

---

tle: "PEP 1"  
thor: "Ricardo Ramírez y Rodrigo Vidal"  
te: "6/10/2019"  
tput: pdf\_document

---

## Introducción

### Tabla de los equipos a participar del campeonato

```
1 library(knitr)
2 library(kableExtra)
3
4 source("source/crear_equipos.R") #Cargar función específica
5 nombre="Ricardo Ramírez" #Emplee el nombre y apellido alumno 1
6 equipos=crear_equipos(nombre) #Función para creación de los equipos
7
8 kable(equipos)%>% # Creación de la tabla
9 kable_styling(latex_options = "striped") # Aplicación de estilos a la tabla
```

EQUIPO	OFF	DEF	SPI
Puerto Pera	1.83	1.73	55.03
Sport Limon	1.93	0.38	22.34
Nuevo Chirimoya	1.90	1.91	36.78
Cruz Maracuya	3.48	1.69	68.10
Independiente de Platano	1.26	1.99	74.49
Audax Rosa Mosqueta	1.89	1.70	31.62
Alianza de Durazno	1.93	0.55	22.50
Liga de Ajo	1.92	1.83	72.36
Real Membrillo	0.48	1.01	24.10
AC Lucuma	0.89	0.90	74.77
A Atletico Mandarina	1.45	0.66	74.15
DC Ciruela	3.31	2.33	77.49
San Uva	1.76	0.75	42.04
New Coliflor	1.89	0.34	32.64
Cerro Papaya	1.41	0.77	25.63
Provincial Manzana	0.84	1.89	23.85
Rio Damasco	1.03	2.32	89.73
Sociedad Frutilla	3.06	2.55	86.89
Deportes Aceituna	1.19	1.01	49.73
Jaguars de Higo	2.95	0.93	37.16

**NOTA:** Según resultado obtenido por la creación de la tabla, nuestro equipo es Rayo Almendra (Primero de la tabla)

- Ataque (OFF)
- Defensa (DEF)
- Poder (SPI)

## Actividad 1

### 1. Realizar análisis cuantitativo y visual de frecuencia absoluta y frecuencia relativa

#### 1.1 Tablas de frecuencia

##### 1.1.1 Distribución de frecuencia de la característica ofensiva (OFF) de los equipos.

```
library(agricolae)
tabla_freq_off=table.freq(hist(equipos$OFF,
                               breaks = "Sturges",plot=FALSE))#Crea Tabla Freq. OFF

kable(tabla_freq_off)%>% # Crea tabla de forma gráfica
kable_styling(latex_options = "striped") # Aplicación de estilos a la tabla
```

Lower	Upper	Main	Frequency	Percentage	CF	CPF
0.0	0.5	0.25	1	5	1	5
0.5	1.0	0.75	2	10	3	15
1.0	1.5	1.25	5	25	8	40
1.5	2.0	1.75	8	40	16	80
2.0	2.5	2.25	0	0	16	80
2.5	3.0	2.75	1	5	17	85
3.0	3.5	3.25	3	15	20	100

De la tabla de frecuencia para la característica ofensiva de los equipos del campeonato, podemos observar que un 50% de los equipos posee entre un 0.0 y hasta 1.5 de puntaje, mientras que el otro 50% posee desde un 1.5 hasta un 3.5 de puntaje, por lo anterior, se podría decir que existe una distribución equitativa, que divide dos hemisferios, los equipos que tienen desde 1.5 hacia abajo y los que tienen desde un 1.5 hacia arriba (50 y 50). También se puede apreciar que el 30% de los equipos solamente tiene entre un puntaje entre 1.0 y 1.5 de ofensiva.

##### 1.1.2 Distribución de frecuencia de la característica defensiva (DEF) de los equipos.

```
library(agricolae)
tabla_freq_def=table.freq(hist(equipos$DEF,
                               breaks = "Sturges",plot=FALSE))#Crea Tabla Freq. DEF

kable(tabla_freq_def)%>% # Crea tabla de forma gráfica
kable_styling(latex_options = "striped") # Aplicación de estilos a la tabla
```

Lower	Upper	Main	Frequency	Percentage	CF	CPF
0.0	0.5	0.25	2	10	2	10
0.5	1.0	0.75	6	30	8	40
1.0	1.5	1.25	2	10	10	50
1.5	2.0	1.75	7	35	17	85
2.0	2.5	2.25	2	10	19	95
2.5	3.0	2.75	1	5	20	100

