $(X_1,Y_1),...,(X_n,Y_n)$ de (X,Y), entonces, Y_i se distribuye $Poisson(\lambda(X_i,\beta))$ siendo $\lambda(X_i,\beta) = e^{\widehat{x}\beta}$. Para estimar los parámetros del modelo se utiliza la función de máxima verosimilitud que veremos seguidamente. Además, también los estimadores de máxima verosimilitud son asintóticamente normales, centrados y su matriz de varianzas-covarianzas es la inversa de la matriz de información (la matriz hessiana cambiada de signo), lo cual permite hacer inferencias sobre los parámetros del modelo.

1.6. Estimación de Máxima Verosimilitud

La función de verosimilitud es uno de los conceptos más básicos en inferencia estadística y las inferencias por verosimilitud se basan solamente en los datos X y el modelo $\{\mathbb{P}_\theta : \theta \in \Theta\}$, es decir, un conjunto de posibles medidas de probabilidad para el sistema que es objeto de estudio. Con estos elementos obtenemos la esencia fundamental de la inferencia basada en la verosimilitud, denominada la función de verosimilitud, por lo que la estimación de máxima verosimilitud es el método de estimación general más popular para la obtención de la estimación de una distribución de una muestra finita.

Definición 1.10 . Sea $X = (x_1, ..., x_n)$ una muestra independiente idénticamente distribuida de una distribución con un parámetro θ pertenecientes a Θ. El estimador de máxima verosimilitud (MLE) es el $\theta \in \Theta$ que maximiza la función de verosimilitud dada por:

$$L : \Theta \longrightarrow \mathbb{R}$$
$$L(\theta) = L(\theta|x_1, ..., x_n) = f_\theta(x_1, ..., x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

donde f_θ es la función de probabilidad de X. El estimador de máxima verosimilitud es denotado por $\hat{\theta}$ o $\hat{\theta}_n$ si queremos hacer hincapié en el tamaño de la muestra. Primero hemos quitado la dependencia en L de $x_1, ..., x_n$ para hacer notar que estamos tratando a la probabilidad como una función de θ. Tengamos en cuenta que x_i y θ pueden ser escalares o vectores (no necesariamente de la misma dimensión).

Por otro lado, es importante considerar las siguientes propiedades con relación a la función de verosimilitud y el estimador de máxima verosimilitud:

1. Sea g una función creciente estrictamente monótona tal que $g \circ L$ este bien definida, entonces g preserva el orden, en el sentido de que el máximo de L coincide con el máximo de $g \circ L$. Como consecuencia de esto podemos encontrar el estimador de máxima verosimilitud de $g \circ L$ en lugar del estimador de máxima verosimilitud de L, de aquí una propiedad muy útil ya que podemos trasformar la multiplicación de probabilidades en una suma de probabilidades (la suma es más fácil de diferenciar que el producto).

2. Cualquier término aditivo o multiplicativo en $g \circ L$ que no sea una función de θ puede ser ignorado ya que al suprimirlo no cambiara el máximo.

3. Si $L(\theta)$ es diferenciable en θ, podemos tratar de encontrar el estimador de máxima verosimilitud resolviendo la ecuación

$$\begin{cases} \frac{d \log L}{d\theta} = 0 & \text{si} \quad \theta \in \mathbb{R} \\ \nabla \log L = 0 \equiv \frac{\partial \log L}{\partial \theta_j} = 0, con\, j = 1, ..., n & \text{si} \quad \theta \in \mathbb{R}^n. \end{cases} \tag{1.3}$$

4. Si el método anterior no funciona, es decir, no podemos resolver $\nabla \log L = 0$, podemos encontrar el estimador de máxima verosimilitud iterativamente siguiendo la dirección del gradiente, primero inicializando θ al azar e iterar $\theta \leftarrow \theta + \alpha \nabla \log L$, donde α

es un tamaño de paso suficientemente pequeño hasta la convergencia, por ejemplo $\| \nabla \log L \| \leq \varepsilon$ o $\| \theta^{t+1} - \theta^t \| < \varepsilon$.

5. Los estimadores de máxima verosimilitud tienen ciertas propiedades en general que a continuación enunciamos:

    a) Son consistentes.

    b) Es eficiente e incluso eficiente de segundo orden tras corregir el sesgo.

    c) Es $\sqrt{n}$-consistente y asintóticamente eficiente.

    d) En caso de existir un estimador insesgado y eficiente (varianza igual a la cota de Cramér-Rao), dicho estimador es el máximo verosímil.

    e) Son asintóticamente normales.

1.7. Método del gradiente de descenso

Método de descenso de gradiente es una manera de encontrar un mínimo local de una función, la forma en que funciona es que empezamos con una estimación inicial de la solución y tomamos el gradiente de la función en ese punto, luego consideramos la solución en la dirección negativa del gradiente y repetimos el proceso. El algoritmo eventualmente convergen donde el gradiente es cero, que corresponde a un mínimo local. Su hermano, el **gradiente de ascenso**, encuentra el máximo local más cercano a la solución actual, intensificando hacia la dirección positiva del gradiente. Ambos son algoritmos de primer orden porque se toman sólo la primera derivada de la función.

1.7.1. Descripción del Método

Sea $f : A \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ una función de clase $C^1(A)$, donde A es abierto. Supongamos que estamos buscando un mínimo de f, supuesto algún valor inicial mínimo $x_0 \in A$ para $x \in A$, para averiguar cuál es la mejor dirección para minimizar f, se toma $-\nabla f$, pues intuitivamente, el gradiente dará la pendiente de la curva en ese x y su dirección apuntará a un máximo de la función. Así que cambiamos x la dirección para encontrar el mínimo de la función y además consideremos la secuencia $x_0, x_1, x_2, \ldots$ de tal forma que

$$x_{k+1} = x_k - \gamma_k \nabla f(x_k), \quad k \geq 0$$

donde $\gamma > 0$ es un número suficientemente pequeño que obliga al algoritmo a hacer pequeños saltos. Eso mantiene el algoritmo estable y su valor óptimo depende de la función. Dadas unas condiciones estables y una determinada elección de γ, Se garantiza que $f x_{k+1} \leq f(x_k)$ y así la secuencia $\{x_k\}_{k\geq 0}$ debe converger al mínimo local deseado.

Un tamaño de paso γ en falso puede causar divergencia, por lo que una selección cuidadosa del tamaño de paso es importante. Demasiado grande será divergir, demasiado pequeño tomará un largo tiempo para converger. Una opción es elegir un tamaño de paso fijo que asegure la convergencia donde quiera que comience el descenso de gradiente, otra opción es elegir un tamaño de paso diferente en cada iteración, es decir, un tamaño de paso de adaptación.

Los pasos asociados a la utilización de este método consisten en:

1. Considere un punto inicial $x = x_0$. Hacer $k = 0$.
2. Escoger una dirección de descenso $d_k = -\nabla f(x_k)$.
3. En caso de no tomar un tamaño de paso γ fijo, realizar una búsqueda lineal que seleccione un paso γ_k tal que $g_k(\gamma_k) = f(x_k + \gamma_k d_k) < f(x_k) = g_k(0)$.
4. Hacer $x_{k+1} = x_k + \gamma_k d_k$.
5. Hacer una prueba de convergencia, por ejemplo $\| \nabla f(X_k) \| < \varepsilon$. Si converge se detiene el método. En caso contrario hacer $k = k+1$ y volver al paso 2.

1.8. Contrastes o pruebas de bondad de ajuste

Los contrastes o pruebas de ajuste tienen como objetivo decidir si puede aceptarse la hipótesis de que una muestra dada procede de una población con una distribución de probabilidad totalmente especificada en la hipótesis nula. Estos contrastes se basan en la comparación de las frecuencias observadas en la muestra con aquellas que habría de esperarse si la hipótesis nula fuera cierta. La hipótesis nula se rechaza si existe una diferencia significativa entre las frecuencias observadas y las esperadas. En este tipo de contrastes la distribución de probabilidad del estadístico de prueba es independiente de la postura en la hipótesis nula y depende solo del tamaño de la muestra o del número de clases en que se agrupa la variable.

1.8.1. Prueba de bondad de ajuste Chi-cuadrado (χ^2)

Esta prueba puede realizarse incluso con datos medibles en una escala nominal. La hipótesis nula de la prueba $Chi-cuadrado$ postula una distribución de probabilidad totalmente especificada como el modelo matemático de la población que ha generado la muestra. Para explicar mejor esta prueba supongamos que al realizar un experimento aleatorio n veces, se presentan los resultados $R_1, ..., R_k$ con frecuencias observadas $O_1, ..., O_k$ y que de acuerdo con las leyes de probabilidad se espera que estos resultados se presentes con frecuencias $E_1, ..., Ek$.

Una medida de las diferencias entre las frecuencias observadas y las esperadas está dada por el estadístico χ^2 definido:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

donde, $\sum_{i=1}^{k} O_i = \sum_{i=1}^{k} E_i$ y k es el número de categorías o tamaño de la muestra.

Si k=2, se debe usar la fórmula corregida de Yates:

$$\chi^2 = \sum_{i=1}^{k} \frac{(\mid O_i - E_i \mid -0,5)^2}{E_i}$$

Este estadístico tiene una distribución $Chi-cuadrado$ con $k-r$ grados de libertad si k es suficientemente grande y r es el número de restricciones, además, $r \geq 1$, ya que $\sum_{i=1}^{k} O_i = \sum_{i=1}^{k} E_i$ es siempre una restricción y cada parámetro que se estima con la información de la muestra es otra restricción más. Si las frecuencias observadas coinciden o se aproximan mucho a las esperadas, el valor del estadístico χ^2 tiende a cero. Por el contrario, si las frecuencias observadas difieren significativamente de las esperadas, el valor del estadístico χ^2 será positivo y tan grande cuando mayores sean las diferencias entre las frecuencias. Bajo estas condiciones se tienen que la región de rechazo es solo la región derecha o también llamada cola derecha. En resumen las hipótesis a considerar son las siguientes:

- H_0 : Los datos provienen de una muestra al azar de una población distribuida de acuerdo al modelo teórico.
- H_A : Los datos no provienen de una muestra al azar de una población distribuida de acuerdo al modelo teórico.

Capítulo 2

Datos Disponibles

A continuación se presentara los datos disponibles de la Primera División del Futbol Venezolano, mediante un estudio de las estadísticas en cada temporada a considerar, obtenidos consultando diferentes fuentes para lograr una data final lo suficientemente confiable.

2.1. Datos del Futbol Venezolano

En este trabajo se utilizaran específicamente datos de la Primera División del Futbol Venezolano, para conseguirlos se ha consultado la página web de la Federación Venezolana de Futbol (F.V.F), la página web de ESPN Deportes y la página web http://www.soccerway.com, esta última página fue fundada en 1994 y se encarga de proporciona resultados en directo, próximos partidos, estadísticas y noticias de más de 824 competiciones deportivas de todo el mundo, estas tres fuentes nos permitió comparar las datas y obtener datos de calidad. La Federación Venezolana de Futbol, es el organismo rector del futbol en Venezuela, está a cargo de la selección de futbol de Venezuela y todas las categorías inferiores, tanto en la rama masculina como en la femenina y desde 1995 organiza los torneos nacionales del futbol profesional y aficionado. En su página web http://www.federacionvenezolanadefutbol.org se pueden encontrar las estadísticas y las fichas de los partidos celebrados en nuestro país desde el año 2012. ESPN Deportes, es un canal de televisión de Estados Unidos, propiedad de la cadena ESPN, la cual opera y produce canales de televisión por cable, satélite y radio, sitios web, revistas y libros relacionados con el deporte, en su página web http://www.espndeportes.com también se pueden encontrar las estadísticas y las fichas de los partidos celebrados en nuestro país desde el año 2005. Para la realización de este trabajo se tendrán en cuenta los resultados desde la temporada 2009-2010 hasta la temporada 2013-2014.

Este Campeonato consta de dos Torneos: Apertura y Clausura y una Fase Final. A partir de 2011, los Torneos de Apertura y Clausura de Primera División se juegan en un sólo grupo de 18 equipos, todos contra todos a una vuelta cada uno, con tabla de clasificación independiente. En el Torneo Apertura se disputan 17 jornadas en partidos considerados de ida y en el Torneo Clausura se disputan 17 jornadas en partidos considerados de vuelta. Si en el Torneo Apertura y Torneo Clausura existiese el mismo ganador, se le declarará automáticamente campeón de la Primera División del Futbol Venezolano y el subcampeón resultaría quién consiga la segunda posición de una Tabla Acumulada elaborada con la suma de los puntos obtenidos en ambos torneos. El Reglamento General de Competiciones de la F.V.F establece en su artículo 15 que todos los Campeonatos, Torneos y competencias se jugarán, asignándole a cada encuentro tres (3) puntos al ganador y uno (1) a cada uno en caso de empate, Además según el artículo 16, el orden de los calendarios, en cualquier Torneo, Campeonato o Competición, se determinará por sorteo, adjudicando un número a cada equipo y procurando, en la medida de lo posible, evitar coincidencias entre equipos de una misma localidad.

2.1.1. Temporada 2009-2010

Resultados de los partidos

En la temporada 2009-2010 jugaron 18 equipos en la Primera División la temporada comenzó el 9 de Agosto del 2009 y finalizó el 30 de Mayo del 2010. En total, se celebraron 306 partidos, a continuación se muestran los equipos que participaron en esta temporada con sus respectivos resultados.

	Casa →	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Caracas		1-0	3-1	1-1	1-1	3-1	4-2	1-6	2-0	0-1	0-3	1-1	0-0	1-2	0-1	1-1	0-1	1-3
2	Dptvo. Táchira	1-0		0-3	2-0	0-0	1-1	3-2	2-2	2-1	0-0	1-2	0-1	0-0	0-0	1-2	0-2	0-2	1-3
3	Dptvo. Petare	0-1	1-0		2-2	1-0	0-2	1-1	1-3	0-1	0-2	2-0	5-3	0-1	3-2	0-1	0-3	0-1	2-2
4	CD Lara	2-0	0-0	0-1		1-3	1-1	2-1	0-2	1-0	2-0	0-1	1-0	2-2	1-1	1-1	0-4	0-0	1-3
5	Trujillanos	3-1	0-0	2-1	2-2		1-2	3-1	2-1	0-2	0-1	0-1	1-0	1-1	3-2	2-0	1-1	1-1	2-1
6	Zamora	4-0	2-0	2-1	1-1	0-0		1-1	2-0	0-2	0-0	0-1	0-2	0-0	6-0	2-1	1-2	1-2	2-0
7	Zulia	4-0	2-0	2-0	2-1	2-0	2-1		3-1	4-3	2-4	2-0	2-2	2-1	2-2	1-5	3-4	0-3	0-1
8	Atl. Vigía	6-0	2-1	2-0	1-0	3-1	0-1	0-1		2-1	2-1	0-1	2-1	3-1	2-0	1-0	1-2	1-0	2-3
9	Dptvo. Anzoátegui	1-1	2-0	2-0	2-1	2-1	0-1	0-1	2-0		0-2	3-2	1-1	0-1	0-0	1-0	0-2	3-1	1-0
10	Aragua	4-0	2-1	1-1	1-1	3-1	2-3	1-0	4-2	1-1		2-3	2-2	1-1	0-2	1-0	2-0	1-0	3-1
11	Dptvo. La Guaira	3-2	0-0	2-0	1-0	2-1	3-2	0-1	3-1	2-3	1-0		5-0	3-3	0-0	0-0	2-0	1-3	1-1
12	Monagas SC	3-1	3-1	7-1	2-2	3-0	4-3	1-0	2-2	2-1	1-0	2-1		3-2	1-2	6-2	0-2	2-1	1-2
13	Carabobo	2-1	2-0	4-0	3-0	7-2	1-0	1-0	4-2	1-0	0-1	0-0	0-0		2-1	2-0	1-1	2-0	2-0
14	Est. de Mérida	1-0	3-0	0-0	1-1	1-0	2-0	2-1	2-2	0-0	1-1	2-2	0-1	0-0		1-1	1-1	2-0	0-0
15	A.C. Mineros	2-1	4-1	1-1	2-0	1-0	1-1	0-0	1-1	1-1	1-0	1-1	1-3	3-0	3-1		3-1	1-1	2-2
16	Llaneros	1-1	3-1	2-1	2-1	2-1	4-3	2-1	3-1	4-1	1-0	2-1	3-0	2-1	2-2	3-2		3-1	1-2
17	Yaracuyanos	2-0	4-3	4-1	1-1	5-1	3-2	2-1	2-2	2-2	2-0	3-0	2-1	2-2	1-2	1-0	2-1		2-3
18	Centro Italo	2-1	1-0	3-1	2-1	0-1	4-1	2-1	4-2	1-0	1-0	2-0	2-1	2-2	1-0	1-1	0-0	0-0	

Figura 2.1: Resultados Temporada 2009/2010

De la figura 2.1 podemos concluir que de los 306 partidos, 159 terminaron con un resultado de victoria en casa (color azul), 78 resultaron con un resultado de empate (color verde) y 69 victorias de visitante (color rojo).

2.1.2. Clasificación

Al finalizar la temporada, el Deportivo Táchira fue el equipo que acumulo el mayor número de puntos en la clasificación general, con 72 puntos, 45 de ellos en partidos jugados en casa, el siguiente equipo fue el Caracas F.C, con un total de 70 puntos, de los cuales 44 fueron obtenidos cuando jugaba de local, el tercer equipo en la clasificación fue el Deportivo Petare con un total de 69 puntos, de los cuales 39 de ellos fueron obtenidos jugando en casa. En la siguiente figura se muestran una tabla con los puntos obtenidos por cada uno de los 18 equipos enfrentados en la Primera División en esta temporada, además del total de puntos, se muestra el total de los partidos ganados, empatados y perdidos, así como también, los goles a favor y en contra. En la clasificación de la figura 2.2 los equipos aparecen ordenados según el puesto alcanzado en la clasificación general con respecto al total de puntos conseguidos por cada uno de los equipos.

Equipo	J	G	E	P	GF	GC	Puntos
Dptvo. Táchira	34	21	9	4	53	22	72
Caracas	34	21	7	6	66	30	70
Dptvo. Petare	34	21	6	7	62	31	69
CD Lara	34	15	14	5	47	30	59
Trujillanos	34	16	8	10	52	36	56
Zulia	34	16	5	13	49	50	53
Zamora	34	14	8	12	44	48	50
Dptvo. Anzoátegui	34	14	7	13	38	39	49
Atl. Vigía	34	14	5	15	50	60	47
A.C. Mineros	34	10	12	12	39	46	42
Aragua	34	11	8	15	31	44	41
Est. de Mérida	34	8	16	10	38	38	40
Monagas SC	34	11	7	16	51	62	40
Dptvo. La Guaira	34	10	8	16	39	48	38
Yaracuyanos	34	9	8	17	39	57	35
Carabobo	34	6	13	15	32	52	31
Centro Italo	34	6	8	20	36	55	26
Llaneros	34	5	7	22	38	67	22

Leyenda
J: Partidos jugados.
G: Partidos ganados.
E: Partidos empatados.
P: Partidos perdidos.
GF: Goles a favor.
GC: Goles en contra.

