

# Análisis Estadístico para Ingeniería

Prueba Parcial Programada I (PEP1)

*Manuel Villalobos Cid*

*junio 2019*

##1.- Instrucciones Este documento contiene las actividades que constituyen la primera **Prueba Especial Programada (PEP1)**. Estas evalúan **mediante la práctica** los conocimientos adquiridos en las dos primeras unidades del curso: (1) **Estadística Descriptiva** y (2) **Probabilidad, variables aleatorias y distribución de probabilidad**.

La PEP debe ser realizada en grupos constituidos por un máximo de **dos alumnos** empleando **R/RStudio**, y el contenido debe ser redactado en un informe usando **RMarkdown**. El plazo máximo de entrega corresponde al día **12 de junio**.

No es necesario que presente un documento en el formato convencional típico de otras asignaturas: introducción, índice, antecedentes, entre otras secciones, preferentemente escriba pregunta y el desarrollo de la respuesta. **Limítese a efectuar las actividades respondiendo las preguntas planteadas, explicar cada uno de los pasos realizados en R, analizando los resultados.**

**Las tablas** deben ser construidas en base a una clase **dataframe**, y **no** está **permitido** efectuar cálculos **manuales** o el uso de **otras herramientas**.

##2- Conjunto de datos Las actividades se centran en el **contexto futbolístico**. Los datos empleados fueron descargados de **Club Soccer Predictions**. La base de datos cuenta con múltiples **indicadores** para equipos de todo el mundo, no obstante, en el transcurso de este documento se emplearán tres de ellos: **puntaje de ataque (OFF, offensive score)**, **puntaje defensivo (DEF, defensive score)**, e **índice de poderío (SPI, Soccer Power Index)**. La descripción específica del cálculo de cada una de estas variables puede ser encontrada en el sitio oficial de **ESPN**.

##3.- Actividades

###3.1 Definición de clubes La primera actividad consiste en la generación de su club deportivo (**primero en la lista**) y otros 19 equipos que formarán un campeonato ficticio de fútbol. Los nombres de los equipos, así como los **parámetros de ataque (OFF)**, **defensa (DEF)** y **poder (SPI)** se estimarán para cada grupo de manera aleatoria. Para ello efectúe lo siguiente, definiendo como entorno de trabajo la carpeta donde ubicó el archivo pdf y las fuentes (carpeta source). **Los resultados mostrados a continuación están realizados a modo de ejemplo**, debe actualizar el campo **nombre** con sus propios datos.

```
source("source/crear Equipos.R") #Cargar función específica
nombre="Datos del primer alumno" #Emplee el nombre y apellido alumno 1
equipos=crear_equipos(nombre) #Función para creación de los equipos
print(equipos) #Mostrar equipos y parámetros
```

```
##          EQUIPO  OFF  DEF  SPI
## 1      Cerro Nispero 3.19 2.24 48.50
## 2      Sao Cebolla 1.31 0.22 55.26
## 3      Audax Nuez 0.64 0.63 63.17
## 4      Real Melon 0.91 1.27 33.27
## 5  Agrupacion Rosa Mosqueta 2.21 1.85 39.72
## 6      Nuevo Acelga 1.29 2.56 78.28
## 7      Club Higo 1.85 1.34 26.86
## 8      Atletico Limon 2.18 1.99 61.24
## 9      FC Coliflor 3.04 2.10 24.29
## 10     Boca Pera 0.90 0.45 16.61
## 11     AC Naranja 1.40 1.45 51.85
## 12     DC Limon 2.23 1.52 70.52
## 13     America de Brocoli 3.49 1.97 61.82
## 14     Rio Sandia 0.54 2.07 87.49
## 15     Rayo Cereza 2.82 0.28 78.66
## 16     Villa Damasco 1.66 2.43 66.67
## 17     Leon de Membrillo 1.46 1.69 27.23
## 18     Cruz Ajo 1.69 1.85 59.84
## 19     Arsenal de Maracuya 1.03 0.60 48.37
## 20     Sport Uva 0.59 2.16 48.06
```

#### Actividad - Realice el análisis cuantitativo y visual de la frecuencia (absolutas y relativas) de cada variable estadística. Efectue un análisis de los resultados considerando que los equipos registrados en la base de datos de **Club Soccer Predictions** con mejor y peor ofensiva (OFF) son el **Real Madrid** (3.5) y el **Bastia** (0.3), los de mejor y peor defensa (DEF) son el **Bourg-Peronnas** (2.6) y el **Bastia** (0.2), y los equipos de mayor y menor poder (SPI) son el **Bayern Munich** (91.5) y el **Bastia** (11.1). - Describa las características de su equipo en comparación al resto de los que conforman la liga.