De la tabla de frecuencia para la característica defensiva (DEF) de los equipos, se puede observar que se divide la distribución de los equipos en partes iguales, es decir, el 50% de los equipos tiene entre un 0.0 hasta un 1.5 de puntaje y el otro 50% tiene desde un 1.5 hasta un 3.0 de puntaje, lo que también indica que existe un rango de 1.5 entre el máximo que alcanza cierto grupo y el mínimo que alcanza el otro grupo. También se puede observar que solamente 1 equipo se encuentra entre el puntaje máximo a obtener (entre 2.5 y 3.0) el que representa al 5% de la población del espacio muestral. Por otra parte, otro dato que se puede observar, es que hay 2 grupos que componen el 50% del total de los equipos, los que poseen entre 1.0 y 1.5 puntos (25%) y los que poseen 2.0 y 2.5 puntos (25%).

### 1.1.3 Distribución de frecuencia de SPI para los equipos

```
library(agricolae)
tabla_freq_spi=table.freq(hist(equipos$SPI,
                               breaks = "Sturges",plot=FALSE)) # Tabla de frecuencia para los SPI
kable(tabla_freq_spi)%>%
kable_styling(latex_options = "striped") # Aplicación de estilos a la tabla
```

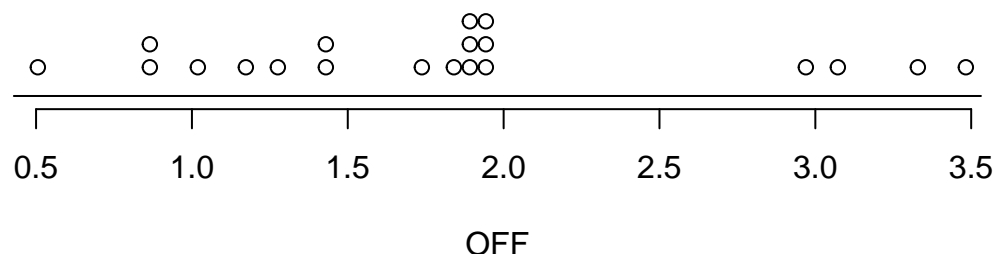
Lower	Upper	Main	Frequency	Percentage	CF	CPF
20	30	25	5	25	5	25
30	40	35	4	20	9	45
40	50	45	2	10	11	55
50	60	55	1	5	12	60
60	70	65	1	5	13	65
70	80	75	5	25	18	90
80	90	85	2	10	20	100

De la tabla de frecuencia obtenida para el puntaje de poder adquisitivo (SPI) de los equipos, se puede observar que un 35% tiene desde 50 a 60 como máximo, lo que representa a 7 de los 20 equipos del total del espacio muestral y que solamente 1 equipo se encuentra entre el rango de puntaje máximo alcanzado (90 a 100 puntos) y solamente 1 equipo se encuentra entre el rango de puntaje mínimo alcanzado (10 a 20 puntos). También se puede visualizar que un 45% de los equipos alcanza entre 10 a 50 puntos como máximo y un 55% alcanza desde 50 puntos a 100 puntos como máximo, siendo 100 el máximo a alcanzar.

## 1.2 Diagramas de puntos

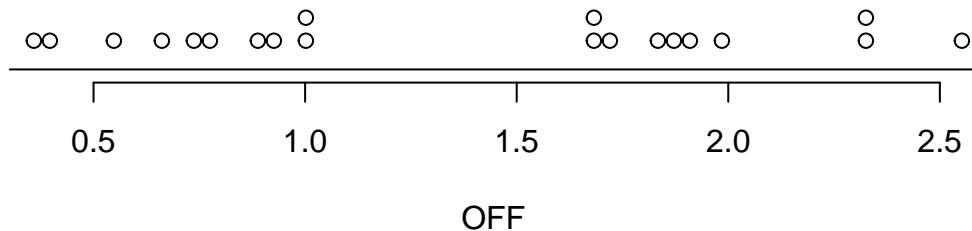
### 1.2.1 Diagrama de puntos para visualizar la distribución de los puntajes ofensivos de los equipos

```
library("BHH2")
dotPlot(equipos$OFF,xlab = "OFF")
```