Figura 2.2: Clasificación Temporada 2009/2010

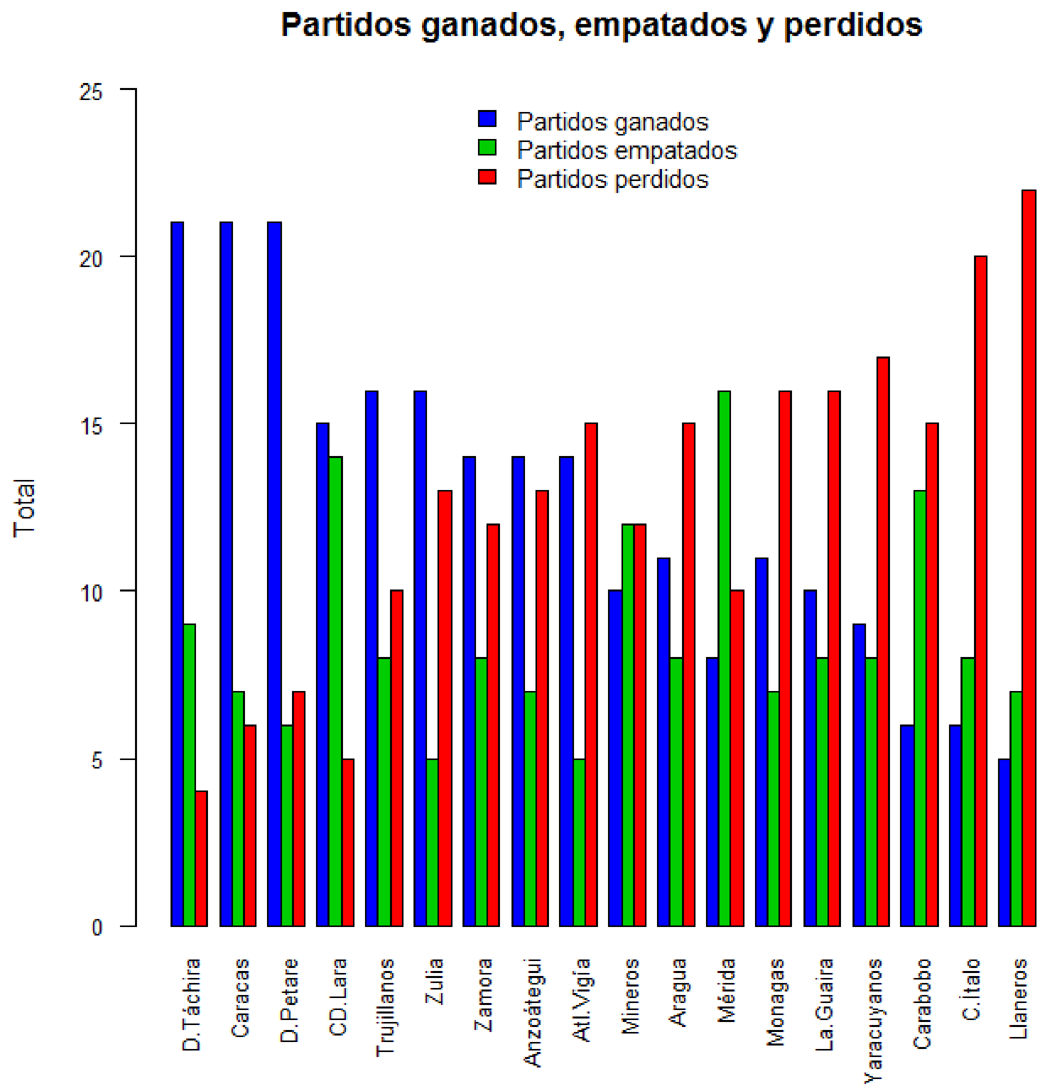

Figura 2.3: Gráfico partidos ganados, empatados y perdidos Temporada 2009/2010

En vista a que los equipos están ordenados de manera descendente según su puntuación obtenida, como es de esperar se observa que los mejores clasificados presentan más partidos ganados y ese número va descendiendo conforme a la posición en la tabla general.

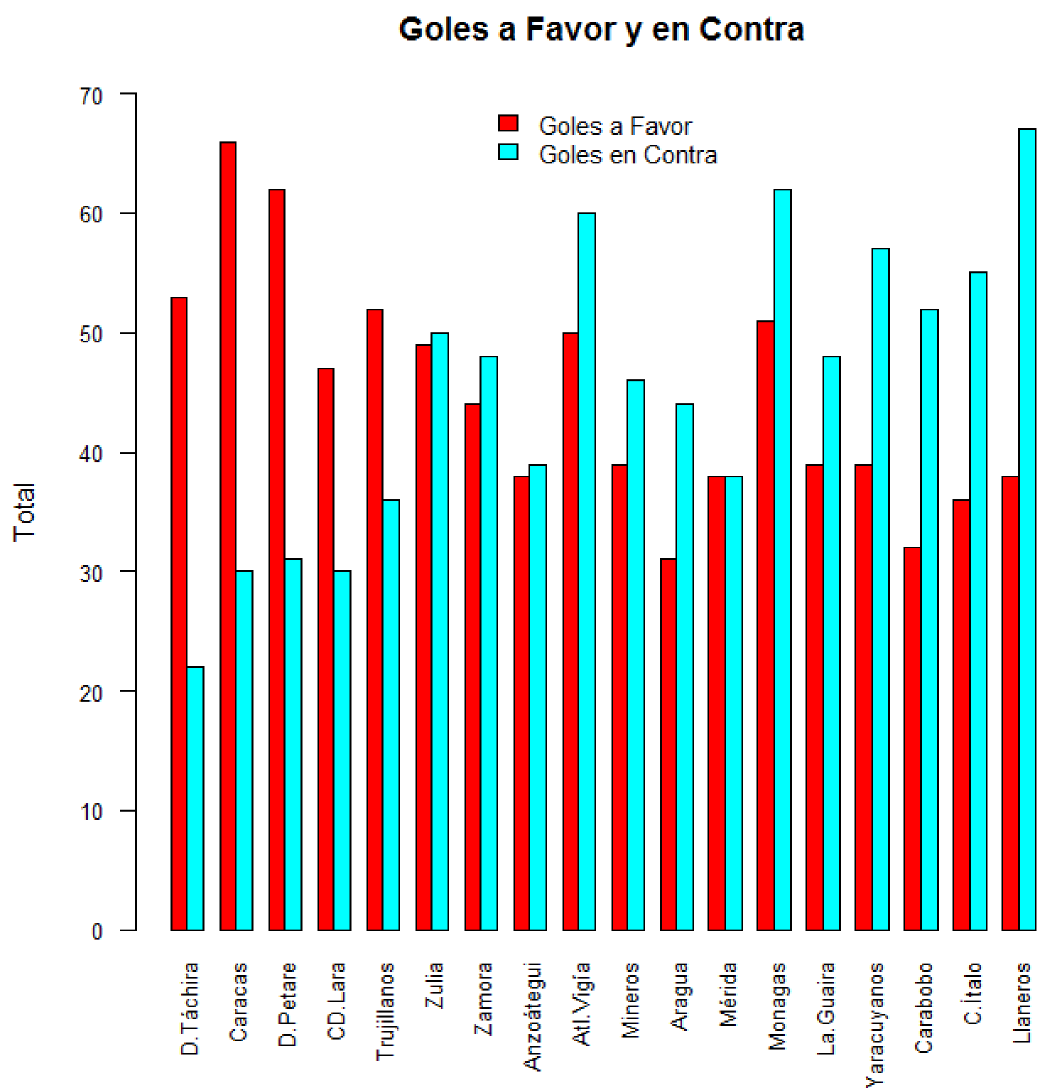

Figura 2.4: Gráfico goles a favor y en contra Temporada 2009/2010

Una tendencia similar se observa en relación a los goles a favor, (ver figura 2.4), aunque el Deportivo Táchira que clasifico de primero en la tabla acumulada, presenta un menor número de goles marcados que los equipos que clasificaron en segundo y tercer lugar. Cuando observamos los goles en contra notamos que el Deportivo Táchira presenta el menor número de goles encajados de todos los equipos, esto nos llevaría a pensar que clasificar de primero no solo depende de las goles marcados, que a primera impresión nos dan una idea del ataque, sino también, de los goles encajados los cuales intuitivamente nos dirían lo buena que es la defensa.

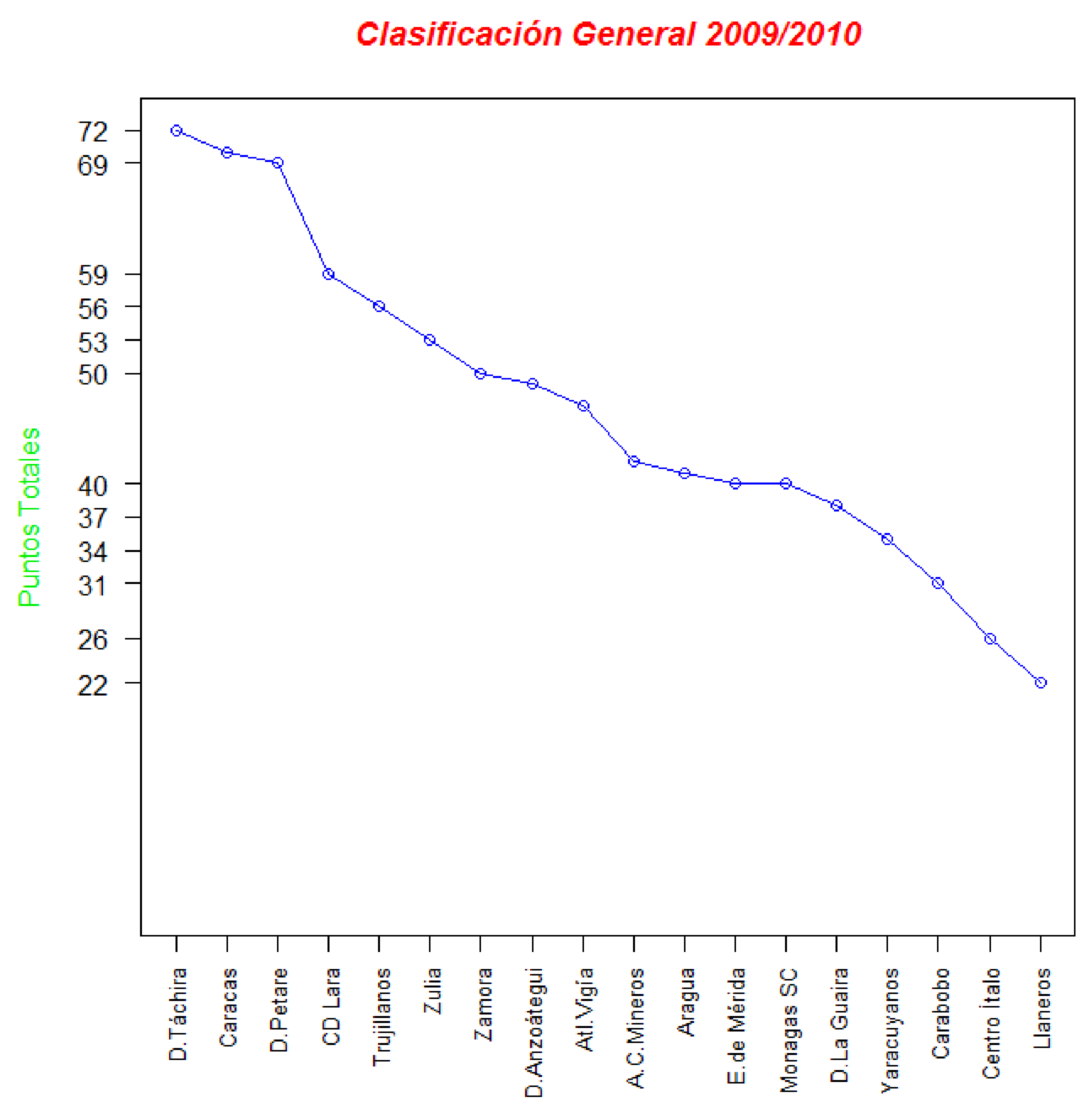

Figura 2.5: Clasificación general según la puntuación Temporada 2009/2010

En la figura se muestra la clasificación general de la temporada 2009/2010 teniendo como base el total de puntos, en este gráfico se observa que los equipos en los tres primeros puestos, a saber Deportivo Táchira, Caracas F.C y Deportivo Petare respectivamente que marcan una diferencia significativa en puntos acumulados, con respectos a un segundo grupo de equipos que se encuentra posteriormente, seguidamente se observa un tercer grupo de equipos cuya diferencia en puntuación se ve estable y por último se encuentra un grupo de equipos en la zona baja de la tabla.

2.1.3. Temporada 2010-2011

Resultados de los partidos

En la temporada 2010-2011 jugaron 18 equipos en la Primera División la temporada comenzó el 8 de Agosto del 2010 y finalizó el 29 de Mayo del 2011. En total, se celebraron

305 partidos, debido a que la Federación Venezolana de Futbol decidió suspender el partido de vuelta entre Caroní y Zulia F.C, ya que el equipo Caroní se presentó en la anterior jornada con un equipo juvenil. A continuación se muestran los equipos que participaron en esta temporada con sus respectivos resultados.

	Casa ⇒	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Caracas		1-2	2-3	0-3	0-1	3-0	2-0	0-1	1-0	0-1	0-1	2-2	1-3	1-0	0-1	1-2	2-3	0-1
2	Dptvo. La Guaira	3-0		0-0	1-0	2-0	2-1	0-0	0-1	1-2	1-3	1-0	0-0	1-3	1-0	2-5	1-0	0-1	0-1
3	Dptvo. Petare	2-2	0-0		0-1	1-2	2-2	0-1	2-1	1-1	2-1	2-4	0-1	3-0	3-3	1-0	1-2	0-2	0-3
4	Zamora	0-1	3-0	0-0		1-2	3-2	2-0	0-1	2-1	3-1	1-3	1-1	2-2	0-3	0-0	0-4	2-1	4-1
5	Dptvo. Táchira	0-0	2-2	0-0	1-1		1-0	1-3	0-1	1-0	1-0	2-0	0-2	2-2	0-0	1-0	2-2	3-2	2-1
6	Dptvo. Anzoátegui	5-1	3-0	0-0	1-1	4-0		2-0	0-0	1-1	3-1	1-0	1-0	1-1	1-1	1-2	2-4	1-2	1-7
7	Trujillanos	1-0	1-1	1-0	3-0	1-0	3-3		2-1	1-1	3-1	1-1	0-0	0-1	0-2	0-0	1-2	1-4	1-0
8	Aragua	0-1	3-0	1-1	1-2	3-1	2-4	2-0		1-0	0-1	2-2	1-1	1-0	0-0	0-4	0-0	0-1	0-2
9	Yaracuyanos	0-1	2-0	2-1	1-0	1-0	0-1	1-1	2-2		1-1	4-1	1-1	1-3	3-1	1-4	0-0	1-2	1-1
10	A.C. Mineros	1-0	3-2	2-1	2-2	1-1	1-0	1-1	1-0	2-1		0-0	0-2	2-2	0-1	2-1	0-1	0-3	0-2
11	Monagas SC	3-0	2-1	1-1	2-1	1-0	3-1	1-0	0-0	0-0	1-0		0-0	1-3	2-1	1-1	0-0	0-2	0-3
12	Carabobo	1-0	1-0	1-0	2-2	3-1	1-1	1-0	1-0	1-1	1-1	0-1		0-3	3-1	1-3	1-1	1-5	1-3
13	CD Lara	1-0	0-0	1-2	3-1	3-1	2-1	0-0	2-2	1-4	1-1	1-1	2-2		1-1	1-0	1-0	0-4	0-3
14	Atl. Vigía	2-1	2-0	1-0	2-1	3-1	2-2	3-2	1-0	1-1	3-1	3-1	2-0	1-1		4-1	0-1	1-1	1-0
15	Zulia	5-1	1-0	1-0	5-1	4-0	1-0	2-1	0-1	2-0	2-3	3-2	2-3	0-1	5-1		2-1	1-1	S
16	Est. de Mérida	1-0	1-0	5-1	1-0	3-1	2-1	0-1	3-1	1-1	1-1	6-0	3-1	1-0	0-0	1-3		3-0	1-1
17	Atl. Venezuela	4-0	1-0	2-1	5-0	4-1	10-0	3-0	1-2	2-0	2-2	2-2	0-2	3-2	4-0	2-1	4-0		0-0
18	Caroní	2-2	3-0	0-0	1-0	3-0	2-1	2-1	2-1	0-0	2-0	4-0	3-1	4-0	15-0	4-0	2-1	1-0	

Figura 2.6: Resultados Temporada 2010/2011

De la figura 2.6 podemos concluir que de los 305 partidos, 142 terminaron con un resultado de victoria en casa (color azul), 81 resultaron con un resultado de empate (color verde) y 82 victorias de visitante (color rojo).

2.1.4. Clasificación

Al finalizar la temporada, el Caracas F.C fue el equipo que acumulo el mayor número de puntos en la clasificación general, con 73 puntos, 36 de ellos en partidos jugados en casa. El siguiente equipo fue el Deportivo La Guaira, con un total de 64 puntos, de los cuales 40 fueron obtenidos cuando jugaba de local, el tercer equipo en la clasificación fue el Deportivo Petare con un total de 60 puntos, de los cuales 31 de ellos fueron obtenidos jugando en casa. En la siguiente figura se muestran una tabla con los puntos obtenidos por cada uno de los 18 equipos enfrentados en la Primera División en esta temporada, además del total de puntos, se muestra el total de los partidos ganados, empatados y perdidos, así como también, los goles a favor y en contra. En la clasificación de la figura los equipos aparecen ordenados según el puesto alcanzado en la clasificación general con respecto al total de puntos conseguidos por cada uno de los equipos.

Equipo	J	G	E	P	GF	GC	Puntos
Caracas	34	23	4	7	55	22	73
Dptvo. La Guaira	34	19	7	8	46	24	64
Dptvo. Petare	34	16	12	6	46	31	60
Zamora	34	17	8	9	57	37	59
Dptvo. Táchira	34	16	8	10	54	31	56
Dptvo. Anzoategui	34	15	10	9	61	48	55
Trujillanos	34	14	10	10	40	31	52
Aragua	34	14	9	11	37	32	51
Yaracuyanos	34	12	14	8	39	36	50
A.C. Mineros	34	13	10	11	45	37	49
Monagas SC	34	11	11	12	47	37	44
Carabobo	34	9	13	12	41	39	40
CD Lara	34	9	12	13	47	47	39
Atl. Vigía	34	9	10	15	56	47	37
Zulia	33	11	4	18	38	62	37
Est. de Mérida	34	7	9	18	30	54	30
Atl. Venezuela	34	6	5	23	30	81	23
Caroní	33	3	6	24	16	85	15

Leyenda
J: Partidos jugados.
G: Partidos ganados.
E: Partidos empatados.
P: Partidos perdidos.
GF: Goles a favor.
GC: Goles en contra.

