**Tarea 5**

**Copa América**

**IN-3401-1**

**Profesor: Marcelo Olivares**

**Integrantes: Diego Bórquez**

**Gonzalo Riquelme**

# Pregunta 1

1.1

Para poder resolver esta parte de la pregunta se comenzó por añadir cinco nuevas columnas a nuestra base de datos, neutralidad, visita y la diferencia de ranking para cada equipo. Primero que nada mirando los datos de la base se puede apreciar que habrán datos relacionados a otros dependiendo si un equipo o selección se encuentra en la columna equipo1 o equipo2, para evitar esta diferenciación se duplicaron los datos y se intercambió equipo1 por equipo, de esta forma cada vez que se haga la regresión se utilizará la totalidad de datos y no solo los datos de una selección cuando está en la columna equipo1 o equipo2. La primera columna señala si los equipos juegan en una parte ajena a los dos países jugando y la segunda señala si el equipo 1 o 2 juega de visita o local (1 y 0 respectivamente). Estas columnas fueron creadas porque se encontró relevante en que parte se juega un partido de fútbol.

La tercera columna creada fue de diferencia de ranking, esta columna será la más importante del modelo dado que será el dato identificador de los equipos que se enfrentan, se utilizó la diferencia del ranking pensando que la cantidad de goles que hace un equipo depende de la posición relativa entre estos dos equipos, por lo tanto se tendrán dos columnas, una con una diferencia positiva y otra con una diferencia negativa.

Ya hechas las columnas se procedió a hacer una regresión de Poisson tomando en cuenta las tres variables ya mencionadas y la variable “importancia del partido” debido la intensidad con la que se juega un partido depende en gran cantidad de el torneo en el que se juega.

La regresión de Poisson entregó los siguientes resultados:



De estos se pueden desprender todas las variables son significantes y que su Máxima verosimilitud es -24609.055

2. Ocupando la misma regresión que se utilizó en la parte uno de la pregunta se selecciono todos los partidos de las 12 selecciones que juegan la Copa América 2015. Para poder determinar el “rate of goals” de cada selección en promedio, se utilizó la siguiente ecuación:

Las constantes utilizadas fueron las siguientes:

|  |  |
| --- | --- |
| **Importancia** | 0,023 |
| **Visita** | -0,466 |
| **Difrank1** | -0,007 |
| **neutralidad** | -0,143 |
| **constante** | 0,351 |

Haciendo este proceso para cada partido jugado por las selecciones y ponderándolos se obtuvo los siguientes resultados:

|  |  |  |
| --- | --- | --- |
|  | **prom** | **error** |
| **chile** | 1,431 | 0,465655004 |
| **bolivia** | 1,002 | 0,2971763 |
| **colombia** | 1,369 | 0,416417677 |
| **brasil** | 1,509 | 0,409901613 |
| **perú** | 1,109 | 0,298869634 |
| **jamaica** | 1,338 | 0,501029047 |
| **uruguay** | 1,488 | 0,430401028 |
| **paraguay** | 1,23 | 0,423303998 |
| **venezuela** | 1,273 | 0,482094132 |
| **argentina** | 1,468 | 0,441471057 |
| **ecuador** | 1,243 | 0,374899282 |
| **mexico** | 1,633 | 0,525934131 |

3. Aplicando la misma fórmula aplicada en la parte anterior se procedió a sacar por cada partido el lambda esperado de cada selección en cada partido. Esto nos arrojo los siguientes resultado.

|  |  |  |  |
| --- | --- | --- | --- |
| **Equipo1** | **LAMBDA1** | **LAMBDA2** | **Equipo2** |
| **Chile** | 1,655329363 | 0,878095431 | **Ecuador** |
| **Chile** | 1,565178696 | 0,804929969 | **Mexico** |
| **Chile** | 2,484322533 | 0,507123863 | **Bolivia** |
| **Ecuador** | 1,247323431 | 1,395147298 | **Mexico** |
| **Ecuador** | 1,979808257 | 0,878973966 | **Bolivia** |
| **Bolivia** | 0,831104284 | 2,093840627 | **Mexico** |

Donde lambda1 se refiere a la proyección que se tiene por el equipo 1 y lambda2 es la proyección que se tiene por el equipo 2.

4.

Para esta parte se comenzó por determinar la fórmula que se ocuparía para encontrar la probabilidad de que chile gane cada partido por si solo. Se determinó ocupar la siguiente fórmula:

En la cual lambda1 es la tasa de chile en el partido y lambda2 es la tasa del contrincante. De esta fórmula se analiza que para que chile gane tiene que ganar por 1 gol mínimo, por esto la segunda sumatoria llega hasta i-1.

Luego se programo la ecuación para sacar la probabilidad de que chile ganara en visual basic en excel.

Hecho esto para cada partido que chile jugará en la fase de grupos se saco la probabilidad de que chile gane todos los partidos multiplicando cada probabilidad sacada. Los resultados fueron los siguientes:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mexico | Ecuador | Bolivia |
| Ganar | 0,5536906 | 0,557365124 | 0,806954602 |
| ProbTotal | 0,249032508 |  |  |

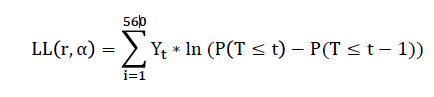
Donde se puede ver que la probabilidad de que chile gane todos los partidos es aproximadamente 25%.

# Pregunta 2

1.

En el apunte del curso de Máxima verosimilitud se indica que:

El enunciado dice que:



Sustituyendo P(T<=t) = 1 –P(T>t) y P(T<=t-1) = 1 –P(T>t-1) en la ecuación anterior y que P(T>t-1) – P(T>t) = P(T=t) se llega a que:

Elevando la ecuación anterior a la Euler se llega a que:

Utilizando la primera ecuación mencionada se puede llegar a que:

Que es la definición de máxima verosimilitud, por lo tanto es verídica la afirmación del enunciado.

2. Sacando la frecuencia de compras para cada tiempo se pudo determinar las probabilidades que se dan en la primera parte del problema para cada t y t-1. Hecho esto se pudo determinar la probabilidad dada en la parte uno, aplicándole a la sumatoria de esta el método solver de excel se llego a la siguiente solución:

|  |  |
| --- | --- |
| r | 12,04869999 |
| alfa | 592,9447848 |
| verosimilitud | -11237,96024 |

3. Aplicando el mismo proceso que se utilizó en la parte anterior, con la diferencia que en esta parte se requiere optimizar un parámetro más se llego a la siguiente solución:

|  |  |
| --- | --- |
| r | 1,726644792 |
| alfa | 375,7058829 |
| c | 1,467104649 |
| verosimilitud | -11164,8151 |

Donde se puede ver que la verosimilitud mejora, pero mejora en un 0,6% lo cual es bastante bajo para sumar un estimador más por lo cual no es muy buena la probabilidad señalada en la parte 3.