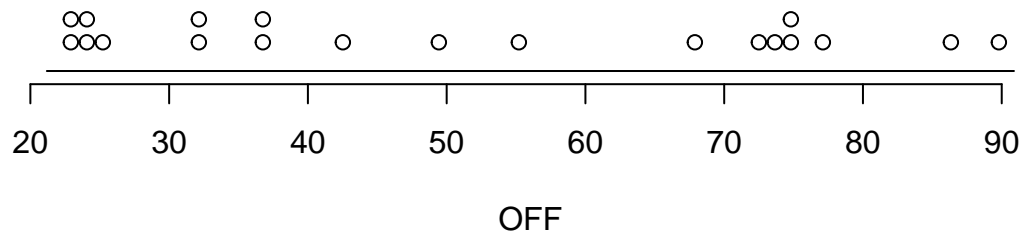
### 1.2.2 Diagrama de puntos para visualizar la distribución de los puntajes defensivos de los equipos

```
library("BHH2")
dotPlot(equipos$DEF,xlab = "OFF")
```



### 1.2.3 Diagrama de puntos para visualizar la distribución de los puntajes de poder adquisitivo de los equipos

```
library("BHH2")
dotPlot(equipos$SPI,xlab = "OFF")
```



## 1.3 Histogramas

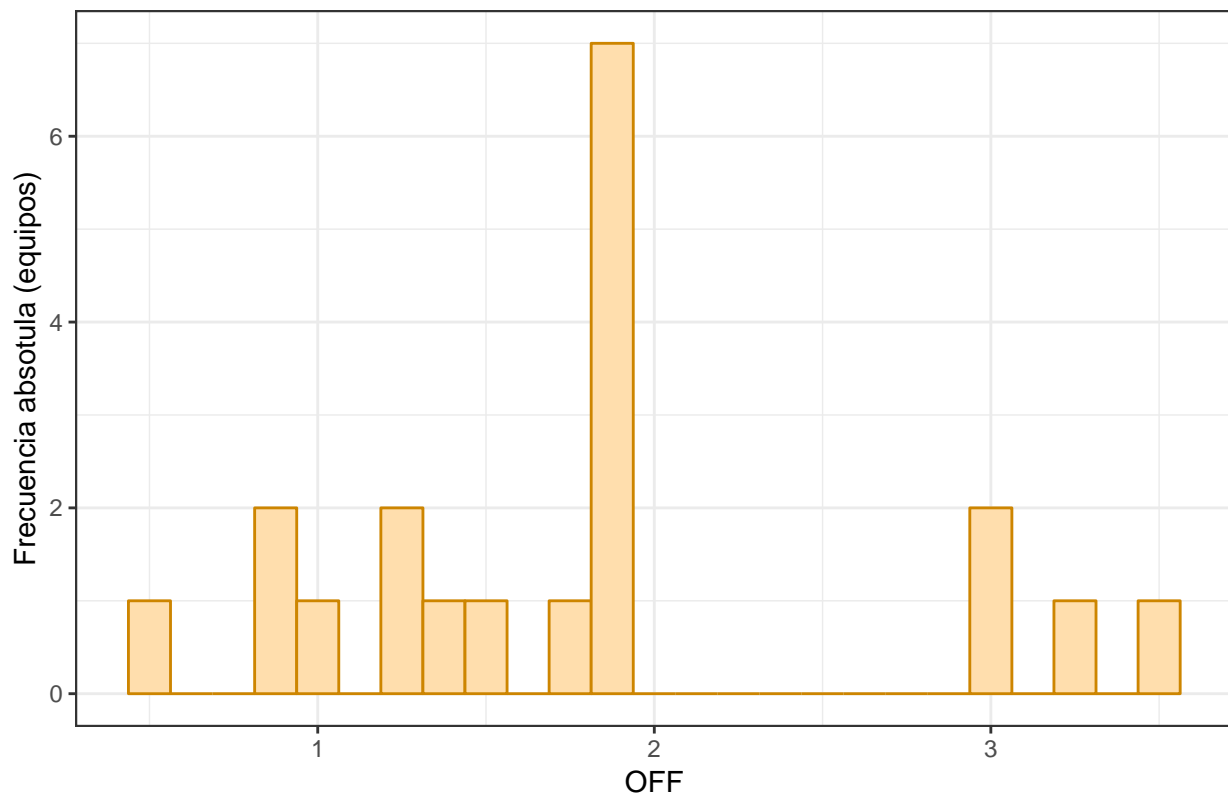
### 1.3.1 Histograma para OFF

```
library("ggplot2")
```

```
## Registered S3 methods overwritten by 'ggplot2':
##   method      from
##   [.quosures  rlang
##   c.quosures  rlang
##   print.quosures rlang
```