### 3.2 Desarrollo del campeonato El campeonato tiene una **modalidad de torneo largo**, donde todos los equipos jugarán todos contra todos estando de local y de visita. El torneo puede ser simulado empleando las siguientes funciones que modelan los resultados aprendiendo de los resultados de partidos reales.

```
#install.packages('gtools') # Instalar biblioteca gtools
library("gtools") # Cargar biblioteca gtools
source("source/jugar_partidos.R") # Cargar función específica
partidos=jugar_partidos(equipos) # Función para jugar todos los partidos
head(partidos,4) # Mostrar sólo los primeros partidos
```

```
##          EQ1          EQ2 OFF1 DEF1  SPI1 OFF2 DEF2  SPI2 GEQ1
## 1 AC Naranja Agrupacion Rosa Mosqueta 1.4 1.45 51.85 2.21 1.85 39.72 2
## 2 AC Naranja America de Brocoli 1.4 1.45 51.85 3.49 1.97 61.82 2
## 3 AC Naranja Arsenal de Maracuya 1.4 1.45 51.85 1.03 0.60 48.37 1
## 4 AC Naranja Atletico Limon 1.4 1.45 51.85 2.18 1.99 61.24 2
##          GEQ2
## 1 2
## 2 3
## 3 1
## 4 2
```

Las primeras dos componentes describen los nombres de equipos que disputan los partidos (filas), las siguientes

exponen sus atributos, y las dos columnas finales expresan el marcador en goles: **GEQ1** (goles del primer equipo) - **GEQ2** (goles del segundo equipo).

#### Actividades Asumiendo probabilidad uniforme:

- ¿Cuántos partidos se jugaron? Formule, explique y resuelva el problema considerando los conceptos de **combinación/permutación**.
- ¿Cuál es la probabilidad de que **su equipo** al finalizar el torneo sea el **ganador**?
- ¿Cuál es probabilidad de que **su equipo** sea campeón del torneo **3 veces consecutivas**?
- ¿Cuál es la probabilidad de que un equipo con un **puntaje de ofensiva (OFF)** sobre la media **NO** sea **ganador** del torneo?

###3.3 Finalización del campeonato El resumen del torneo puede ser visto mediante las siguiente funciones. Los campos **PPER**, **PEMP**, **PGAN** corresponden a los partidos perdidos, empatados y ganados del total de univierso de partidos jugados (**PJUG**). Los campos **GCON** y **GFAV** equivalen a los goles en favor y encontra respectivamente, mientras que el puntaje final del torneo es representado por el campo **PTJE**. Las victorias suman 3 puntos y los empates 1 punto ganando el torneo aquél equipo que obtenga el mayor puntaje (**GAN**).

```
source("source/resumen_torneo.R")           # Cargar función específica
tabla_resumen=resumen_torneo(partidos)      # Función para ver el resumen del torneo
print(tabla_resumen[c(1,5,6,7,8,9,10,11,12,13)]) # Ver campos específicos de la tabla final
```

##		EQ	PPER	PEMP	PGAN	PJUG	GCON	GFAV	PTJE	POS	COND
## 15	Rayo Cereza	0	1	37	38	25	111	112	1	GAN	
## 3	America de Brocoli	3	6	29	38	70	123	93	2	---	
## 8	Cerro Nispero	6	7	25	38	80	110	82	3	---	
## 11	DC Limon	5	9	24	38	52	79	81	4	---	
## 18	Sao Cebolla	6	8	24	38	35	63	80	5	---	
## 12	FC Coliflor	7	9	22	38	86	109	75	6	---	
## 9	Club Higo	12	9	17	38	73	73	60	7	---	
## 4	Arsenal de Maracuya	10	12	16	38	44	49	60	8	---	
## 2	Agrupacion Rosa Mosqueta	11	13	14	38	80	79	55	9	---	
## 7	Boca Pera	15	8	15	38	53	49	53	10	---	
## 5	Atletico Limon	11	14	13	38	72	73	53	11	---	
## 1	AC Naranja	11	16	11	38	61	58	49	12	---	
## 10	Cruz Ajo	14	15	9	38	69	61	42	13	---	
## 6	Audax Nuez	15	15	8	38	43	31	39	14	---	
## 13	Leon de Membrillo	19	12	7	38	81	57	33	15	---	
## 20	Villa Damasco	20	13	5	38	79	56	28	16	---	
## 16	Real Melon	22	13	3	38	66	36	22	17	---	
## 14	Nuevo Acelga	32	4	2	38	77	35	10	18	---	
## 17	Rio Sandia	30	7	1	38	65	18	10	19	DES	
## 19	Sport Uva	34	3	1	38	83	24	6	20	DES	