Figura 2.7: Clasificación Temporada 2010/2011

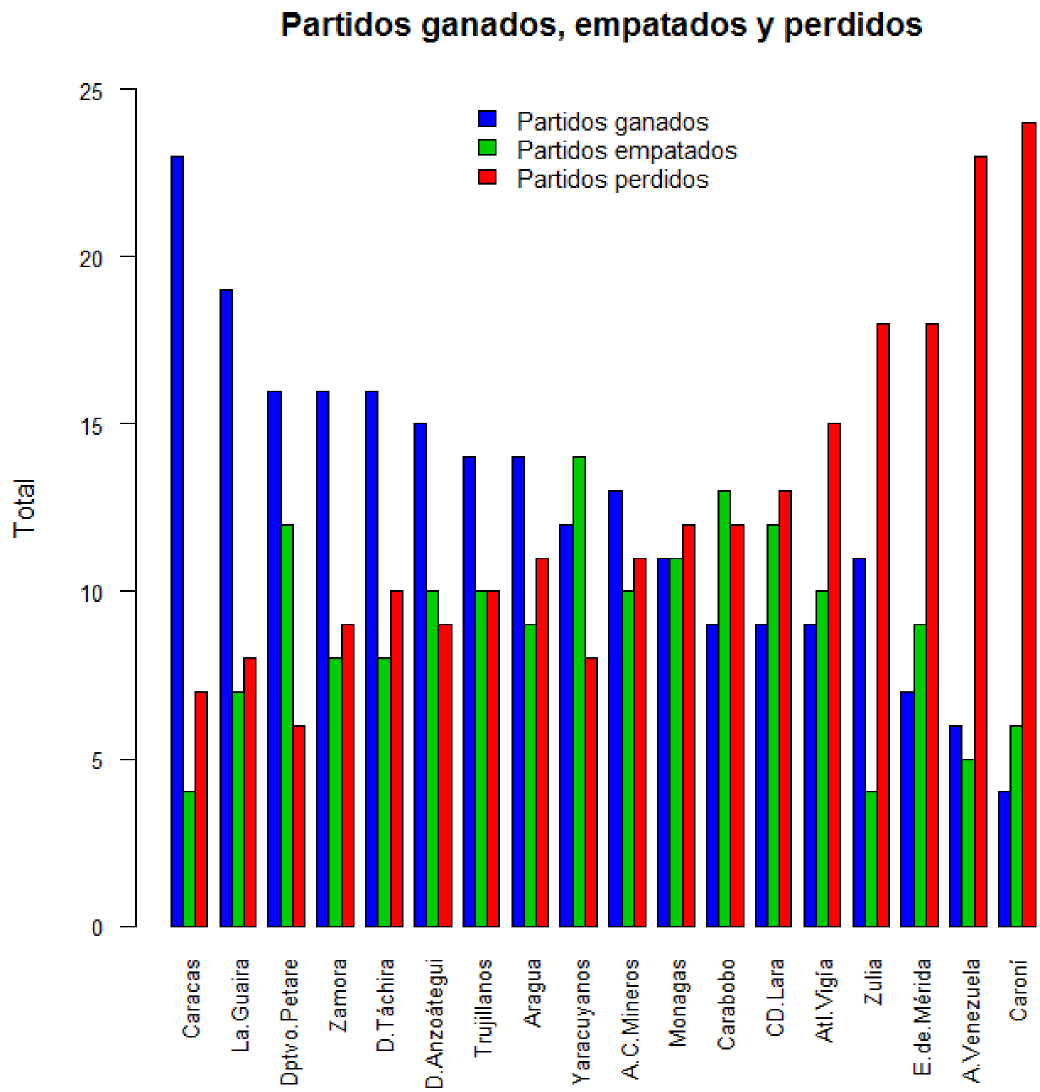

Figura 2.8: Gráfico partidos ganados, empatados y perdidos Temporada 2010/2011

En vista a que los equipos están ordenados de manera descendente según su puntuación obtenida, como es de esperar se observa que los mejores clasificados presentan más partidos ganados y ese número va descendiendo conforme a la posición en la tabla general.

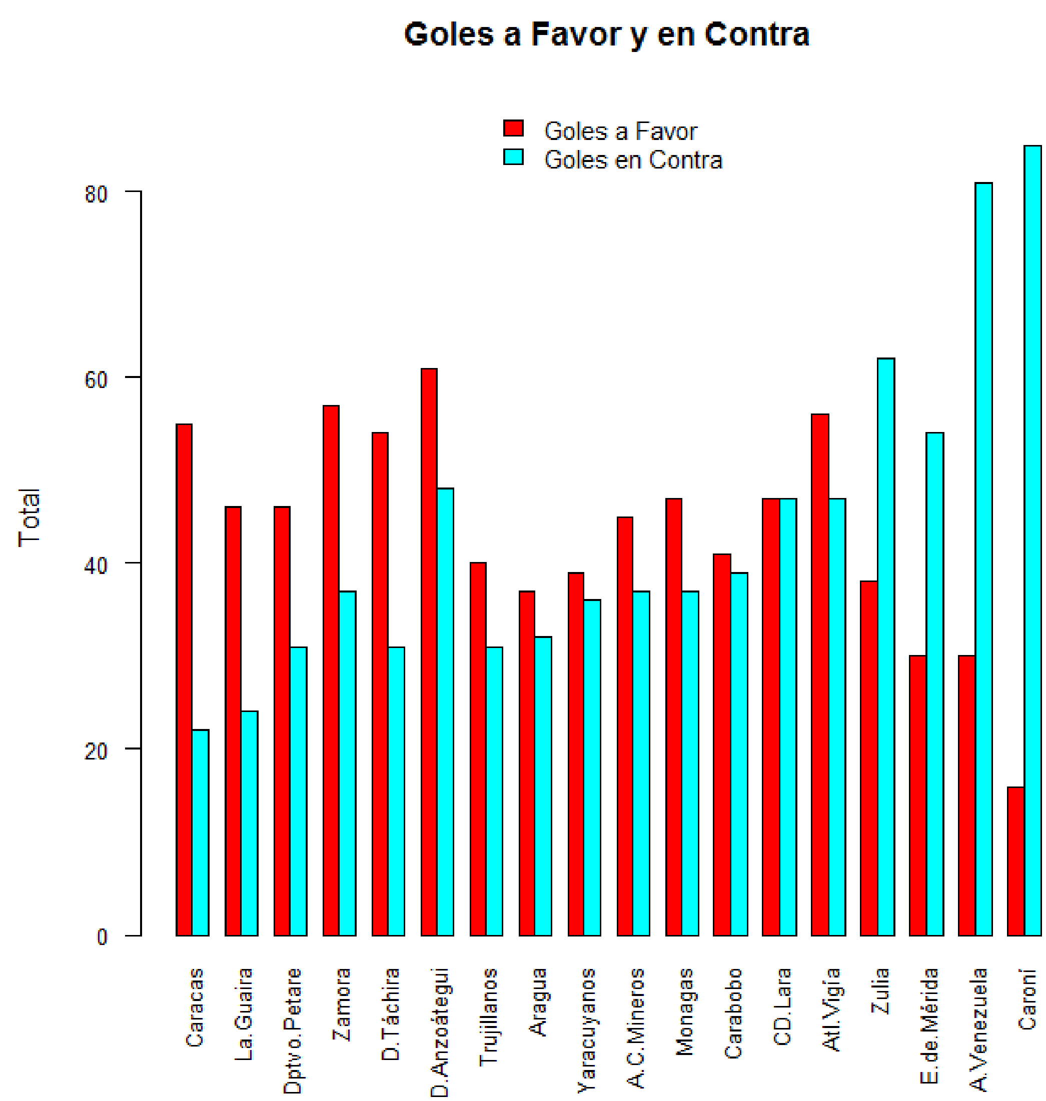

Figura 2.9: Gráfico goles a favor y en contra Temporada 2010/2011

Una tendencia similar se observa en relación a los goles a favor, (ver figura 2.9). Aunque el Caracas F.C que clasifico de primero en la tabla acumulada, presenta un menor número de goles marcados que los equipos que clasificaron en cuarto, quinto y décimo cuarto lugar, cuando observamos los goles en contra notamos que el Caracas F.C presenta el menor número de goles encajados de todos los equipos, esto nos llevaría a pensar que clasificar de primero no solo depende de las goles marcados, que a primera impresión nos dan una idea del ataque, sino también, de los goles encajados los cuales intuitivamente nos dirían lo buena que es la defensa.

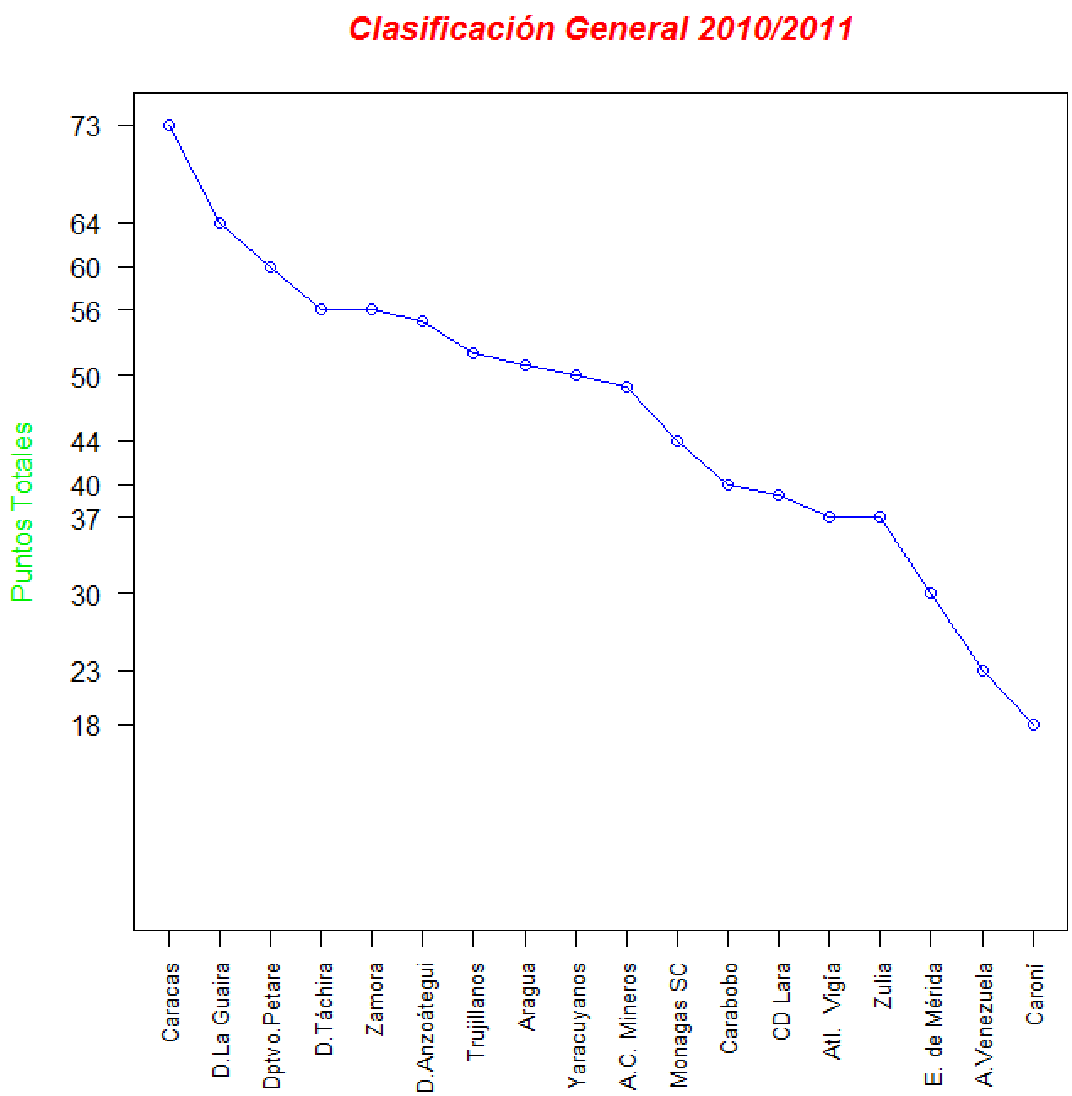

Figura 2.10: Clasificación general según la puntuación Temporada 2010/2011

En la figura 2.10 se muestra la clasificación general de la temporada 2010/2011 teniendo como base el total de puntos, en este gráfico se observa como el Caracas F.C obtuvo una ventaja de puntos significativa con respecto a un grupos de 5 equipos que ocupan el segundo, tercer, cuarto, quinto y sexto lugar, a saber Deportivo La Guaira, Deportivo Petare, Zamora F.C, Deportivo Táchira y Deportivo Anzoátegui respectivamente, los cuales marcan una diferencia en puntos acumulados, con respectos a un segundo grupo de equipos que se encuentra posteriormente cuya diferencia en puntuación se ve estable y por último se encuentra un grupo de equipos en la zona baja de la tabla.

2.1.5. Temporada 2011-2012

Resultados de los partidos

En la temporada 2011-2012 jugaron 18 equipos en la Primera División la temporada comenzó el 11 de Agosto del 2011 y finalizó el 13 de Mayo del 2012. En total, se celebraron 306 partidos, a continuación se muestran los equipos que participaron en esta temporada con sus respectivos resultados.

	Casa →	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	CD Lara		2-3	1-0	0-1	1-2	3-4	1-3	0-2	1-1	1-3	0-2	1-1	1-2	0-0	0-2	1-1	2-3	1-1
2	Caracas	0-0		1-0	1-1	2-0	1-0	0-0	3-3	0-0	2-1	2-0	2-1	1-2	0-1	2-3	0-2	1-2	1-4
3	Dptvo. Anzoátegui	0-0	0-0		2-0	1-0	0-1	1-0	1-0	0-2	0-0	1-1	0-2	0-1	1-1	1-0	0-0	0-1	0-1
4	A.C. Mineros	5-1	4-0	0-1		1-0	2-4	1-2	2-2	1-0	1-1	0-1	1-2	1-1	0-1	1-2	1-1	2-2	0-2
5	Zulia	1-0	2-1	3-2	0-2		1-2	1-1	1-2	1-4	1-0	1-0	1-0	1-0	1-0	1-1	2-2	0-1	0-1
6	Dptvo. Petare	2-1	1-4	1-3	0-1	0-1		1-2	0-2	0-0	0-0	2-0	3-3	1-0	0-3	3-2	1-1	2-0	2-2
7	Zamora	4-0	1-0	1-0	3-1	0-1	2-2		0-0	2-1	1-3	3-1	0-0	2-2	1-2	1-2	1-1	0-0	1-3
8	Yaracuyanos	1-0	2-0	2-1	1-1	2-0	1-0	1-2		0-3	2-2	0-0	1-1	2-1	0-0	1-1	1-0	1-1	0-1
9	Aragua	1-1	2-1	0-1	2-2	1-3	1-0	2-2	1-0		2-1	0-0	2-0	0-0	1-1	2-1	1-1	1-1	0-1
10	Trujillanos	2-0	1-0	2-1	2-0	1-0	2-2	3-1	0-0	2-0		1-1	2-0	1-1	1-1	0-2	1-0	1-1	1-3
11	Monagas SC	2-1	2-1	2-0	3-1	1-0	1-1	1-0	1-0	0-2	2-0		1-0	1-1	2-0	2-0	0-0	2-1	3-1
12	Dptvo. Táchira	3-1	1-1	2-1	2-0	1-1	1-0	1-1	2-0	1-0	1-1	3-0		3-0	2-0	0-0	2-1	1-0	1-2
13	Llaneros	1-0	1-0	1-0	2-1	3-3	0-0	0-0	4-0	2-1	4-0	1-1	3-1		3-2	2-1	1-2	0-0	1-1
14	Dptvo. La Guaira	4-0	2-0	0-0	1-0	1-1	0-0	4-0	1-2	1-1	0-2	1-0	0-1	2-0		1-0	0-5	2-0	2-0
15	Atl. Vigía	3-0	2-1	2-1	4-0	1-0	0-0	3-0	2-0	3-2	3-0	2-1	0-0	2-0	0-0		1-0	1-1	1-2
16	Est. de Mérida	3-1	2-1	3-1	1-1	1-0	1-1	3-2	5-1	0-0	2-0	3-2	3-0	1-1	0-0	1-1		2-2	5-1
17	Carabobo	2-0	2-1	0-0	2-0	3-1	2-1	2-1	3-2	0-0	1-1	2-0	2-1	1-0	0-2	1-0	2-2		1-0
18	Tucanes	5-1	3-0	6-1	1-1	3-2	1-0	2-2	2-3	1-0	2-0	3-2	1-0	2-1	3-1	1-1	2-0	1-0	

Figura 2.11: Resultados Temporada 2011/2012

De la figura 2.11 podemos concluir que de los 306 partidos, 147 terminaron con un resultado de victoria en casa (color azul), 92 resultaron con un resultado de empate (color verde) y 67 victorias de visitante (color rojo).

2.1.6. Clasificación

Al finalizar la temporada, el C.D Lara fue el equipo que acumulo el mayor número de puntos en la clasificación general, con 83 puntos, 45 de ellos en partidos jugados en casa, el siguiente equipo fue el Caracas F.C, con un total de 64 puntos, de los cuales 41 fueron obtenidos cuando jugaba de local, el tercer equipo en la clasificación fue el Deportivo Anzoátegui con un total de 62 puntos, de los cuales 38 de ellos fueron obtenidos jugando en casa. En la siguiente figura se muestran una tabla con los puntos obtenidos por cada uno de los 18 equipos enfrentados en la Primera División en esta temporada, observe que el Carabobo F.C tiene un punto menos debido a una sanción de la F.V.F, además del total de puntos, se muestra el total de los partidos ganados, empatados y perdidos, así como también, los goles a favor y en contra. En la clasificación de la figura 2.12 los equipos aparecen ordenados según el puesto alcanzado en la clasificación general con respecto al total de puntos conseguidos por cada uno de los equipos.

Leyenda
J: Partidos jugados.
G: Partidos ganados.
E: Partidos empatados.
P: Partidos perdidos.
GF: Goles a favor.
GC: Goles en contra.