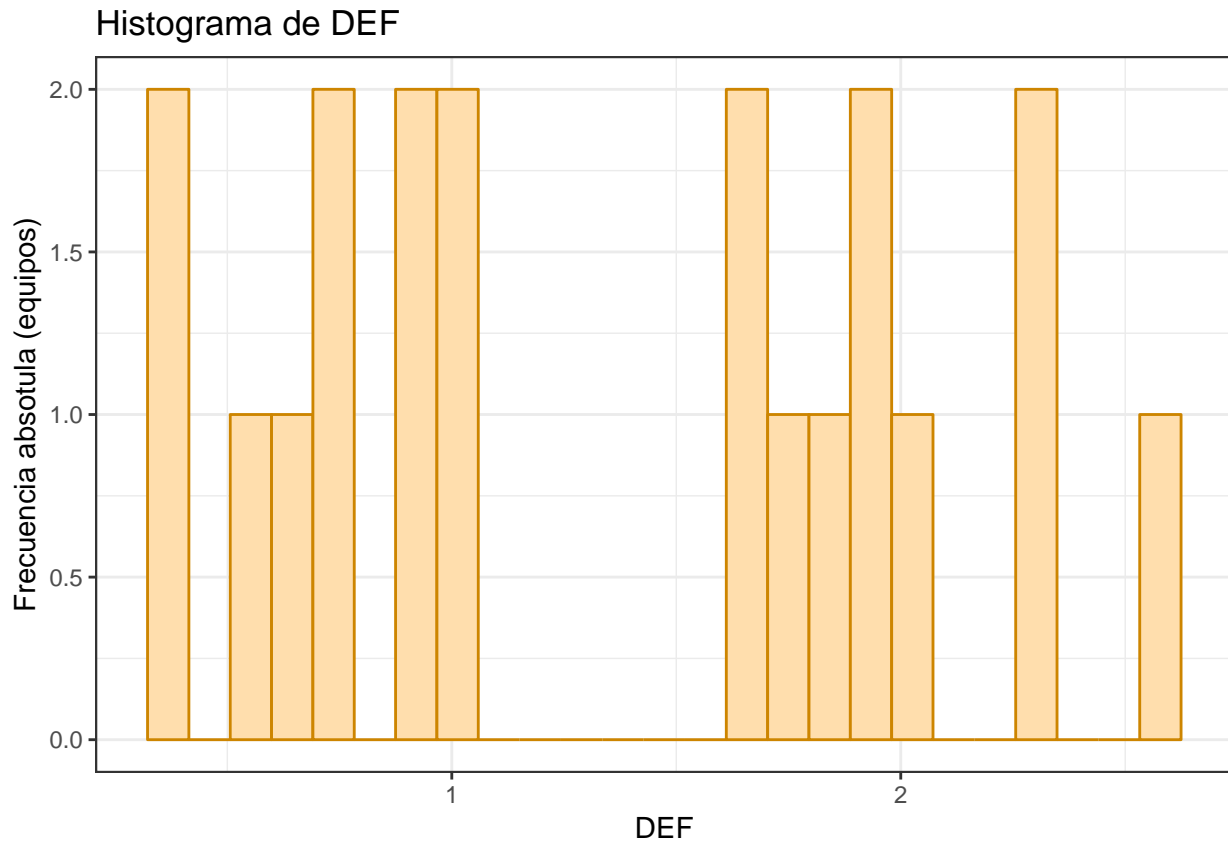
```
grafico=ggplot(equipos,aes(equipos$OFF)) # Gráfico y datos base
#Histograma (25 niveles) (colores- http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf)
grafico = grafico + geom_histogram(bins=25,fill="navajowhite",color="orange3")
grafico = grafico + theme_bw() # Visualización estándar en blanco y negro
grafico = grafico + ylab("Frecuencia absoluta (equipos)") + xlab("OFF")
grafico = grafico + ggtitle("Histograma de OFF")
plot(grafico)
```