#### Actividades - **Grafique el histograma y la distribución normal** de los goles en contra (**GCON**) obtenido por todo el conjunto de equipos en una **misma figura y compare**. Ajuste la **presentación visual** del gráfico. - ¿Cuál es la probabilidad de que un equipo haya recibido entre 80 y 100 goles empleando el **histograma/tabla de frecuencia**? - ¿Cuál es la probabilidad de que un equipo haya recibido entre 80 y 100 goles asumiendo una **distribución normal**?

###3.4 Equipos descendidos, contrataciones y venta de jugadores Las reglas indican que al finalizar el torneo los **dos** últimos equipos deben **descender (DES)** a una división menor y que los clubes pueden iniciar la **venta y compra de jugadores**. La siguiente función muestra la actualización del estado de los equipos luego de cerrar el libro de pases.

```
source("source/actualizar_equipos.R")      # Cargar función específica
nombre="Datos del segundo alumno"         # Emplee el nombre y apellido alumno 2
equipos_nuevos=actualizar_equipos(nombre)  # Función para actualizar equipos
print(equipos_nuevos)                     # Mostrar equipos y parámetros
```

```
##          EQUIPO  OFF  DEF   SPI COND
## 1      Cerro Nispero 1.21 1.37 41.42 ANT
## 2      Sao Cebolla 0.35 0.33 59.15 ANT
## 3      Audax Nuez 3.46 1.99 48.74 ANT
## 4      Real Melon 1.31 1.39 13.24 ANT
## 5  Agrupacion Rosa Mosqueta 3.11 1.46 34.35 ANT
## 6      Nuevo Acelga 1.08 0.74 73.16 ANT
## 7      Club Higo 3.24 0.69 49.75 ANT
## 8      Atletico Limon 0.96 1.96 62.60 ANT
## 9      FC Coliflor 3.44 0.96 58.74 ANT
## 10     Boca Pera 2.95 2.11 75.04 ANT
## 11     AC Naranja 2.13 0.79 79.20 ANT
## 12     DC Limon 1.48 2.27 91.00 ANT
## 13  America de Brocoli 1.10 0.23 16.11 ANT
## 14     Puerto Pera 0.59 2.03 25.39 NUE
## 15     Rayo Cereza 2.13 0.53 32.70 ANT
## 16     Villa Damasco 0.52 2.06 69.78 ANT
## 17     Leon de Membrillo 2.00 0.90 47.55 ANT
## 18     Cruz Ajo 1.38 2.27 64.54 ANT
## 19  Arsenal de Maracuya 1.96 2.22 70.28 ANT
## 20     Agrupacion Maqui 2.32 2.37 80.02 NUE
```

El parámetro **COND** indica si un equipo es nuevo (**NUE**) o antiguo (**ANT**).

#### Actividades - Asumiendo probabilidad uniforme ¿Cuál es la probabilidad de que un equipo **NO salga campeón**, pero **tampoco** vaya a la **segunda división**? - Emplee estadística descriptiva (*biblioteca psych*), identificación de valores atípicos y coeficientes de variación para: - Caracterizar los parámetros de los equipos (OFF, DEF y SPI) **durante el torneo**. - Caracterizar los parámetros de los equipos (OFF, DEF y SPI) una vez que éstos **se actualizaron**. - Con los resultados anteriores establezca una comparación para estas tres variables durante (clase 1) y después del torneo (clase 2). Apoye la comparación empleando gráficos de cajas. - Si su equipo no descendió: compare sus características al jugar el torneo y en la actualidad.