Equipo	J	G	E	P	GF	GC	Puntos
CD Lara	34	25	8	1	71	19	83
Caracas	34	19	7	8	50	27	64
Dptvo. Anzoátegui	34	18	8	8	37	17	62
A.C. Mineros	34	17	10	7	50	22	61
Zulia	34	16	6	12	42	20	54
Dptvo. Petare	34	13	12	9	44	29	51
Zamora	34	12	12	10	46	30	48
Yaracuyanos	34	11	11	12	42	27	44
Aragua	34	10	14	10	31	24	44
Trujillanos	34	10	12	12	39	23	42
Monagas SC	34	11	8	15	34	22	41
Dptvo. Táchira	34	10	10	14	32	26	40
Llaneros	34	9	12	13	35	21	39
Dptvo. La Guaira	34	9	11	14	27	21	38
Atl. Vigía	34	8	9	17	28	31	33
Est. de Mérida	34	5	16	13	32	26	31
Carabobo	34	6	12	16	31	22	29
Tucanes	34	5	6	23	36	43	21

Figura 2.12: Clasificación Temporada 2011/2012

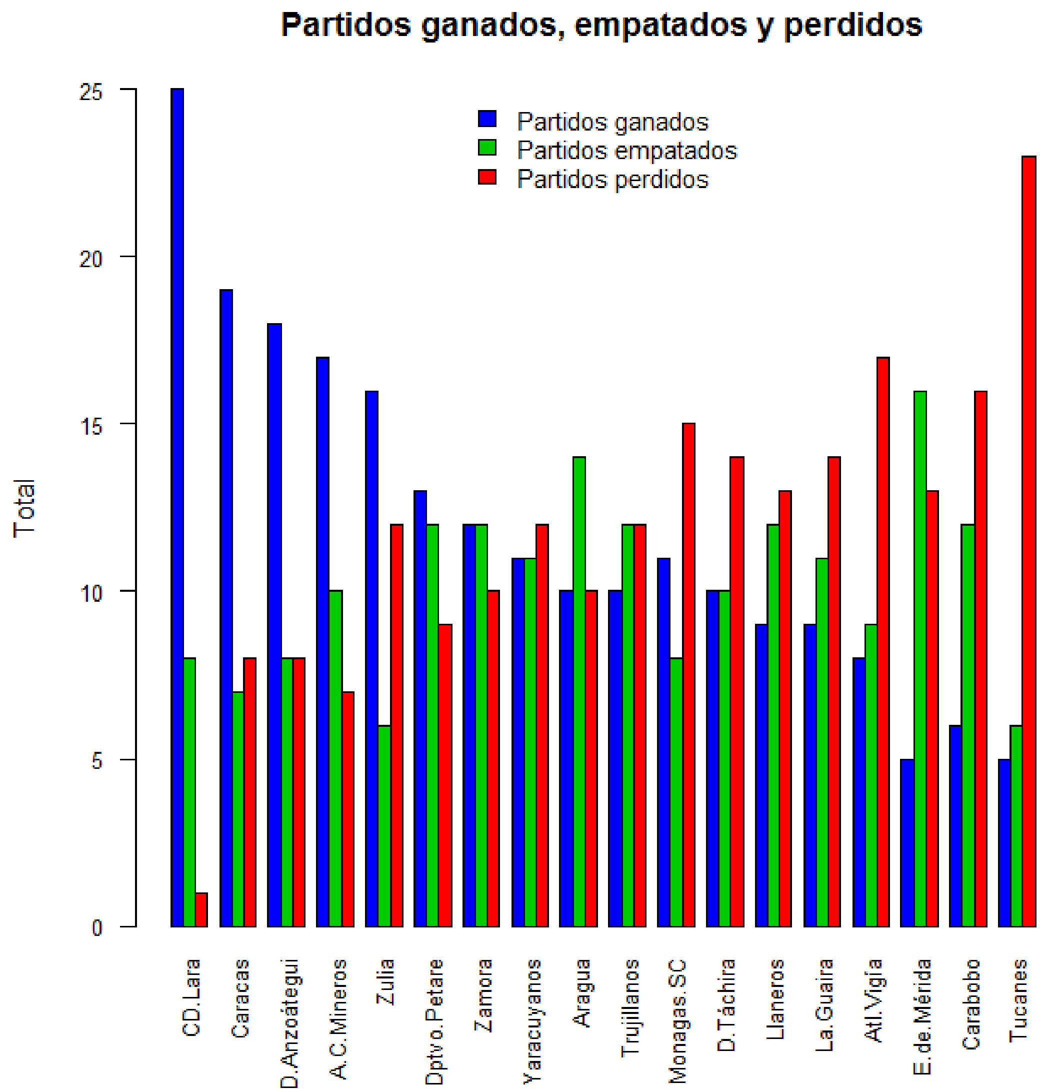

Figura 2.13: Gráfico partidos ganados, empatados y perdidos Temporada 2011/2012

En vista a que los equipos están ordenados de manera descendente según su puntuación obtenida, como es de esperar se observa que los mejores clasificados presentan más partidos ganados y ese número va descendiendo conforme a la posición en la tabla general.

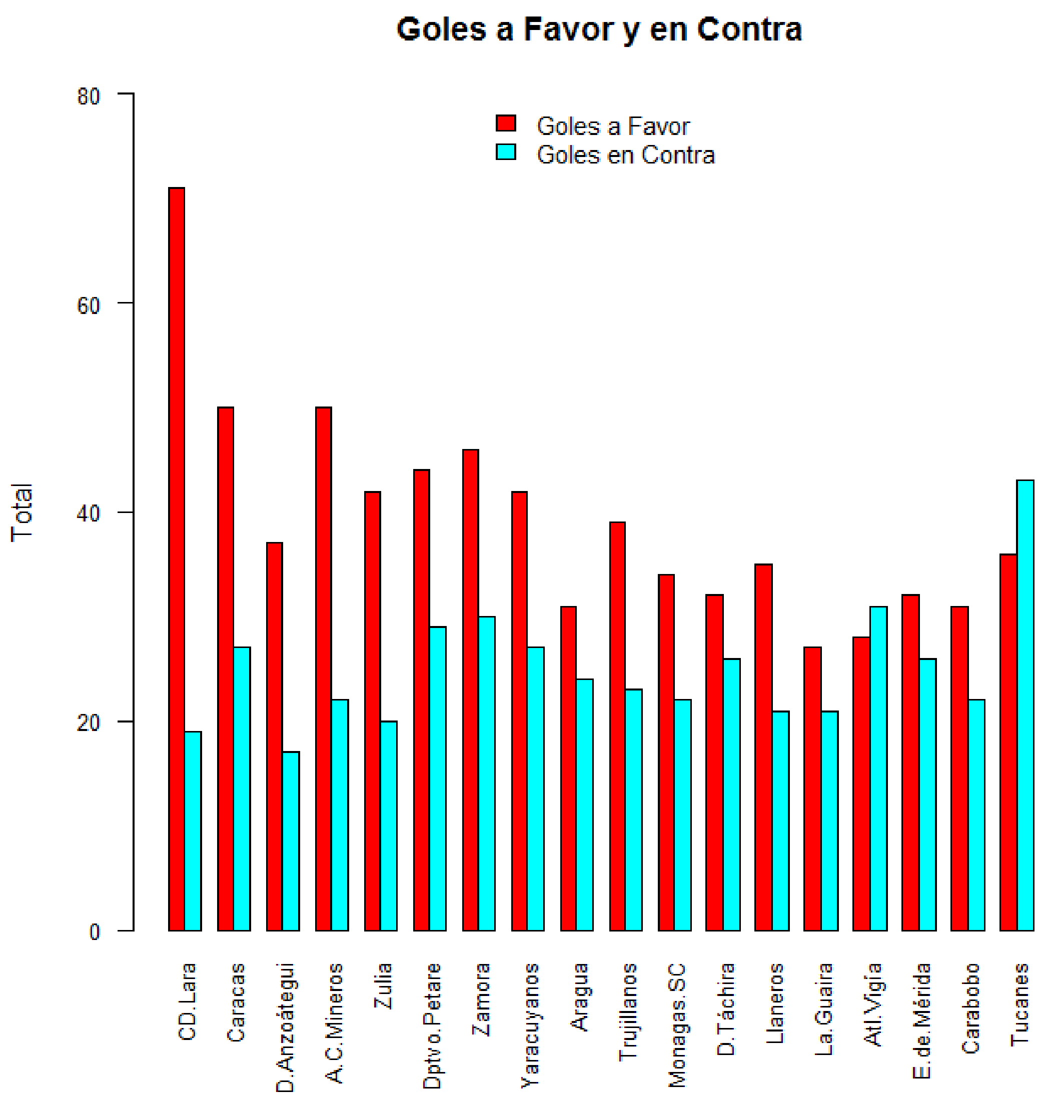

Figura 2.14: Gráfico goles a favor y en contra Temporada 2011/2012

Una tendencia similar se observa en relación a los goles a favor, (ver figura 2.14). El C.D Lara que clasifico de primero en la tabla acumulada, presenta un mayor número de goles marcados que los demás equipos, y un menor número de goles encajados comparado con casi todos los equipos que se encuentran más abajo en la tabla, ya que cuando observamos los goles en contra equipo por equipo notamos que el Deportivo Anzoátegui presenta el menor número de goles encajados de todos los equipos, esto nos llevaría a pensar que clasificar de primero no solo depende de las goles marcados, que a primera impresión nos dan una idea del ataque, sino también, de los goles encajados los cuales intuitivamente nos dirían lo buena que es la defensa.

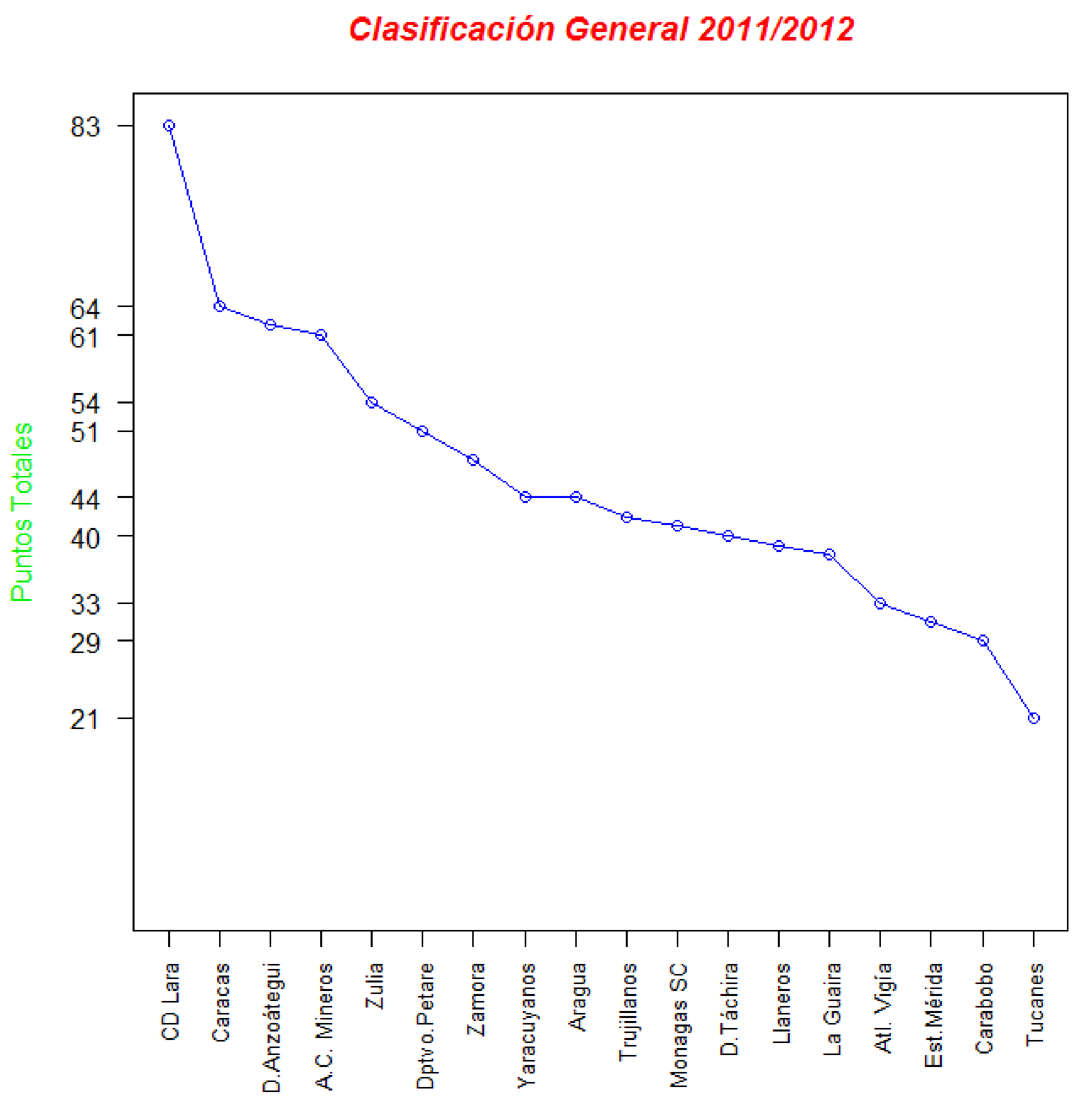

Figura 2.15: Clasificación general según la puntuación Temporada 2011/2012

En la figura 2.15 se muestra la clasificación general de la temporada 2011/2012 teniendo como base el total de puntos, en este gráfico se observa como el C.D Lara obtuvo una ventaja de puntos significativa con respecto a un grupos de 3 equipos que ocupan el segundo, tercer y cuarto lugar, a saber Caracas F.C, Deportivo Anzoátegui y A.C Mineros respectivamente los cuales marcan una diferencia en puntos acumulados, con respectos a un segundo grupo de equipos que se encuentra posteriormente, luego se presentan un tercer grupo de equipos con puntuaciones similares y por último se encuentra un grupo de equipos en la zona baja de la tabla.

2.1.7. Temporada 2012-2013

Resultados de los partidos

En la temporada 2012-2013 jugaron 18 equipos en la Primera División la temporada comenzó el 11 de Agosto del 2012 y finalizó el 12 de Mayo del 2013. En total, se celebraron 306 partidos. A continuación se muestran los equipos que participaron en esta temporada con sus respectivos resultados.

	Casa	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Dptvo. Anzoátegui		2-1	1-1	1-2	5-0	0-2	2-0	0-0	0-0	0-1	2-0	1-3	1-2	3-1	0-2	1-1	0-1	2-2
2	Caracas	1-0		1-1	1-2	0-0	1-1	0-0	0-0	1-3	0-2	1-1	1-1	0-2	0-2	4-1	4-1	3-4	1-2
3	Zamora	1-0	1-0		0-3	0-0	1-0	0-1	0-2	2-0	0-2	1-2	0-5	0-0	0-0	1-1	0-0	1-3	2-3
4	CD Lara	1-1	3-0	0-3		1-1	2-1	3-3	1-2	3-3	2-3	0-1	0-1	1-1	0-1	0-1	0-1	1-0	2-2
5	Trujillanos	1-0	0-0	5-2	1-0		2-0	3-0	0-4	1-1	1-2	1-3	1-1	0-0	1-0	1-1	0-2	1-2	2-1
6	A.C. Mineros	3-2	0-1	3-0	1-0	1-0		4-0	1-0	0-1	3-1	1-2	3-1	0-0	2-2	0-2	2-2	0-2	0-2
7	Llaneros	2-0	4-1	0-2	0-0	3-0	2-2		1-1	1-2	1-1	0-0	2-0	1-2	1-0	0-0	2-2	3-1	0-1
8	Atl. Venezuela	1-0	0-0	3-1	3-0	1-2	2-0	1-3		0-0	1-1	2-1	0-1	1-0	2-1	0-0	1-4	0-0	0-2
9	Dptvo Táchira	3-2	0-0	1-1	0-0	3-1	2-1	1-1	2-0		2-1	1-0	3-4	3-2	0-0	1-1	1-0	0-2	1-2
10	Aragua	2-1	1-1	1-1	0-0	2-1	1-1	0-0	1-1	3-0		1-0	1-3	2-0	0-1	3-2	1-0	0-1	2-1
11	Dptvo. La Guaira	1-0	3-0	1-0	3-2	0-3	1-1	2-0	2-4	3-0	3-1		1-0	1-2	2-3	3-1	0-0	3-2	0-1
12	Atl. Vigía	1-1	1-0	2-1	2-1	2-0	4-0	2-1	0-0	4-0	0-0	1-2		0-0	4-0	0-1	0-0	4-1	3-0
13	Zulia	3-1	2-1	3-2	3-2	1-1	3-1	1-1	2-0	1-1	1-0	0-0	2-0		1-1	1-2	1-0	0-0	2-1
14	Yaracuyanos	2-0	1-0	4-0	1-0	1-0	1-0	1-0	2-1	1-0	3-1	0-2	1-0	3-3		0-1	1-0	0-2	0-2
15	Dptvo. Petare	2-0	3-0	2-0	3-2	3-0	1-3	1-3	1-1	1-3	2-0	1-0	1-0	4-0	0-1		5-1	0-0	1-2
16	Est. de Mérida	2-1	1-0	3-0	2-2	2-1	3-1	3-1	1-0	1-0	3-3	1-1	1-1	3-1	1-0	1-0		0-0	1-0
17	Portuguesa	5-1	1-0	2-1	4-2	1-0	4-2	1-0	3-3	2-1	2-0	1-2	3-0	1-0	0-1	0-0	1-1		0-1
18	Monagas SC	2-1	2-1	2-0	3-2	1-0	0-0	2-3	4-1	3-0	1-2	2-1	2-0	2-0	2-2	0-0	0-0	1-0	

Figura 2.16: Resultados Temporada 2012/2013

De la figura 2.16 podemos concluir que de los 306 partidos, 145 terminaron con un resultado de victoria en casa (color azul), 82 resultaron con un resultado de empate (color verde) y 79 victorias de visitante (color rojo).

2.1.8. Clasificación

Al finalizar la temporada, el Deportivo Anzoátegui fue el equipo que acumulo el mayor número de puntos en la clasificación general, con 73 puntos, 47 de ellos en partidos jugados en casa, el siguiente equipo fue el Caracas F.C, con un total de 68 puntos, de los cuales 40 fueron obtenidos cuando jugaba de local, el tercer equipo en la clasificación fue el Zamora F.C con un total de 66 puntos, de los cuales 39 de ellos fueron obtenidos jugando en casa. En la siguiente figura se muestran una tabla con los puntos obtenidos por cada uno de los 18 equipos enfrentados en la Primera División en esta temporada, además del total de puntos, se muestra el total de los partidos ganados, empatados y perdidos, así como también, los goles a favor y en contra. En la clasificación de la figura 2.17 los equipos aparecen ordenados según el puesto alcanzado en la clasificación general con respecto al total de puntos conseguidos por cada uno de los equipos.

Leyenda

J: Partidos jugados.
G: Partidos ganados.
E: Partidos empatados.
P: Partidos perdidos.
GF: Goles a favor.
GC: Goles en contra.

Equipo	J	G	E	P	GF	GC	Puntos
Dptvo. Anzoátegui	34	22	7	5	52	32	73
Caracas	34	19	11	4	48	25	68
Zamora	34	19	9	6	57	26	66
CD Lara	34	18	10	6	53	40	64
Trujillanos	34	16	9	9	47	31	57
A.C. Mineros	34	16	8	10	48	40	56
Llaneros	34	12	12	10	42	40	48
Atl. Venezuela	34	12	11	11	40	38	47
Dptvo Táchira	34	11	10	13	45	42	43
Aragua	34	10	10	14	38	43	40
Dptvo. La Guaira	34	11	6	17	36	47	39
Atl. Vigía	34	10	8	16	31	50	38
Zulia	34	8	12	14	37	42	36
Yaracuyanos	34	10	6	18	31	39	36
Dptvo. Petare	34	9	9	16	31	47	36
Est. de Mérida	34	6	13	15	32	44	31
Portuguesa	34	8	7	19	33	52	31
Monagas SC	34	7	6	21	32	55	27

Figura 2.17: Clasificación Temporada 2012/2013

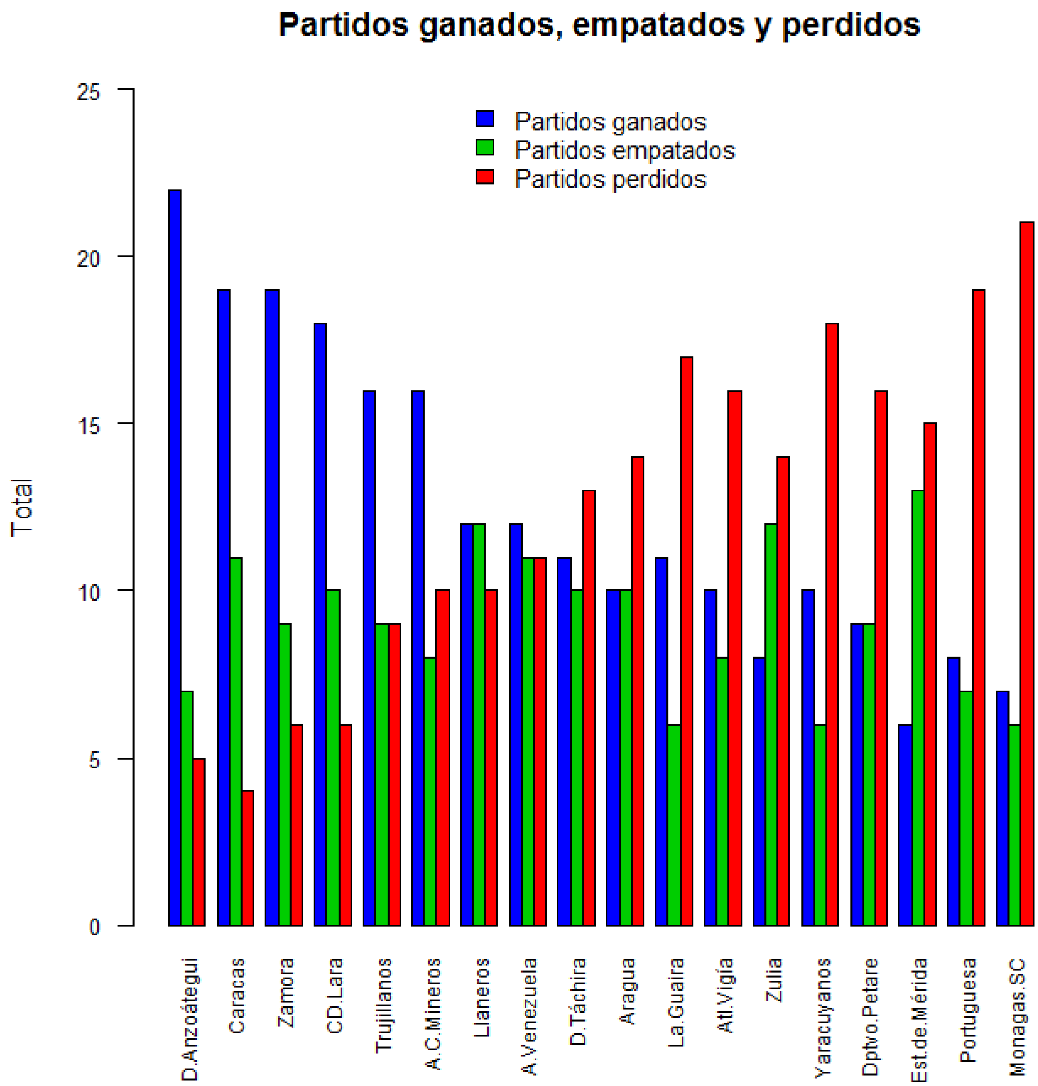

Figura 2.18: Gráfico partidos ganados, empatados y perdidos Temporada 2012/2013

En vista a que los equipos están ordenados de manera descendente según su puntuación obtenida, como es de esperar se observa que los mejores clasificados presentan más partidos ganados y ese número va descendiendo conforme a la posición en la tabla general.