Histograma de OFF



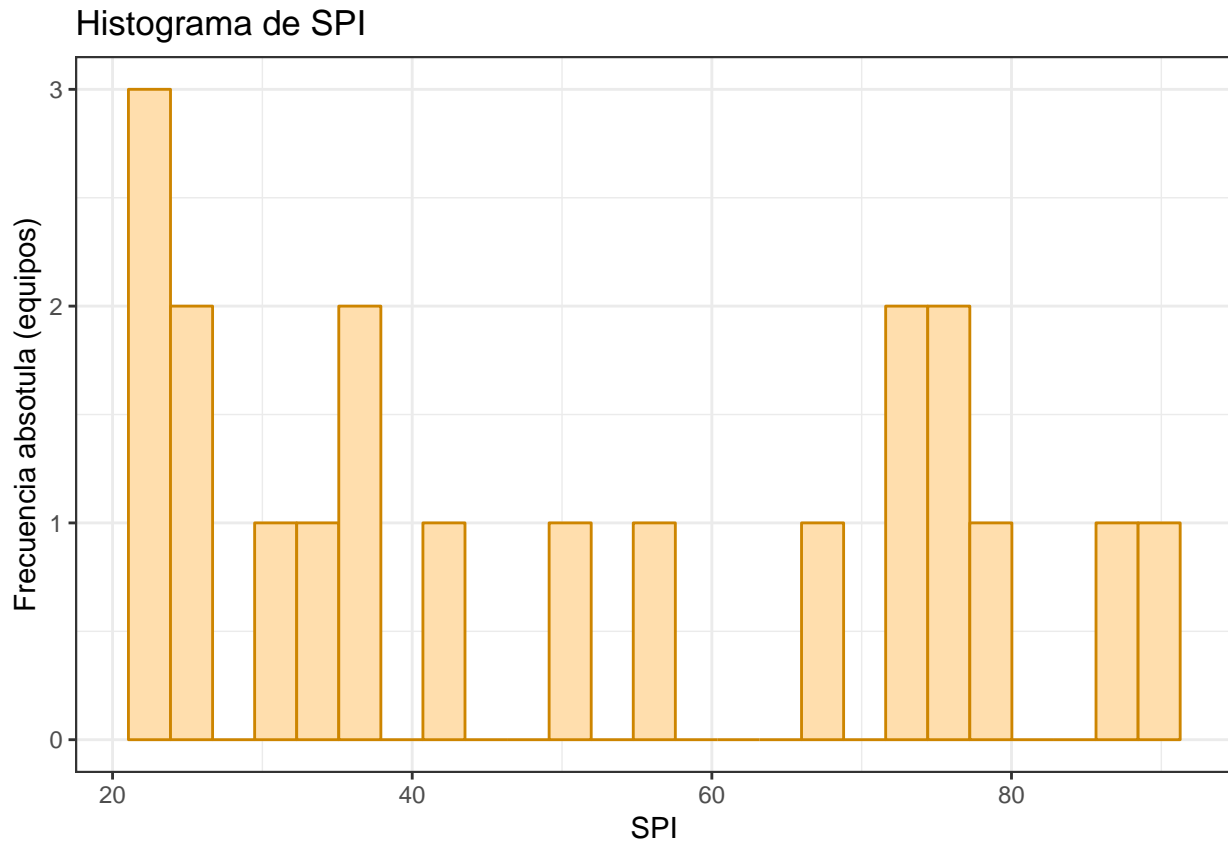
### 1.3.2 Histograma para DEF

```
library("ggplot2")
grafico=ggplot(equipos,aes(equipos$DEF)) # Gráfico y datos base
#Histograma (25 niveles) (colores- http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf)
grafico = grafico + geom_histogram(bins=25,fill="navajowhite",color="orange3")
grafico = grafico + theme_bw() # Visualización estándar en blanco y negro
grafico = grafico + ylab("Frecuencia absoluta (equipos)") + xlab("DEF")
grafico = grafico + ggtitle("Histograma de DEF")
plot(grafico)
```



### 1.3.3 Histograma para SPI

```
library("ggplot2")
grafico=ggplot(equipos,aes(equipos$SPI)) # Gráfico y datos base
#Histograma (25 niveles) (colores- http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf)
grafico = grafico + geom_histogram(bins=25,fill="navajowhite",color="orange3")
grafico = grafico + theme_bw() # Visualización estándar en blanco y negro
grafico = grafico + ylab("Frecuencia absoluta (equipos)") + xlab("SPI")
grafico = grafico + ggtitle("Histograma de SPI")
plot(grafico)
```



## 1.4 Análisis y comparativa entre equipos de mejor y peor rendimiento y nuestro equipo (Rayo Almendra)

### 1.4.1 Mínimo y máximo de puntaje ofensivo de los equipos

```
rango=range(equipos$OFF) # Rango mínimo y máximo de OFF de los equipos
rango
```

```
## [1] 0.48 3.48
```

El puntaje que posee el equipo con más bajo rendimiento en ofensiva es de 0.48 puntos, mientras que el equipo que posee el más alto puntaje de ofensiva es 3.43. El equipo Rayo Almendra tiene un puntaje ofensivo de 0.54, lo que lo aleja solamente 0.06 puntos del más bajo y una diferencia de 2.89 del equipo con mejor rendimiento ofensivo, lo que posiciona al Rayo Almendra entre el 20% con más bajo puntaje de la tabla.

### 1.4.2 Mínimo y máximo de puntaje defensivo de los equipos

```
rango=range(equipos$DEF) # Rango mínimo y máximo de DEF de los equipos
rango
```

```
## [1] 0.34 2.55
```

El puntaje mínimo obtenido para la característica defensiva de los equipos es de 0.26, mientras el máximo es de 2.51 puntos. El equipo Rayo Almendra, para esta categoría tiene un puntaje de 1.67, alejándolo 1.41 puntos del mínimo obtenido y 0.84 del puntaje máximo alcanzado, lo que posiciona a nuestro equipo entre el 50% con mejor rendimiento defensivo del campeonato.

### 1.4.3 Mínimo y máximo de SPI de los equipos

```
rango=range(equipos$SPI) # Rango mínimo y máximo de SPI de los equipos
rango
```

```
## [1] 22.34 89.73
```

El equipo que posee el mínimo SPI tiene 15.52 puntos, mientras que el que posee el máximo puntaje tiene 90.99 puntos. El equipo Rayo Almendra tiene para esta categoría 52.06 puntos, lo que lo aleja 36.54 puntos del peor y 38.93 puntos del mejor. Lo que posiciona a nuestro equipo entre el 55% de mejor puntaje en SPI del campeonato y perteneciendo su puntaje a un 35% del total del espacio muestral.

Generalmente la diferencia para las categorías entre los peores y mejores rendimientos es la siguiente:

- El peor equipo tiene una diferencia de **2.95** puntos del equipo con mejor ofensiva.
- El peor equipo tiene una diferencia de **2.25** puntos del equipo con mejor defensiva.
- El peor equipo tiene una diferencia de **75.47** puntos del equipo con mejor poder adquisitivo.

## 1.5 Análisis de equipos aplicando medidas de centralidad (Media)

### 1.5.1 Media de OFF

```
1 mediana_off = median(equipos$OFF)
2 print(paste(mediana_off))
```

```
## [1] "1.86"
```

El equipo Rayo Almendra tiene un puntaje de 0.54, lo que lo posiciona por debajo de la puntuación media para la característica ofensiva.

### Equipos por sobre la Media de (OFF)

```
1 library("dplyr", warn.conflicts = F)
2 filtro1=filter(equipos, OFF > mediana_off)
3 filtro1=filtro1[order(filtro1$OFF , decreasing = TRUE),]
4 #kable(filtro1)
5 kable(filtro1)%>%
6 kable_styling(latex_options = "striped") # Aplicación de estilos a la tabla
```



	EQUIPO	OFF	DEF	SPI
3	Cruz Maracuya	3.48	1.69	68.10
7	DC Ciruela	3.31	2.33	77.49
9	Sociedad Frutilla	3.06	2.55	86.89
10	Jaguars de Higo	2.95	0.93	37.16
1	Sport Limon	1.93	0.38	22.34
5	Alianza de Durazno	1.93	0.55	22.50
6	Liga de Ajo	1.92	1.83	72.36
2	Nuevo Chirimoya	1.90	1.91	36.78
4	Audax Rosa Mosqueta	1.89	1.70	31.62
8	New Coliflor	1.89	0.34	32.64

```
1 paste(count(filtro1)$n, "equipos por sobre la media OFF de un total de", count(equipos)$n )
```

```
## [1] "10 equipos por sobre la media OFF de un total de 20"
```

## 1.5.2 Media de DEF

```
1 mediana_def = median(equipos$DEF)
2 print(paste(mediana_def))
```

```
## [1] "1.35"
```

El equipo Rayo Almendra para la puntuación de defensiva tiene un puntaje de 1.67, lo que lo posiciona por sobre el puntaje medio (1.475).