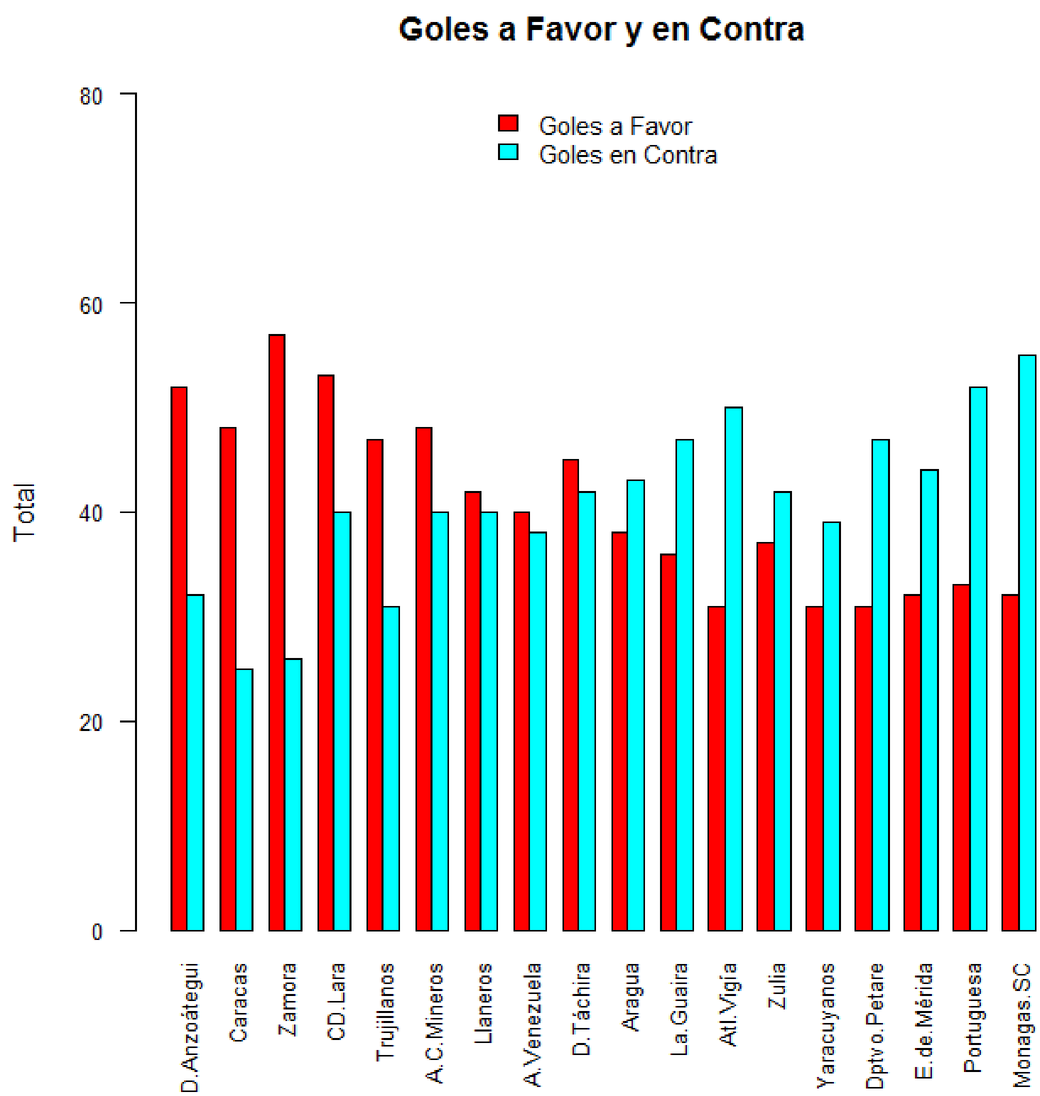

Figura 2.19: Gráfico goles a favor y en contra Temporada 2012/2013

Una tendencia similar se observa en relación a los goles a favor, (ver figura 2.19). El Deportivo Anzoátegui que clasifico de primero en la tabla acumulada, presenta un mayor número de goles marcados que la mayoría de los demás equipos, pero el Zamora F.C y el C.D Lara que ocuparon el tercero y el cuarto lugar de la tabla respectivamente, marcaron más goles que el Deportivo Anzoátegui, además, el Caracas F.C y el Zamora F.C, segundo y tercero respectivamente en la tabla acumulada recibieron menos goles que el Deportivo Anzoátegui, siguiendo la misma línea, la intuición nos dice que el clasificar primero trae como consecuencia tener el mayor número de goles marcados, es decir mayor ataque, o tener el menor número de goles encajados, es decir mejor defensa, este pensamiento nos llevaría a pensar a que el Zamora F.C, tercero en la tabla acumulada debería estar entre los dos primeros lugares de la clasificación, todo esto nos dice que de una forma u otra existen otros parámetros que influyen en el desenvolvimiento de un partido de futbol, también podríamos pensar que los marcadores bajos (1-0) o una diferencia pequeña entre los goles marcados y recibidos (diferencia de un gol) en un partido a favor de un equipo lo podría llevar a los primeros lugares de la clasificaciones que es lo que debió haber pasado en este caso.

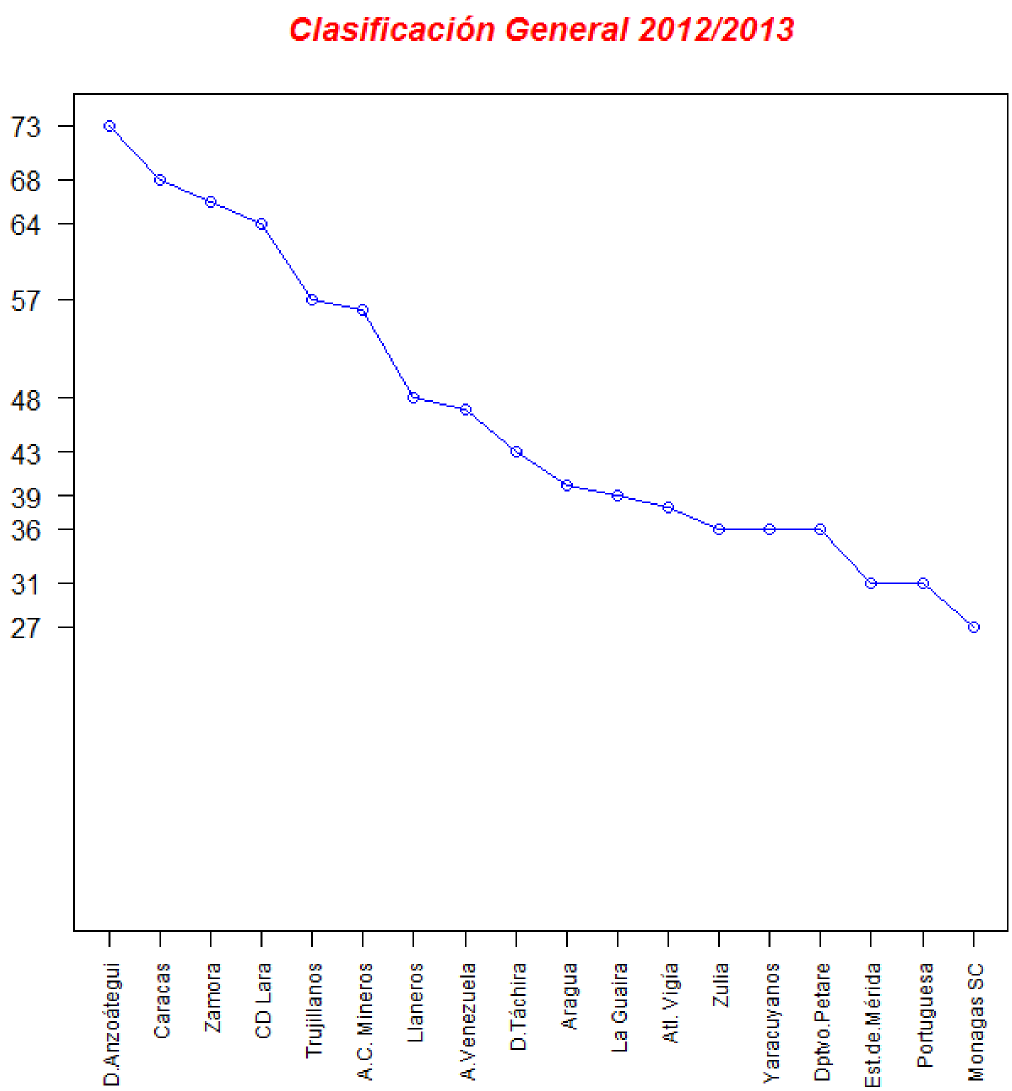

Figura 2.20: Clasificación general según la puntuación Temporada 2012/2013

En la figura 2.19 se muestra la clasificación general de la temporada 2012/2013 teniendo como base el total de puntos, en este gráfico se observa como el Deportivo Anzoátegui, el Caracas F.C, El Zamora F.C y el C.D Lara obtuvieron una ventaja de puntos significativa con respecto a un grupos de 2 equipos que ocupan el quinto y sexto lugar, a saber Trujillanos y A.C Mineros respectivamente los cuales marcan una diferencia en puntos acumulados, con respectos a un tercer grupo de equipos que se encuentra posteriormente, luego se presentan un cuarto grupo de equipos con puntuaciones similares y por último se encuentra un grupo de equipos en la zona baja de la tabla.

2.1.9. Temporada 2013-2014

Resultados de los partidos

En la temporada 2013-2014 jugaron 18 equipos en la Primera División la temporada comenzó el 10 de Agosto del 2013 y finalizó el 11 de Mayo del 2014. En total, se celebraron 306 partidos. A continuación se muestran los equipos que participaron en esta temporada con sus respectivos resultados.

	Casa	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	A.C. Mineros		0-0	1-1	0-0	0-0	0-0	2-3	4-0	1-3	1-0	0-1	1-1	0-2	0-1	3-0	0-2	1-2	1-2
2	Zamora	4-2		2-0	2-2	1-1	0-0	1-1	0-1	1-2	2-3	1-0	0-3	1-3	1-0	0-2	1-1	0-2	1-1
3	Dptvo Táchira	1-0	3-0		1-0	2-0	1-0	1-1	0-1	1-1	3-3	1-0	0-1	1-2	2-4	1-1	1-1	2-3	0-2
4	Caracas	1-1	2-1	3-2		0-0	1-1	1-1	2-2	0-0	2-3	1-1	0-2	2-1	0-3	0-1	1-1	1-2	0-1
5	Dptvo. Anzoátegui	4-3	0-0	2-1	3-2		3-1	2-0	1-2	1-2	1-0	3-0	1-2	0-1	2-3	0-1	1-1	0-2	1-2
6	Trujillanos	1-0	5-2	2-2	4-0	1-1		1-0	0-1	2-2	1-1	1-0	0-0	0-3	0-1	0-0	0-0	0-5	0-4
7	Tucanes	5-1	3-1	2-1	2-0	3-2	2-2		0-0	3-2	1-2	1-1	0-1	0-0	1-2	3-2	3-1	0-1	0-2
8	Carabobo	1-0	6-2	3-1	3-4	2-1	1-1	0-0		2-1	1-2	2-4	0-2	0-0	1-1	0-1	2-3	3-1	1-1
9	Atl. Venezuela	3-0	0-0	4-1	0-0	1-1	0-0	4-0	0-1		0-0	0-1	0-0	1-1	0-1	2-0	0-0	0-2	2-3
10	Aragua	2-1	2-0	2-1	2-1	1-2	2-0	0-0	1-0	1-1		0-0	0-1	0-0	3-0	0-2	0-2	0-1	1-1
11	CD Lara	3-1	2-0	0-0	2-2	2-2	1-4	1-0	3-2	1-3	0-0		0-0	0-0	4-1	0-0	0-2	4-1	1-0
12	Dptvo. La Guaira	1-1	2-1	3-1	4-1	3-0	1-1	0-2	2-0	0-0	0-0	0-1		1-1	0-0	1-1	0-0	2-0	0-2
13	Dptvo. Petare	2-1	2-1	1-1	2-2	1-0	1-0	1-1	0-1	0-0	2-1	0-1	3-0		1-1	1-2	0-2	2-0	2-4
14	Zulia	2-0	3-2	2-0	1-0	1-0	1-1	2-2	0-2	2-0	4-1	3-3	2-0	1-1		0-1	1-1	1-1	2-0
15	Llaneros	0-0	3-1	3-1	2-1	6-1	3-1	1-0	4-0	1-0	3-0	2-1	1-0	0-1	3-3		3-3	1-2	1-2
16	Est. de Mérida	3-0	4-2	4-0	1-1	1-1	4-1	3-1	1-0	1-3	2-3	0-1	2-1	1-0	4-3	5-1		0-1	0-1
17	Atl. Vigía	4-1	4-0	5-2	2-0	3-1	4-0	3-1	1-0	2-1	1-2	4-0	2-1	1-1	2-1	1-3	3-3		1-0
18	Yaracuyanos	1-0	2-0	4-2	1-0	1-0	5-2	2-0	3-1	2-1	2-1	1-1	2-2	1-1	1-0	3-1	1-1	3-2	

Figura 2.21: Resultados Temporada 2013/2014

De la figura 2.21 podemos concluir que de los 306 partidos, 179 terminaron con un resultado de victoria en casa (color azul), 93 resultaron con un resultado de empate (color verde) y 34 victorias de visitante (color rojo).

2.1.10. Clasificación

Al finalizar la temporada, el A.C Mineros fue el equipo que acumulo el mayor número de puntos en la clasificación general, con 75 puntos, 45 de ellos en partidos jugados en casa, el siguiente equipo fue el Zamora F.C, con un total de 72 puntos, de los cuales 45 fueron obtenidos cuando jugaba de local, el tercer equipo en la clasificación fue el Deportivo Táchira con un total de 66 puntos, de los cuales 43 de ellos fueron obtenidos jugando en casa. En la siguiente figura se muestran una tabla con los puntos obtenidos por cada uno de los 18 equipos enfrentados en la Primera División en esta temporada, además del total de puntos, se muestra el total de los partidos ganados, empatados y perdidos, así como también, los goles a favor y en contra. En la clasificación de la figura 2.22 los equipos aparecen ordenados según el puesto alcanzado en la clasificación general con respecto al total de puntos conseguidos por cada uno de los equipos.

Leyenda
J: Partidos jugados.
G: Partidos ganados.
E: Partidos empatados.
P: Partidos perdidos.
GF: Goles a favor.
GC: Goles en contra.

Equipo	J	G	E	P	GF	GC	Puntos
A.C. Mineros	34	22	9	3	56	27	75
Zamora	34	21	9	4	67	31	72
Dptvo. Táchira	34	19	9	6	63	38	66
Caracas	34	16	14	4	55	33	62
Dptvo. Anzoátegui	34	17	9	8	52	38	60
Trujillanos	34	13	15	6	53	34	54
Tucanes	34	13	11	10	44	37	50
Carabobo	34	14	7	13	47	42	49
Atl. Venezuela	34	11	14	9	32	39	47
Aragua	34	11	10	13	39	39	43
CD Lara	34	9	12	13	38	40	39
Dptvo. La Guaira	34	8	13	13	27	38	37
Dptvo. Petare	34	7	14	13	28	39	35
Zulia	34	8	10	16	34	52	34
Llaneros	34	8	7	19	37	56	31
Est. de Mérida	34	6	13	15	37	60	31
Atl. Vigía	34	7	3	24	36	71	24
Yaracuyanos	34	3	7	24	29	63	16

Figura 2.22: Clasificación Temporada 2013/2014

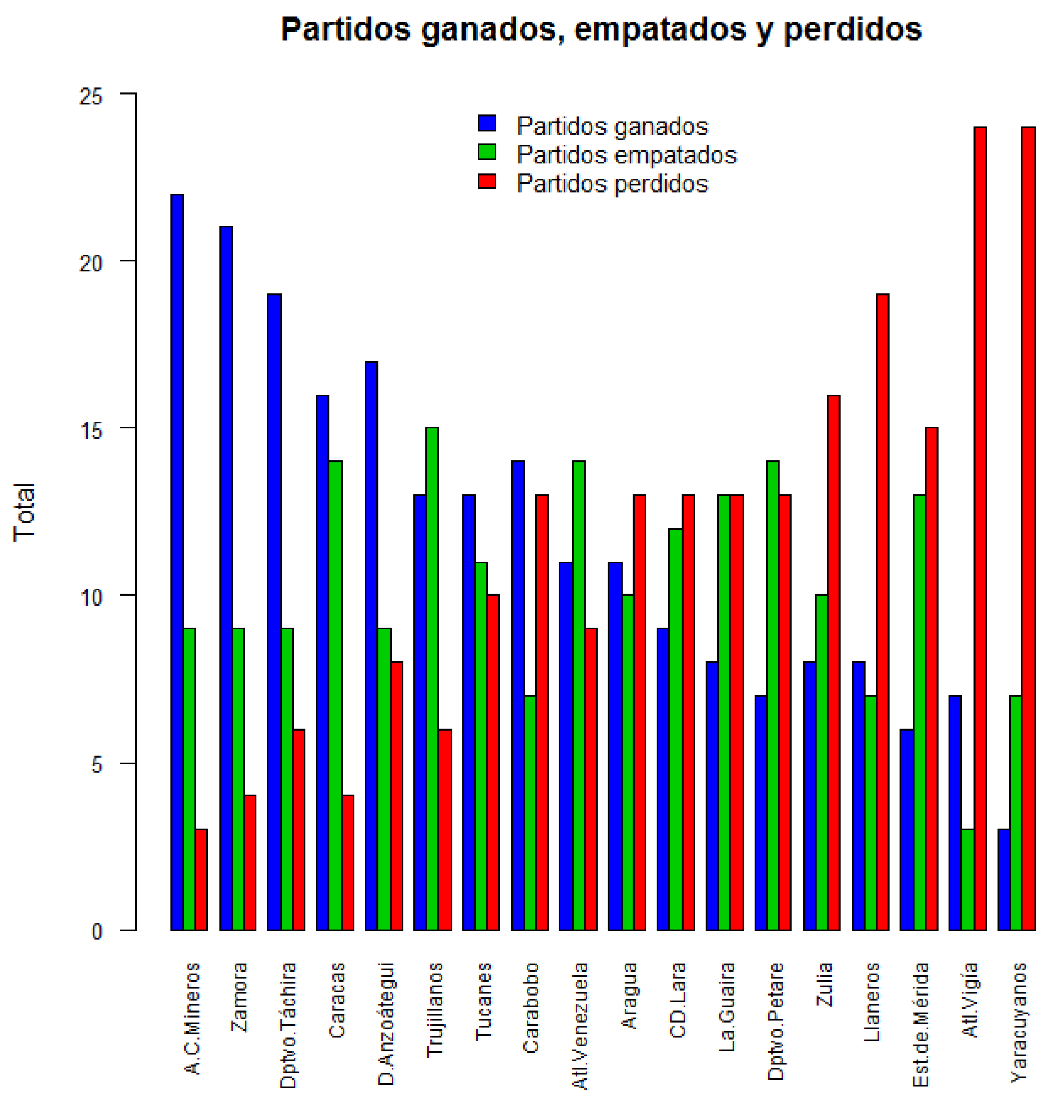

Figura 2.23: Gráfico partidos ganados, empatados y perdidos Temporada 2013/2014

En vista a que los equipos están ordenados de manera descendente según su puntuación obtenida, como es de esperar se observa que los mejores clasificados presentan más partidos ganados y ese número va descendiendo conforme a la posición en la tabla general.