## Equipos por sobre la Media de (DEF)

```
1 library("dplyr", warn.conflicts = F)
2 filtro2=filter(equipos, DEF > mediana_def)
3 filtro2=filtro2[order(filtro2$DEF, decreasing = TRUE),]
4 #kable(filtro2)
5 kable(filtro2)%>%
6 kable_styling(latex_options = "striped") # Aplicación de estilos a la tabla
```

	EQUIPO	OFF	DEF	SPI
10	Sociedad Frutilla	3.06	2.55	86.89
7	DC Ciruela	3.31	2.33	77.49
9	Rio Damasco	1.03	2.32	89.73
4	Independiente de Platano	1.26	1.99	74.49
2	Nuevo Chirimoya	1.90	1.91	36.78
8	Provincial Manzana	0.84	1.89	23.85
6	Liga de Ajo	1.92	1.83	72.36
1	Puerto Pera	1.83	1.73	55.03
5	Audax Rosa Mosqueta	1.89	1.70	31.62
3	Cruz Maracuya	3.48	1.69	68.10

```
1 paste(count(filtro2)$n, "equipos por sobre la media DEF de un total de", count(equipos)$n )
```

```
## [1] "10 equipos por sobre la media DEF de un total de 20"
```

Se puede observar al equipo en la posición nº9 de la tabla de equipos sobre la media en DEF.

## Media de SPI

```
1 mediana_spi = median(equipos$SPI)
2 print(paste(mediana_spi))
```

```
## [1] "45.885"
```

El equipo Rayo Almendra tiene un puntaje de 52.06 para SPI, lo que lo posiciona sobre la media la cual es 51.11.

## Equipos por sobre la Media de (SPI)

```
1 library("dplyr",warn.conflicts = F)
2 filtro3=filter(equipos, SPI > mediana_spi)
3 filtro3=filtro3[order(filtro3$SPI, decreasing = TRUE),]
4 #kable(filtro3)
5 kable(filtro3)%>%
6 kable_styling(latex_options = "striped") # Aplicación de estilos a la tabla
```

	EQUIPO	OFF	DEF	SPI
8	Rio Damasco	1.03	2.32	89.73
9	Sociedad Frutilla	3.06	2.55	86.89
7	DC Ciruela	3.31	2.33	77.49
5	AC Lucuma	0.89	0.90	74.77
3	Independiente de Platano	1.26	1.99	74.49
6	Atletico Mandarina	1.45	0.66	74.15
4	Liga de Ajo	1.92	1.83	72.36
2	Cruz Maracuya	3.48	1.69	68.10
1	Puerto Pera	1.83	1.73	55.03
10	Deportes Aceituna	1.19	1.01	49.73

```
1 paste(count(filtro3)$n, "equipos por sobre la media SPI de un total de", count(equipos)$n )
```

```
## [1] "10 equipos por sobre la media SPI de un total de 20"
```

El equipo Rayo Almendra queda en la posición nº7 de la tabla de los equipos sobre la media de SPI.

## Equipos por sobre la Media de (OFF), Media de (DEF) y Media de (SPI)

```

1 filtro_final=filter(equipos, OFF > mediana_off)
2 filtro_final=filter(filtro_final, DEF > mediana_def)
3 filtro_final=filter(filtro_final, SPI > mediana_spi)
4 #kable(filtro_final)
5 kable(filtro_final)%>%
6 kable_styling(latex_options = "striped") # Aplicación de estilos a la tabla

```

EQUIPO	OFF	DEF	SPI
Cruz Maracuya	3.48	1.69	68.10
Liga de Ajo	1.92	1.83	72.36
DC Ciruela	3.31	2.33	77.49
Sociedad Frutilla	3.06	2.55	86.89

```

1 paste(count(filtro_final)$n, "equipos por sobre todas las medias de un total de", count(equipos)$n )

```

```
## [1] "4 equipos por sobre todas las medias de un total de 20"
```

```

#install.packages('gtools')
library("gtools")
source("source/jugar_partidos.R")
partidos=jugar_partidos(equipos)
head(partidos,40) # Muestra los primeros 20 partidos