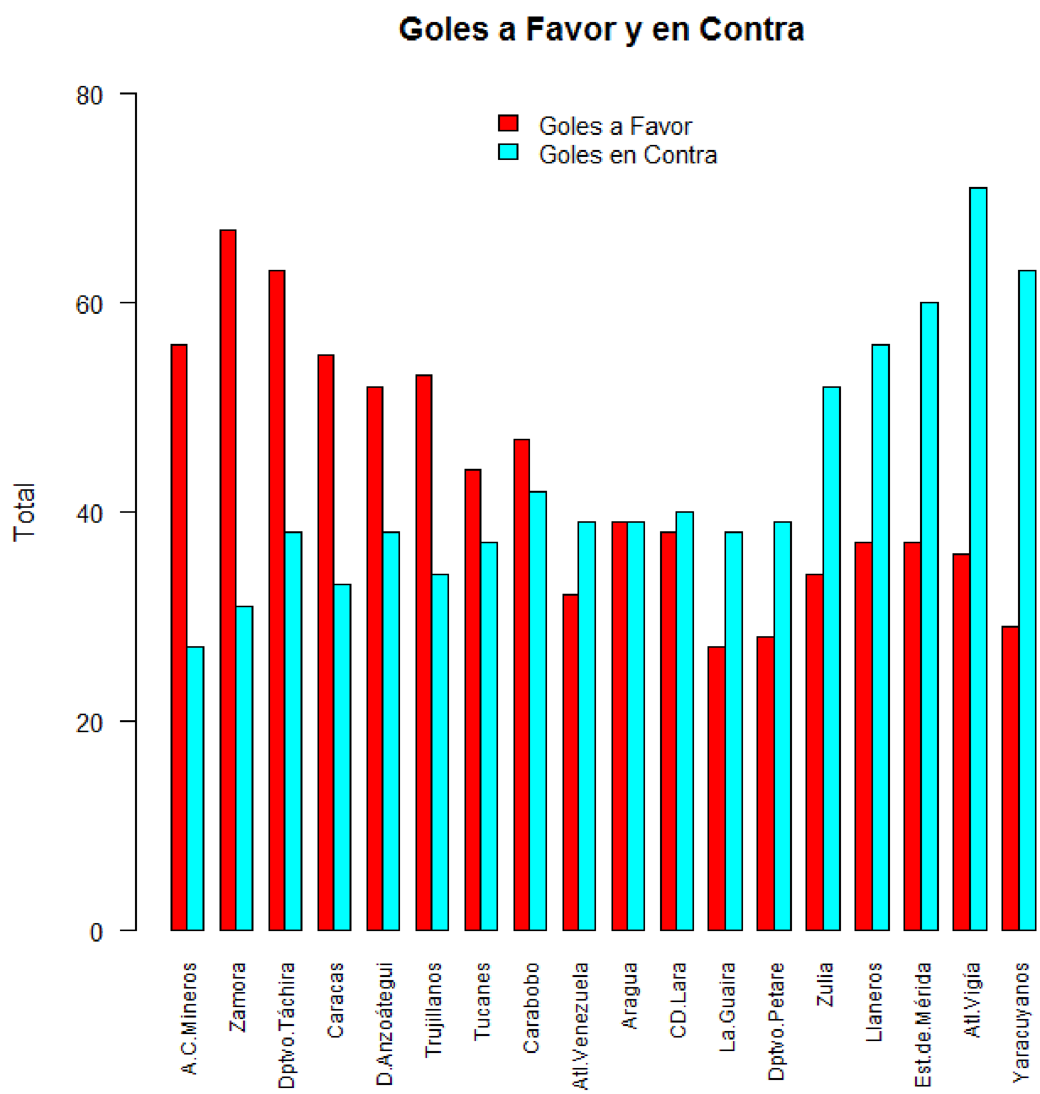

Figura 2.24: Gráfico goles a favor y en contra Temporada 2013/2014

Una tendencia similar se observa en relación a los goles a favor, (ver figura 2.24). El A.C Mineros que clasifico de primero en la tabla acumulada, presenta un mayor número de goles marcados que la mayoría de los demás equipos, por otro lado, el Zamora y el Deportivo Táchira que clasificaron de segundo y tercero respectivamente, tienen una mayor cantidad de goles marcados que el A.C Mineros, pero el A.C Mineros tiene el menor número de goles encajados de todos los equipos, esto nos llevaría a pensar que clasificar de primero no solo depende de las goles marcados, que a primera impresión nos dan una idea del ataque, sino también, de los goles encajados los cuales intuitivamente nos dirían lo buena que es la defensa.

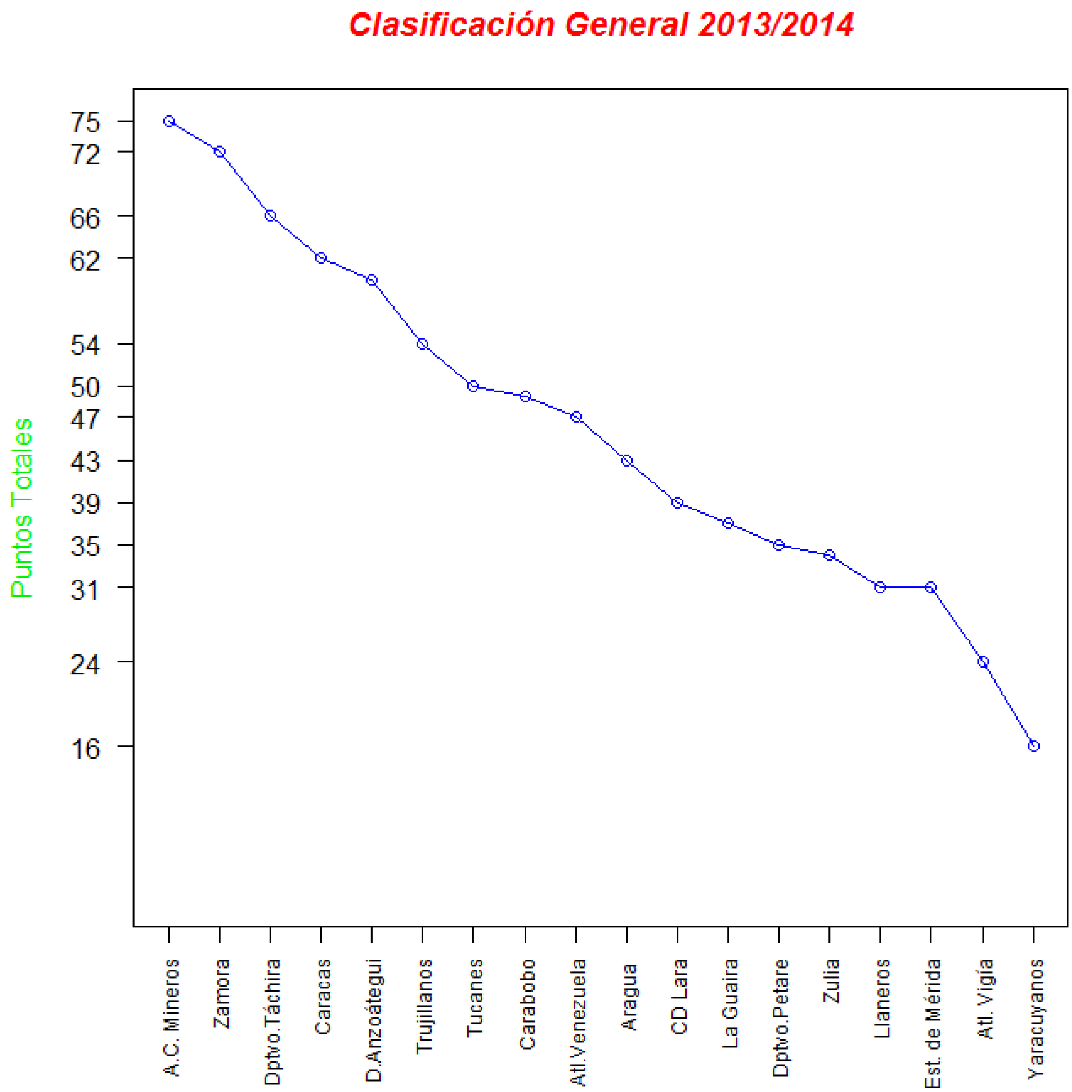

Figura 2.25: Clasificación general según la puntuación Temporada 2013/2014

En la figura 2.25 se muestra la clasificación general de la temporada 2013/2014 teniendo como base el total de puntos, en este gráfico se observa como el Zamora F.C y el A.C Mineros obtuvieron una ventaja de puntos significativa con respecto a un grupos de 3 equipos que ocupan el tercero, cuarto y quinto lugar, a saber Deportivo Táchira, Caracas y Deportivo Anzoátegui respectivamente los cuales marcan una diferencia en puntos acumulados, con respectos a un tercer grupo de equipos que se encuentra posteriormente, luego se presentan un cuarto grupo de equipos que tienen una diferencia en puntos acumulados relativamente similar y por último se encuentra un grupo de equipos en la zona baja de la tabla.

2.1.11. Resultados por temporadas

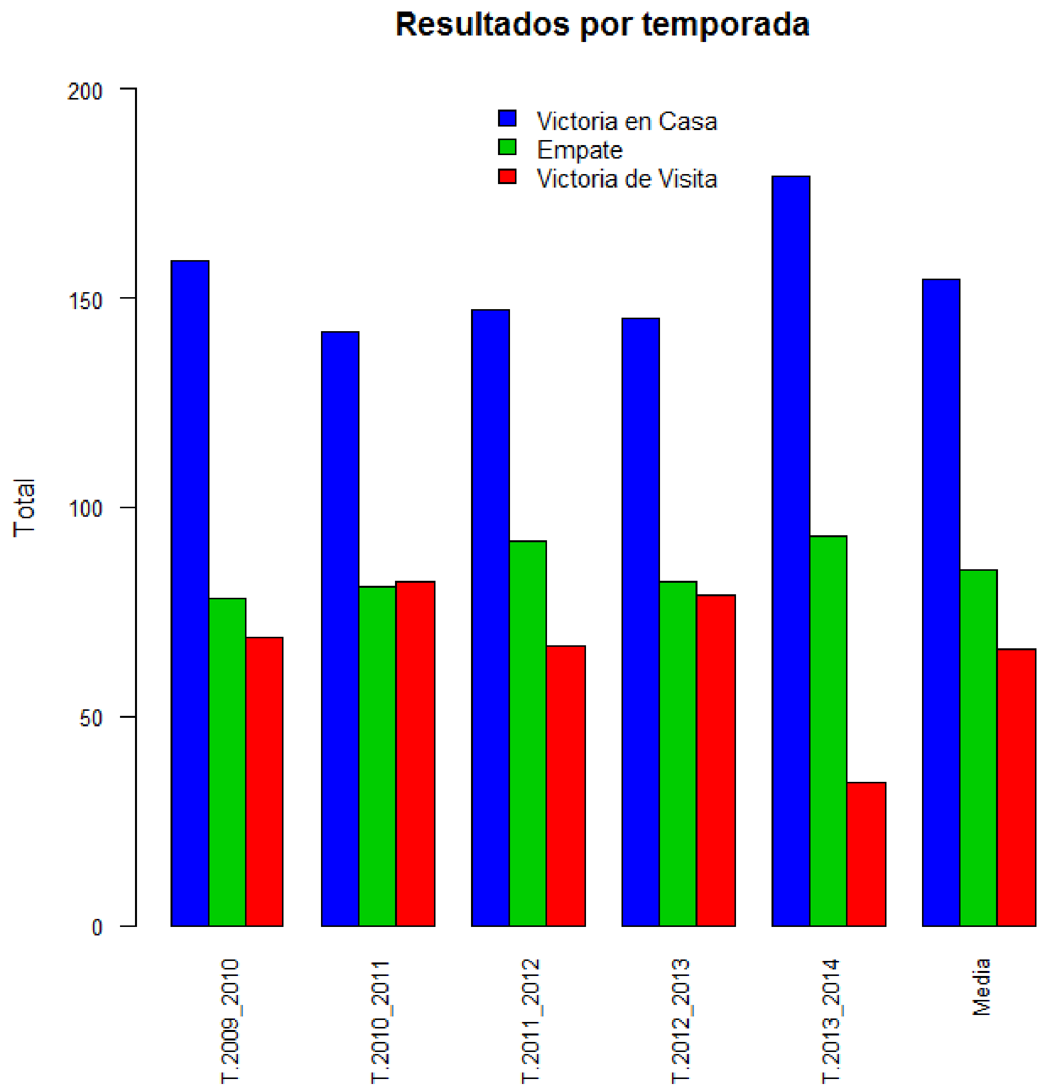

Figura 2.26: Victoria en casa, empate y victoria de visita por temporada

Observando la figura 2.26 se puede evidenciar que existe una mayor probabilidad de que los equipos ganen en casa, y lo más natural es pensar que debería haber un parámetro que influye para que estos resultados se den con mayor frecuencia, a este parámetro comúnmente se le llama en la literatura efecto casa, en la siguiente tabla se muestra el porcentaje de los tres resultados posibles de un partido durante las cinco temporadas a estudiar y la media o "valor esperado"para la siguiente temporada.

Resultado	2009/2010	2010/2011	2011/2012	2012/2013	2013/2014	Media
Victoria en casa	51,96 %	46,56 %	48,04 %	47,39 %	58,50 %	50,49 %
Empate	25,49 %	26,56 %	30,07 %	26,50 %	30,39 %	27,86 %
Victoria de visita	22,55 %	26,88 %	21,89 %	25,81 %	11,11 %	21,65 %

2.1.12. Tiempo de anotación

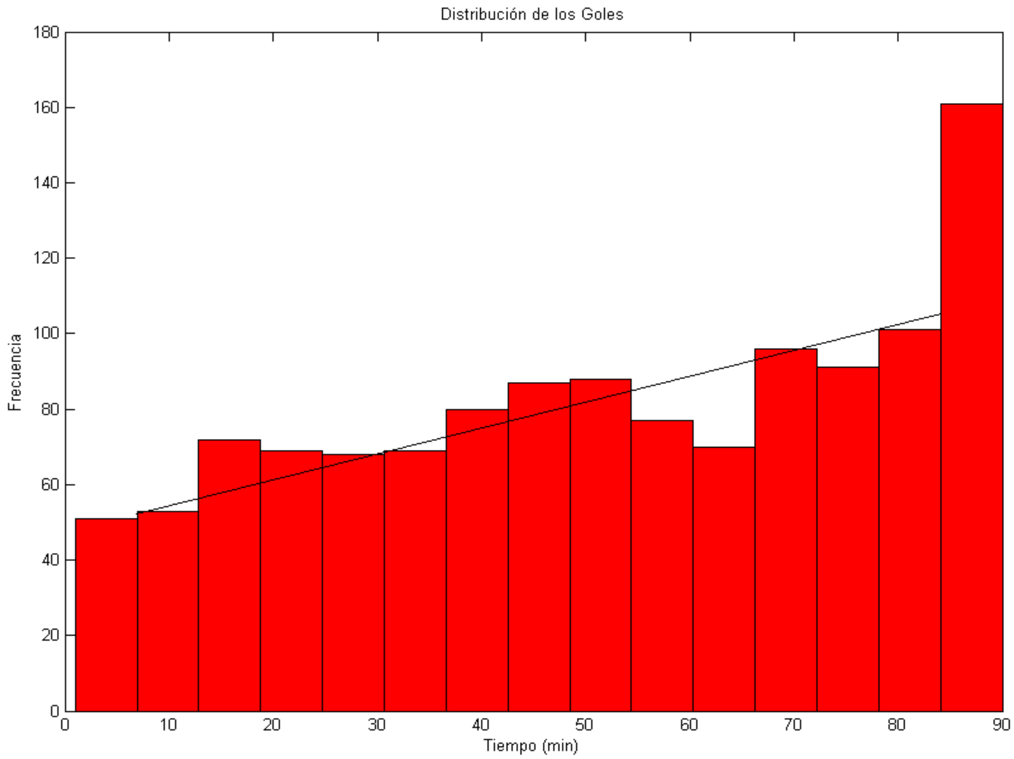

Figura 2.27: Distribución de los goles en el tiempo

Las dos características más evidentes que podemos observar en la gráfica son: En primer lugar el elevado número de goles en el final de cada medio tiempo, es decir en torno a los 45 y 90 minutos. Este aumento en el número de goles se debe al tiempo de descuento, por lo general entre 1 y 5 minutos agregados por el árbitro y también es debido a que los goles en el tiempo de descuento se contabilizan como goles en el minuto 45 o 90 para la primera y segunda mitad respectivamente. La segunda característica importante es el gran número de goles en el minuto 90, observando también el gráfico podríamos concluir que para cada partido la tasa de goles a aumentado a través del tiempo, tal vez debido al cansancio de los jugadores, sin embargo esto puede ser un efecto debido al promedio, por ejemplo, si la tasa de goles se mantuvo constante, mientras el marcador era $(0,0)$ pero aumento tan pronto como sucede un gol, entonces tomando en cuenta más partidos, en promedio podría conducir a aumentos graduales, tal como los observados en la figura 2.27. Así, dos posibles razones de la falta de homogeneidad en los minutos finales son:

1. Un aumento gradual en las tasas de goles
2. Variación debido a la dependencia en la puntuación actual.

2.1.13. Número de goles durante las temporadas

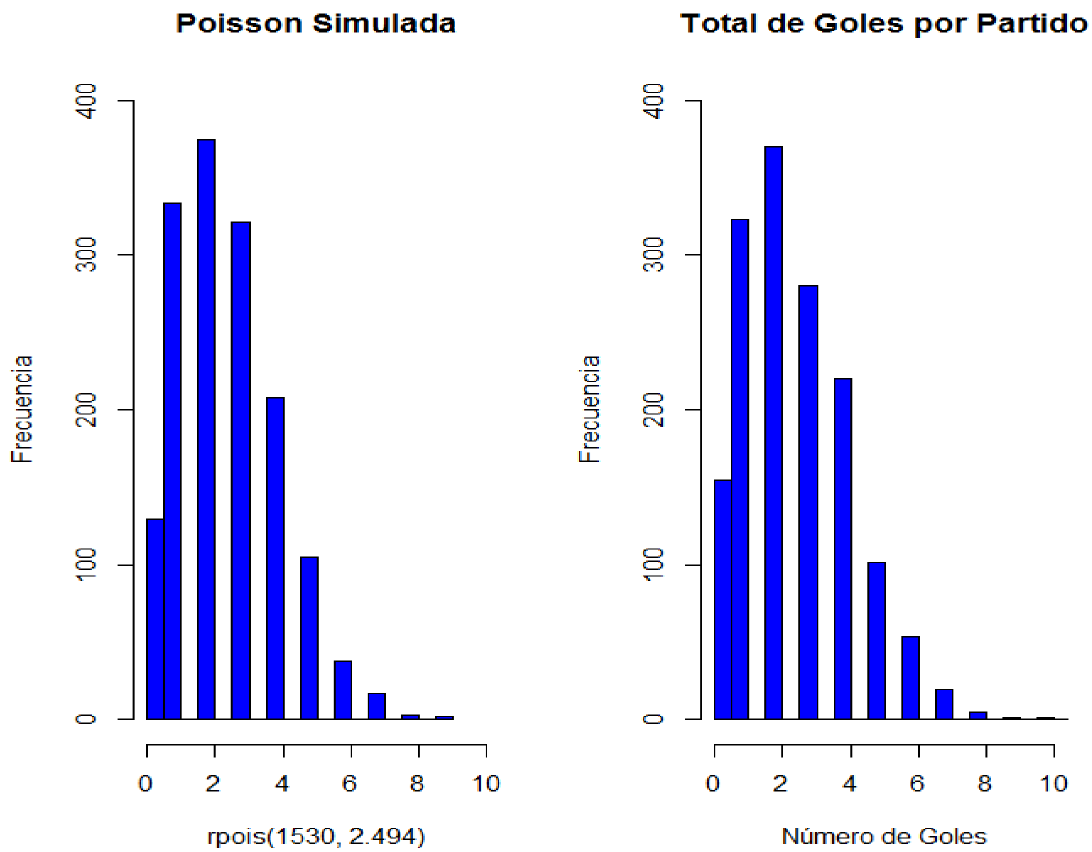

Figura 2.28: Total de goles

En primer lugar tengamos en cuenta que el número de simulaciones del gráfico de la derecha es igual a la cantidad de datos, además, la media ($\lambda = 2{,}494$) con que se hizo las simulaciones también es igual a la media de los datos, entonces observando la figura 2.28 nuestra primera conclusión seria decir que el número de goles en un partido de futbol tienen una distribución Poisson.

2.1.14. Prueba de bondad de ajuste Chi-cuadrado (χ^2)

Tomando en consideración las conclusiones anteriores, en lo que sigue nuestra meta será ver que tan cerca el número de goles en un partido de futbol se ajusta a una distribución de Poisson, esto se hará con un Test de bondad de ajuste, como la distribución de Possion y los datos de futbol son discretos, el test de bondad de ajuste Chi-cuadrado se puede utilizar, y así, verificar si los datos observados de las temporadas, 2009-2010, 2010-2011, 2011-2012, 2012-2013 y 2013-2014 siguen una distribución en particular, en nuestro caso la Distribución Poisson. Sea

- H_0 : La distribución del número de goles en un partido de futbol sigue una Distribución de Poisson.

- H_A : La distribución del número de goles en un partido de futbol no sigue una Distribución de Poisson.

Número de Goles	0	1	2	3	4	5	6	7	8	9	10
Frecuencia	155	323	370	280	220	101	53	19	5	1	1

Así, aplicando el Test mediante el programa estadístico R, y especificando el método de estimación como (ML) Máxima Verosimilitud y un nivel de significancia del 1 % obtenemos los siguientes resultados

Número	0	1	2	3	4	5	6	7	8	9	10
Observado	155	323	370	280	220	101	53	19	5	1	1
Esperado	127.2	316.3	393.0	325.7	202.4	100.6	41.7	14.8	4.6	1.3	0.3

Ahora bien, ya hecho el contraste con la distribución teórica, es posible calcular el estadístico de prueba χ^2 con los valores de la tabla anterior mediante el software R, obteniendo que $\chi^2 = 20{,}53827$ con 9 grados de libertad y que el $p - valor = 0{,}0148661$, por lo tanto no rechazamos la hipótesis nula y podemos decir que la distribución del número total de goles de las temporadas estudiadas, siguen una distribución Poisson, en consecuencia, la distribución de Poisson para modelar el número de goles es una buena opción.

Capítulo 3

Métodos y Resultados

En este capítulo se describirán los modelos estadísticos presentados en [1], [3] y [4], seguidamente, al terminar cada descripción se mostraran los resultados obtenidos.

3.1. Modelo de Poisson enfoque Karlis-Ntzoufras

El modelado de los goles en un partido de futbol adopta diversas formas, la más básica y más usada en la literatura es la distribución de Poisson. A partir de ella se han creado modelos como el propuesto en [4] que toman en cuenta otros parámetros que influyen en el desenvolvimiento de un partido, lo cual permite mejores predicciones. El desarrollo de uno de los modelos presentados en [4] se presenta seguidamente. En primer lugar, consideremos

$$X_{j,k} = n_{1,j,k} \sim Poisson(\lambda_{1,j,k}), \quad j,k = 1,...,N, \tag{3.1}$$

$$Y_{j,k} = n_{2,j,k} \sim Poisson(\lambda_{2,j,k}), \quad j,k = 1,...,N, \tag{3.2}$$

$$\ln(\lambda_{1,j,k}) = \mu + h + a_j + d_k, \quad j,k = 1,...,N, \tag{3.3}$$

$$\ln(\lambda_{2,j,k}) = \mu + a_j + d_k, \quad j,k = 1,...,N, \tag{3.4}$$

donde N es en número de equipos en la liga, $n_{i,j,k}$ y $\lambda_{i,j,k}$ son el número observado y esperado respectivamente, de los goles marcados por el equipo j, en contra del equipo k, jugando en el campo del equipo i, donde $i = 1$ indica casa e $i = 2$ indica visita; μ es un parámetro constante o intercepto, h es el parámetro de efecto casa, que comúnmente se cree que es causada por apoyo de la afición y de los familiares de los jugadores, a_j es el parámetro de rendimiento ofensivo para el equipo j y d_k es un parámetro que recoge la información defensiva del equipo k.

La principal motivación para usar 3.3 y 3.4 es el supuesto de que las habilidades ofensivas y defensivas de cada equipo cambian según se juegue en casa o de visita. Por otro lado, gracias a la formulación anterior podemos estudiar dos modelos distintos, uno para juegos en casa y otro para juegos de visita.

Además, consideremos la siguiente restricción:

$$\sum_{j=1}^{N} a_j = \sum_{k=1}^{N} d_k = 0, \tag{3.5}$$

la restricción 3.5 facilita la interpretación del modelo implicando que μ es el promedio en log-media del número de goles para los partidos de visita, h es una medida promedio del efecto en casa y está dada por la diferencia entre el promedio en log-media de las cantidad de goles como local y la cantidad de goles como visita, a_j es la habilidad ofensiva del equipo j expresada en desviaciones de μ y por último d_k es la habilidad defensiva del equipo k expresada en desviaciones de μ. Es claro que cuanto mayor es el parámetro ofensivo, mejor será el rendimiento ofensivo del equipo correspondiente, mientras que cuando menor sea el parámetro defensiva mejor será el rendimiento defensivo del equipo correspondiente

En el modelo antes planteado los autores muestran que se puede usar la independencia en la formulación del modelo, es decir, que podemos suponer que $X_{j,k}$ e $Y_{j,k}$ son independientes, así la independencia nos dice que la probabilidad de los resultados de un partido está dada por el producto de la probabilidad de los resultados del equipo de casa con la probabilidad de los resultados del equipo visitante, esto es:

$$\mathbf{P}(X_{j,k}, Y_{j,k}) = \mathbf{P}(X_{j,k})\mathbf{P}(Y_{j,k}) \tag{3.6}$$

Ahora bien, como indican las ecuaciones 3.1 y 3.2 el número de goles puede ser modelado por una distribución de Poisson, por lo cual la probabilidad del resultado de un partido puede ser calculada por:

$$\mathbf{P}(X_{j,k} = x, Y_{j,k} = y) = \frac{\lambda_{1,j,k}^{x} e^{-\lambda_{1,j,k}}}{x!} \frac{\lambda_{2,j,k}^{y} e^{-\lambda_{2,j,k}}}{y!}.$$

3.1.1. Máxima Verosimilitud

En primer lugar del modelo anterior sabemos que las medias son $\lambda_{1,j,k} = e^{\mu+h+a_j+d_k}$ y $\lambda_{2,j,k} = e^{\mu+a_j+d_k}$, lo que asegura que sean positivas. Ahora bien, para estimar los parámetros μ, h, a_j y d_k se utiliza el método de estimación de máxima verosimilitud. En efecto, supongamos que N es el número de equipos que participan en la liga, entonces se deben estimar $2N+2$ parámetros. Estas estimaciones se obtienen usando la función de verosimilitud, gracias a la independencia de los eventos, esta función se convierte en un producto de funciones de densidad, luego, para $i = 1,..,N$ juegos, con marcadores respectivos (x_i, y_i) y gracias a que la función logaritmo natural es una función estrictamente creciente, podemos tomar logaritmo natural para simplificar los cálculos. En conclusión la función de verosimilitud nos queda:

$$\mathcal{L}(a_j, d_k, \mu, h; j,k = 1,...,N | x_i, y_i; i = 1,...,N) = \ln\left(\prod_{i=1}^{N} f(x_i, y_i)\right)$$

Por lo tanto, usando 3.1 y 3.2 y considerando las ecuaciones 3.3 y 3.4 respectivamente, obtenemos que:

$$\mathcal{L}(a_j,d_k,\mu,h;j,k=1,...,N|x_i,y_i;i=1,...,N) = \sum_{i=1}^{N} \ln f(x_i,y_i)$$
$$= \sum_{i=1}^{N} \ln\left(\frac{\lambda_{1,j,k}^{x_i} e^{-\lambda_{1,j,k}}}{x_i!} \frac{\lambda_{2,j,k}^{y_i} e^{-\lambda_{2,j,k}}}{y_i!}\right)$$

$$= \sum_{i=1}^{N}\left(x_i(\mu+h+a_j+d_k)-[\mu+h+a_j+d_k+\ln(x_i!)]+y_i(\mu+a_j+d_k)-[\mu+a_j+d_k+\ln(y_i!)]\right)$$

luego, esta función de verosimilitud se maximiza usando métodos numéricos, y así se obtienen los parámetros deseados.

3.1.2. Parámetros estimados

A continuación, se muestran los parámetros estimados por el modelo ordenado de manera descendente según el mejor ataque. Para este modelo se considero los datos de la temporada 2010-2011 hasta la temporada 2013-2014.

$\mu = 0{,}066$		$h = 1{,}424$	
Posición	Equipo	Ataque a	Defensa d
1	Zamora F.C	0.484	-0.444
2	Caracas F.C	0.436	-0.527
3	A.C Mineros	0.418	-0.439
4	Dptvo. Táchira	0.416	-0.358
5	Dptvo. Anzoátegui	0.401	-0.450
6	C.D. Lara	0.390	-0.410
7	Trujillanos F.C	0.322	-0.481
8	Tucanes	0.259	-0.155
9	Llaneros	0.244	-0.258
10	Zulia F.C	0.243	-0.233
11	Carabobo F.C	0.241	-0.323
12	Aragua F.C	0.217	-0.411
13	Dptvo.Petare	0.155	-0.370
14	Dptvo. La Guaira	0.152	-0.415
15	Atl. Venezuela	0.127	-0.238
16	Est. de Mérida	0.127	-0.185
17	Portuguesa F.C	0.110	-0.275

La razón de presentar los parámetros de forma descendente según el mejor ataque se debe, a que es este parámetro el que nos da más información en cuanto a los goles de los equipos, también, como era de esperar al mirar la tabla anterior el Caracas F.C un equipo que históricamente ha estado en los primeros puesto presenta el mejor parámetro de defensa y el segundo mejor parámetro de ataque. Es importante destacar también que el equipo Mineros de Guayana que fue el equipo que ocupo la primera posición en la tabla acumulada de la temporada 2013-2014, para nosotros la temporada más reciente, presenta parámetros de defensa y ataques muy significativos con respecto a la mayoría de equipos, además, los equipos que por su rendimiento pasado se esperaban que estuvieran en la zona baja de la tabla como lo son Portuguesa F.C y Estudiantes de Mérida presentan los menores valores de parámetros de ataque, ultimo y penúltimo respectivamente y más aun, estos parámetros de ataque resultaron ser muy débiles.

Por último podemos acotar que tomando los valores de los parámetros mostrados, es posible calcular las probabilidades dadas por 3.1 y así se puede obtener la probabilidad de los distintos resultados en un partido de futbol, como es sabido, o gana visita o empate o gana casa.

3.1.3. Simulación de Clasificación

En la siguiente tabla se presenta la clasificación mas probable, obtenida mediante la simulación de una serie de juegos.

Posición	Equipo	Porcentaje
1	Caracas F.C	85.3 %
2	Zamora F.C	79.4 %
3	Trujillanos F.C	55.9 %
4	Dptvo. Anzoátegui	41.2 %
5	C.D Lara	64.7 %
6	Aragua F.C	32.4 %
6	Dptvo. Táchira	44.1 %
8	A.C Mineros	38.2 %
9	Dptvo. La Guaira	67.6 %
10	Carabobo F.C	47.1 %
11	Dptvo.Petare	64.7 %
12	Llaneros	52.9 %
13	Zulia F.C	58.8 %
14	F.C.Potuguesa	38.2 %
15	Tucanes	38.2 %
16	Atl. Venezuela	47.1 %
17	Est. de Mérida	85.3 %

Es importante señalar que el Trujillanos F.C equipo que en el torneo apertura 2014-2015, ocupo la primera posición de la tabla acumulada, posee un mayor porcentaje de quedar entre los primero lugares de la tabla acumulada.

3.2. Modelo de Regresión de Poisson

El modelo de regresión de Poisson viene directamente de la distribución de Poisson mediante el uso de la misma función de masa y el uso de variables que describen las ocurrencias de los acontecimientos. En este modelo y_i es un vector $n-dimensional$ de variables dependientes escalares que corresponden con el número de veces que el evento en cuestión se produce y x_i es una matriz de tamaño $d \times n$ de regresores lineales independientes, es decir, una variable que tiene un efecto sobre los valores y_i. Por lo tanto, como y_i depende de x_i, el modelo se estudia como la variable aleatoria $y_i \mid x_i$, es decir y_i dado x_i, que siguiendo la distribución de Poisson tienen densidad

$$f(Y_i = y_i \mid x_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i}, \quad i = 1, \ldots, n, \tag{3.7}$$

donde,

$$\mu_i = \mu(x_i, \beta) = \mathbf{E}[y_i \mid x_i], \quad i = 1, \ldots, n, \tag{3.8}$$

son los vectores de media, con $y_i \geq 0$. Además se debe tomar en cuenta la dimensiones de los términos, de lo contrario la multiplicación de la matriz x_i^T por el vector β, es decir, $x_i^T\beta$ no tendría sentido y el modelo fracasaría. Así, $x_i^T = [x_{1i}, \ldots, x_{di}]$ es una matriz $n \times d$ de los regresores $x_{j,i}$, con $j = 1, \ldots, d$. Por otro lado, β es un vector $d-dimensional$, es decir,$d \times 1$, que contiene los parámetros del modelo, entonces $x_i^T\beta$ es un vector $n-dimensional$.

En nuestro caso, n es el número de equipos y g el número de juegos, además consideremos los equipos i y j, el equipo de casa y visitante respectivamente, entonces, tenemos n parámetros de ataque, n parámetros de defensa y un parámetro de efecto casa, es decir, $d = 2n+1$ parámetros. Consideremos Y el vector con las puntuaciones de los equipos involucrados en los partidos, así, este vector tiene la siguiente forma:

$$Y^T = (y^1_{i_1,j_1}, y^1_{j_1,i_1}, \ldots, y^g_{i_g,j_g}, y^g_{j_g,i_g})$$

donde, $y^g_{i_g,j_g}$ representa el número de goles marcados por el equipo i_g contra el equipo j_g en el juego g y Y se distribuye Poisson con parámetro $e^{{x_i}^T\beta}$,

por otro lado,

$$\beta^T = (\alpha_1, \ldots, \alpha_n, \gamma_1, \ldots, \gamma_n, \delta)$$

con α_k, γ_k y δ los parámetros de ataque, defensa y efecto casa respectivamente.

3.2.1. Máxima Verosimilitud

En este modelo los vectores de media $\mu_i = e^{x_i^T\beta}$ son ahora funciones de media exponenciales para la distribución de Poisson, lo cual nos asegura que sean positivas como se requiera para datos de conteo, por esta razón este modelo también es llamado *modelo de regresión de Poisson log-lineal*. Para estimar β se utiliza el método de máxima verosimilitud, así usando la función de densidad y para un vector de observaciones independientes tenemos que la función de verosimilitud se convierte en un producto de funciones de densidades y más aun como se explico anteriormente, podemos tomar logaritmo para obtener:

$$\mathcal{L}(\beta; x_i, y_i; i = 1, \ldots, n) = \ln\left(\prod_{i=1}^{n} f(Y_i = y_i \mid x_i)\right)$$

Por lo tanto, usando 3.7 y 3.8 y sabiendo que $\mu_i = e^{x_i^T\beta}$ obtenemos:

$$\begin{aligned}\mathcal{L}(\beta; x_i, y_i; i=1,...,n) &= \sum_{i=1}^{n} \ln(f(Y_i = y_i \mid x_i)) \\ &= \sum_{i=1}^{n} \ln\left(\frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}\right) \\ &= \sum_{i=1}^{n} \left(y_i x_i^T \beta - \mu_i - \ln(y_i!)\right)\end{aligned}$$

luego, esta función de verosimilitud se maximiza usando métodos numéricos ya que al igual que ocurre en la regresión logística, las ecuaciones de verosimilitud no son lineales en los parámetros y es necesario recurrir a procedimientos iterativos para el cálculo de sus estimaciones.

3.2.2. Parámetros estimados

A continuación, se muestran los parámetros estimados por el modelo, ordenados de manera descendente según el mejor ataque. Para este modelo se considero los datos de la temporada 2010-2011 hasta la temporada 2013-2014.

$\delta = 0{,}302$			
Posición	Equipo	Ataque α	Defensa γ
1	Zamora F.C	0.107	-0.063
2	C.D. Lara	0.074	-0.111
3	Caracas F.C	0.071	0.151
4	A.C Mineros	0.000	-0.033
5	Dptvo. Anzoátegui	-0.007	-0.017
6	Dptvo. Táchira	-0.025	-0.083
7	Tucanes	-0.145	-0.444
8	Trujillanos F.C	-0.158	0.0004
9	Zulia F.C	-0.173	-0.323
10	Carabobo F.C	-0.187	-0.193
11	Llaneros	-0.221	-0.268
12	Dptvo.Petare	-0.234	-0.090
13	Aragua F.C	-0.278	-0.081
14	Dptvo. La Guaira	-0.325	-0.043
15	Est. de Mérida	-0.344	-0.454
16	Atl. Venezuela	-0.372	-0.391
17	F.C.Potuguesa	-0.449	-0.384

Como era de esperar al mirar la tabla anterior los equipos Zamora F.C, C.D Lara y Caracas F.C tienen un ataque mayor que la mayoría de los equipos, incluso que A.C Mineros el equipo que obtuvo la primera posición en la tabla acumulada de la temporada 2013-2014, pero el A.C Mineros tiene una mejor defensa que el Zamora F.C y el C.D Lara, también es importante destacar que el Trujillanos F.C presenta la segunda mejor defensa, esto podría explicar el primer lugar que ocupo durante el torneo apertura 2014-2015, y como era de esperar los equipos que frecuentas las últimas posiciones presentaron parámetros de ataque y defensa muy bajos.

3.2.3. Tabla de probabilidades

En la siguiente tabla podremos apreciar las probabilidades de los resultados entre dos equipos, hemos tomado como equipo de casa al Trujillanos F.C que es el equipo que en el torneo apertura 2014-2015 ocupó el primer lugar de la tabla; y el Caracas F.C que históricamente ha sido uno de los equipos más fuertes.

	0	1	2	3	4	5	6	7+
0	12.67	12.58	6.24	2.07	0.51	0.10	0.02	0
1	13.60	13.50	6.70	2.22	0.55	0.11	0.02	0
2	7.30	7.24	3.59	1.19	0.30	0.06	0.01	0
3	2.61	2.59	1.29	0.43	0.11	0.02	0.00	0
4	0.70	0.70	0.34	0.11	0.03	0.01	0.00	0
5	0.15	0.15	0.07	0.02	0.01	0.00	0.00	0
6	0.03	0.03	0.01	0.00	0.00	0.00	0.00	0
7+	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0

Así, el resultado más probable es Trujillanos F.C 0 Caracas F.C 1 con un porcentaje de 13,60 %, el segundo resultado con más porcentaje es Trujillanos F.C 1 Caracas F.C 1 con 13,50 % y el tercer resultado más probable es Trujillanos F.C 0 Caracas F.C 0 con 12,67 %. Ahora los porcentajes de la tabla anterior pueden ser utilizadas para calcular el porcentaje de los tres posibles resultados, a saber, victoria casa, empate o victoria visita, recordando que Trujillanos F.C es el equipo de casa. En el siguiente cuadro se muestra los porcentajes de los tres resultados posibles

Gana casa	Empate	Gana visita
32.83 %	30.22 %	36.95 %

3.3. Modelo de Poisson enfoque Dixon-Coles

El enfoque de [1] es un modelo predictivo, que utiliza solo los goles de partidos anteriores para predecir la probabilidad de resultados en un partido de futbol. Históricamente los goles se han considerado como una medida para la ofensiva y defensiva de un equipo, ya que un equipo que marca muchos goles se considera ofensivamente fuerte, pero un equipo que recibe muchos goles se considera un equipo con una defensiva muy débil, a continuación se presenta el enfoque Dixon-Coles.

Consideremos dos equipos i y j, el equipo de casa y visitante respectivamente, α_i y α_j son los parámetros de ataque de los dos equipos, y β_i y β_j, son los parámetros de defensa. Sea $X_{i,j}$ e $Y_{i,j}$ el número de goles marcados por el equipo de casa(i) y el equipo visitante(j) respectivamente y γ un parámetro de ventaja en casa, donde

$$X_{i,j} \sim Poissn(\alpha_i, \beta_j, \gamma)$$

$$Y_{i,j} \sim Poissn(\alpha_j, \beta_i)$$

de forma independiente, además $\alpha_i\beta_j\gamma$ y $\alpha_j\beta_i$ representan el valor medio para el equipo de casa y visita respectivamente.

Ahora bien, como $X_{i,j}$ e $Y_{i,j}$ son independientes por la definición 1.7 obtenemos que:

$$\mathbb{P}(X_{i,j} = x, Y_{i,j} = y) = \frac{\lambda^x e^{-\lambda}}{x!} \frac{\mu^y e^{-\mu}}{y!}$$

donde, $\lambda = \alpha_i\beta_j\gamma$ y $\mu = \alpha_j\beta_i$.

Por otro lado, según [1], el número de goles marcados por dos equipos, no son completamente independientes. Una dependencia entre los goles en casa y de visita se identificó en juegos con resultados $(0,0)$, $(1,0)$, $(0,1)$ y $(1,1)$. Por lo tanto, se considera la siguiente modificación

$$\tau_{\lambda,\mu}(x,y) = \begin{cases} 1-\lambda\mu\rho, & \text{si } x = y = 0. \\ 1+\lambda\rho, & \text{si } x = 0, y = 1. \\ 1+\mu\rho, & \text{si } x = 1, y = 0. \\ 1-\rho, & \text{si } x = y = 1. \\ 1, & \text{en otro caso.} \end{cases}$$

y la correlación ρ está determinada por,

$$\text{máx}\left(\frac{-1}{\lambda}, \frac{-1}{\mu}\right) \leq \rho \leq \text{mín}\left(\frac{1}{\lambda\mu}, 1\right).$$

Observe que si $\rho = 0$, entonces $X_{i,j}$ $Y_{i,j}$ son independientes puesto que $\tau_{\lambda,\mu}(x,y) = 1$. La ecuación final para el cálculo de probabilidades del número de goles nos queda

$$\mathbb{P}(X_{i,j} = x, Y_{i,j} = y) = \tau_{\lambda,\mu}(x,y)\frac{\lambda^x e^{-\lambda}}{x!} \frac{\mu^y e^{-\mu}}{y!},$$

pero también, este enfoque puede ser utilizado para evaluar las probabilidades de resultados de un partido de futbol, basado únicamente en datos estadísticos de partidos anteriores y así encontrar la probabilidad que exista victoria en casa, empate y victoria visitante. Por ejemplo la probabilidad de una victoria en casa es:

$$\mathbb{P}(X_{i,j} = x, Y_{i,j} = y \mid x > y) = \sum_{x>y} \tau_{\lambda,\mu}(x,y)\frac{\lambda^x e^{-\lambda}}{x!} \frac{\mu^y e^{-\mu}}{y!}. \tag{3.9}$$

3.3.1. Parámetros

Con un total de n equipos se tienen n parámetros de ataque, n parámetros de defensa, la correlación ρ y γ el parámetro de ventaja en casa. Esto quiere decir que para n equipos se tienen $2n+2$ parámetros que deben ser estimados; pero, puesto que el modelo no debe ser sobreparametrizado, se impone la siguiente restricción sobre los parámetros de ataque, que nos permitirá encontrar un conjunto único de parámetros que maximiza la verosimilitud

$$\frac{1}{n}\sum_{i=1}^{n}\alpha_i = 1. \tag{3.10}$$

Para estimar los parámetros se utiliza la función de verosimilitud, entonces, para un conjunto de juegos $k = 1,...,N$ con marcadores (x_k, y_k) se tiene la siguiente proporcionalidad, que para efectos de cálculos se toma como una igualdad:

$$L(\alpha_1,...,\alpha_n,\beta_1,...,\beta_n,\rho,\gamma) = \prod_{k=1}^{N}\tau_{\lambda_k,\mu_k}(x_k,y_k)\lambda_k^{x_k}e^{-\lambda_k}\mu_k^{y_k}e^{-\mu_k} \tag{3.11}$$

donde, $\lambda_k = \alpha_i(k)\beta_j(k)\gamma$ y $\mu_k = \alpha_j(k)\beta_i(k)$, aquí $i(k)$ y $j(k)$ denotan los índices de local y de visita en el partido k, mientras que x_k y y_k denota el número de goles marcados por cada equipo, luego, si hacemos una estimación por máxima verosimilitud de 3.11 es posible estimar los parámetros, pero en el futbol, los equipos cambian con el tiempo y se pueden producir derrotas sucesivas y rachas ganadoras debido a cambios en los jugadores, en el terreno de juego y muchos factores más que inciden en el desenvolvimiento del juego. El enfoque hasta aquí planteado no toma en cuenta estos cambios o factores, y da el mismo peso a todos los partidos, por lo que se plantea un factor de desvanecimiento en 3.11 que permite dar mayor peso a los partidos más recientes. Así, obtenemos una pseudoverosimilitud que nos queda de la siguiente forma:

$$L(\alpha_1,...,\alpha_n,\beta_1,...,\beta_n,\rho,\gamma) = \prod_{k=1}^{N}\left(\tau_{\lambda_k,\mu_k}(x_k,y_k)\lambda_k^{x_k}e^{-\lambda_k}\mu_k^{y_k}e^{-\mu_k}\right)^{\phi(t-t_k)} \tag{3.12}$$

donde, t es el tiempo en que se realiza la evaluación y t_k es el tiempo en que el partido k se jugó, tal función de desvanecimiento ϕ debe producir un factor más pequeño a medida que t aumenta de tal manera que los partido más reciente tienen un mayor peso, esto es debido a que $\lim\limits_{x\to 0} a^x = 1$, lo que implica que para partidos muy viejos con pequeños valores de t tiene un peso muy pequeño. Existen muchas maneras de elegir esta función y todas pueden ser usadas, pero los autores han examinado algunas de las posibilidades y recomiendan el uso de:

$$\phi(t) = e^{\xi t}$$

aquí, ξ es un valor pequeño mayor que cero, además ξ debe ser estimado empíricamente porque no es posible optimar 3.12 con respecto a ξ.

3.3.2. Máxima Verosimilitud

Con el fin de encontrar los parámetros óptimos la función de verosimilitud debe ser maximizada, en este caso, por métodos numéricos debido a las características de la función, el método para hacerlo es el conocido método de máxima verosimilitud, pero es posible simplificar la derivación implícita en el método, utilizando el logaritmo de la verosimilitud en lugar de la verosimilitud y obtener:

$$
\begin{aligned}
\mathcal{L}(\alpha_1,...,\alpha_n,\beta_1,...,\beta_n,\rho,\gamma) = \\
&= \ln \prod_{k=1}^{N} \left(\tau_{\lambda_k,\mu_k}(x_k,y_k)\lambda_k^{x_k} e^{-\lambda_k} \mu_k^{y_k} e^{-\mu_k} \right)^{\phi(t-t_k)} \\
&= \sum_{k=1}^{N} \phi(t-t_k) \ln \left(\tau_{\lambda_k,\mu_k}(x_k,y_k)\lambda_k^{x_k} e^{-\lambda_k} \mu_k^{y_k} e^{-\mu_k} \right) \\
&= \sum_{k=1}^{N} \phi(t-t_k)(\ln(\tau_{\lambda_k,\mu_k}(x_k,y_k)) + x_k \ln(\lambda_k) - \lambda_k + y_k \ln(\mu_k) - \mu_k)
\end{aligned}
$$

luego, para encontrar los parámetros óptimos, esta función de verosimilitud se debe maximizar.

3.3.3. Parámetros estimados

Los siguientes resultados se obtuvieron mediante la implementación de un modelo de regresión considerando el enfoque Dixon-Coles pero sin tomar en cuenta la perspectiva temporal, además

$$\lambda_s = e^{\alpha_i - \beta_j + \gamma};$$
$$\mu_s = e^{\alpha_j - \beta_i};$$

con $s = 1,...,n$ y se consideraron solo los datos de la temporada 2013-2014.

$\rho = -0{,}123$		$\gamma = 0{,}305$	
Posición	Equipo	Ataque α	Defensa β
1	Zamora F.C	1.494	1.270
2	Dptvo. Táchira	1.407	1.098
3	A.C Mineros	1.308	1.441
4	Caracas F.C	1.279	1.247
5	Trujillanos F.C	1.259	1.238
7	Dptvo. Anzoátegui	1.245	1.108
8	Carabobo F.C	1.139	1.012
9	Tucanes	1.128	1.041
10	Zulia F.C	0.987	0.796
11	Aragua F.C	0.969	1.093
12	Est. de Mérida	0.932	0.674
13	Atl. Vigía	0.924	0.496
14	C.D. Lara	0.914	1.073
15	Llaneros	0.906	0.733
16	Atl. Venezuela	0.781	1.100
17	Dptvo.Petare	0.664	1.089
18	Dptvo. La Guaira	0.553	1.138

En primer lugar observando la tabla anterior Zamora F.C, Dptvo. Táchira, A.C Mineros y Caracas F.C tienen los cuatro mejores ataques respectivamente, justamente estos equipos fueron los que estuvieron peleando por los primeros puestos durante la temporada 2013-2014, por otro lado, como sabemos el A.C Mineros fue el equipo que obtuvo la primera posición en la tabla acumulada de la temporada 2013-2014, y justamente este equipo tiene la mejor defensa de todos los equipos, también es importante destacar que el Trujillanos F.C presenta la cuarta mejor defensa no muy distinta a la del Zamora F.C y la Caracas F.C que son la segunda y tercera mejor defensa respectivamente, además, Trujillanos F.C presenta un ataque considerable en comparación con la mayoría de los equipos, todo esto podría explicar el primer lugar que ocupo en el torneo apertura 2014-2015, en segundo lugar, es importante señalar que los equipos que frecuentan las últimas posiciones tienen parámetros de ataque o defensa muy bajo.

3.3.4. Tabla de probabilidades

En la siguiente tabla podremos apreciar las probabilidades de los tres posibles resultados entre dos equipos, hemos tomado como equipo de casa al Trujillanos F.C que es el equipo que en la temporada 2014-2015 está en el primer lugar de la tabla y el Caracas F.C que históricamente ha sido uno de los equipos más fuertes.

Gana casa	Empate	Gana visita
0.4349437	0.3033571	0.2610538

Así, con un porcentaje de 43,54 % el resultado más probable es que Trujillanos F.C gane, pero no habría que descartar la posibilidad de un empate, pues el porcentaje de este resultado es 30,3 %.

Conclusiones

Los autores de los trabajos [1] y [4], son referencia mundial en lo que se refiere a modelos matemáticos deportivos, mientras que en [3] y [4] se muestra que los goles en un partido de futbol siguen la distribución de Poisson. En [1] toman esto como un hecho, asociado a esto cada uno por su parte incorporan herramientas que mejoran las predicciones. El modelo de regresión de Poisson [3] arrojo mejores resultados para predecir el marcador del partido Trujillanos F.C contra Caracas F.C, el cual se tomó para analizar las predicciones, pero no hay que menospreciar los resultados que se obtuvieron con [1], ya que este solo se implemento con los datos de la temporada 2013-2014. Más aun, las predicciones obtenidas con este modelo en cuanto a los tres posibles resultados, coincidieron con el resultado real.

Las simulaciones hechas a partir del trabajo [4], que arrojó la clasificación final de temporadas futuras se acerco mucho a la realidad, ya que el actual campeón del torneo apertura 2014-2015 fue Trujillanos F.C, y nuestros resultados pronosticaron que este equipo tenía mayores probabilidades de ocupar las primeras posiciones. Por otro lado, gracias al estudio que se hizo en el capítulo 2, se pudo tener evidencia significativa de que los datos siguen una distribución de Poisson, y así dar más validez a la implementación de los tres modelos. En este capítulo también se observó el notable efecto casa en la Primera División del Futbol Venezolano, como es de esperar cuando se estudia datos de futbol. Luego, se estimaron los parámetros de los equipos para cada modelo, pero se piensa que existen otros parámetros que influyen en el desenvolvimiento de un partido, como lo son, un parámetro de distancia, goles anotados en la primera y segunda parte, tipo de grama del campo y una última pero no menos importante, las condiciones meteorológicas, pero estudiar este último parámetro sería una tarea que consumiría mucho tiempo a la hora de recopilar estos datos. Todo esto se podría hacer teniendo cuidado de no sobreparametrizar el modelo.

En este sentido también sería interesante ampliar los resultados que tenemos hasta el momento e implementar en su totalidad el trabajo presentado en [1], considerando también más temporadas para ver su ajuste, y de seguro obtendríamos mejores resultados al considerar el enfoque dinámico. De la misma forma, probablemente se podrían elegir otros deportes de equipo y ver si resulta factible el uso estos modelos. Es importante tener en cuenta que aun tomando todas las consideraciones anteriores y lograr obtener un modelo más robusto, la estimación de los parámetros pueden no ser totalmente exacta. Esto se debe a que muchas cosas pueden pasar en el futbol que no es posible simular. Por ejemplo, cuando se obtiene la predicción de toda una temporada, una lesión grave en un jugador clave puede afectar a los equipos y estos no pueden ser incluidos en la estimación. Aunque se podría intentar predecir cada conjunto de juegos, uno a la vez, durante la temporada, con la finalidad de ir ajustando los parámetros en consonancia con los juegos, pero es una tarea muy difícil ya que no hay forma de saber la capacidad en la que estos cambios afectarán a un equipo.

Otro importante trabajo que se podría hacer con estos modelos es estudiar su bondad de ajuste, se podrían considerar las medidas de bondad de ajuste conocidas como AIC y BIC para determinar cuál de los tres modelos se ajusta mejor a los datos. Por último, es indispensable señalar que los resultados previstos en este trabajo no se deben utilizar de manera precisa para las apuestas de futuros partidos de futbol. Sin embargo dan una excelente indicación de lo que probablemente pueda suceder.

Bibliografía

[1] Dixon M, J. and Coles S.G. (1997), "Modelling Association Football Scores y Inefficiencies in the Football Betting Market".

[2] Eva, M. Garcia, Q. y María del Carmen, Iglesias, P. (2014), "Aplicación de Modelos de Regresión de Poisson Bivariados a los resultados de los partidos de la Liga Española de Futbol".

[3] James, Gardner. and Jochen, Voss. (2011), "Modeling and Simulating Football Results".

[4] Karlis, D. and Ntzoufras, I. (2000), "On Modelling Soccer Data Studen", **3**, 229-241.

[5] Karlis, D. and Ntzoufras. I. (2003), "Analysis of Sports data by using Bivariate Poisson Models", Journal of the Royal Statistical Society **52**, 301-393.

[6] Landany, S.P. and Machol, R.E (eds). (1997), "Optimal Strategies in Sport ", Amsterdam: North-Holland.

[7] Michael J. Evans. and Jeffrey S. Rosenthal. Probabilidad y Estadística. Editorial Reverté, S.A., 2005.

[8] Ridder, G. Cramer, J.S. and Hopstaken.P. (1994), "Estimating the effect of a red card in soccer", J. Am. Statist, Ass., **89**, 1124-1127.

[9] Stefani, R. and Clarkes. (1992), "Predictions and home advantage for Australian rules Footbal", J Appl.Statist, **19**, 251-261.

[10] Smith, R.L. (1988), "Forecasting records by Maximum Likelihood", J.Am.Statist Ass, **83**, 331-338.