```

```

##          EQ1          EQ2 OFF1 DEF1  SPI1 OFF2 DEF2
## 1      AC Lucuma  Alianza de Durazno 0.89 0.90 74.77 1.93 0.55
## 2      AC Lucuma  Atletico Mandarina 0.89 0.90 74.77 1.45 0.66
## 3      AC Lucuma  Audax Rosa Mosqueta 0.89 0.90 74.77 1.89 1.70
## 4      AC Lucuma  Cerro Papaya 0.89 0.90 74.77 1.41 0.77
## 5      AC Lucuma  Cruz Maracuya 0.89 0.90 74.77 3.48 1.69
## 6      AC Lucuma  DC Ciruela 0.89 0.90 74.77 3.31 2.33
## 7      AC Lucuma  Deportes Aceituna 0.89 0.90 74.77 1.19 1.01
## 8      AC Lucuma  Independiente de Platano 0.89 0.90 74.77 1.26 1.99
## 9      AC Lucuma  Jaguares de Higo 0.89 0.90 74.77 2.95 0.93
## 10     AC Lucuma  Liga de Ajo 0.89 0.90 74.77 1.92 1.83
## 11     AC Lucuma  New Coliflor 0.89 0.90 74.77 1.89 0.34
## 12     AC Lucuma  Nuevo Chirimoya 0.89 0.90 74.77 1.90 1.91
## 13     AC Lucuma  Provincial Manzana 0.89 0.90 74.77 0.84 1.89
## 14     AC Lucuma  Puerto Pera 0.89 0.90 74.77 1.83 1.73
## 15     AC Lucuma  Real Membrillo 0.89 0.90 74.77 0.48 1.01
## 16     AC Lucuma  Rio Damasco 0.89 0.90 74.77 1.03 2.32
## 17     AC Lucuma  San Uva 0.89 0.90 74.77 1.76 0.75
## 18     AC Lucuma  Sociedad Frutilla 0.89 0.90 74.77 3.06 2.55
## 19     AC Lucuma  Sport Limon 0.89 0.90 74.77 1.93 0.38
## 20 Alianza de Durazno  AC Lucuma 1.93 0.55 22.50 0.89 0.90
## 21 Alianza de Durazno  Atletico Mandarina 1.93 0.55 22.50 1.45 0.66
## 22 Alianza de Durazno  Audax Rosa Mosqueta 1.93 0.55 22.50 1.89 1.70
## 23 Alianza de Durazno  Cerro Papaya 1.93 0.55 22.50 1.41 0.77
## 24 Alianza de Durazno  Cruz Maracuya 1.93 0.55 22.50 3.48 1.69
## 25 Alianza de Durazno  DC Ciruela 1.93 0.55 22.50 3.31 2.33
## 26 Alianza de Durazno  Deportes Aceituna 1.93 0.55 22.50 1.19 1.01

```

## 27	Alianza de Durazno	Independiente de	Platano	1.93	0.55	22.50	1.26	1.99
## 28	Alianza de Durazno	Jaguars de Higo		1.93	0.55	22.50	2.95	0.93
## 29	Alianza de Durazno	Liga de Ajo		1.93	0.55	22.50	1.92	1.83
## 30	Alianza de Durazno	New Coliflor		1.93	0.55	22.50	1.89	0.34
## 31	Alianza de Durazno	Nuevo Chirimoya		1.93	0.55	22.50	1.90	1.91
## 32	Alianza de Durazno	Provincial Manzana		1.93	0.55	22.50	0.84	1.89
## 33	Alianza de Durazno	Puerto Pera		1.93	0.55	22.50	1.83	1.73
## 34	Alianza de Durazno	Real Membrillo		1.93	0.55	22.50	0.48	1.01
## 35	Alianza de Durazno	Rio Damasco		1.93	0.55	22.50	1.03	2.32
## 36	Alianza de Durazno	San Uva		1.93	0.55	22.50	1.76	0.75
## 37	Alianza de Durazno	Sociedad Frutilla		1.93	0.55	22.50	3.06	2.55
## 38	Alianza de Durazno	Sport Limon		1.93	0.55	22.50	1.93	0.38
## 39	Atletico Mandarina	AC Lucuma		1.45	0.66	74.15	0.89	0.90
## 40	Atletico Mandarina	Alianza de Durazno		1.45	0.66	74.15	1.93	0.55
##	SPI2	GEQ1	GEQ2					
## 1	22.50	1	1					
## 2	74.15	0	1					
## 3	31.62	1	1					
## 4	25.63	1	1					
## 5	68.10	1	2					
## 6	77.49	1	2					
## 7	49.73	1	1					
## 8	74.49	1	1					
## 9	37.16	1	2					
## 10	72.36	1	1					
## 11	32.64	1	1					
## 12	36.78	1	1					
## 13	23.85	2	0					
## 14	55.03	1	1					
## 15	24.10	1	0					
## 16	89.73	1	0					
## 17	42.04	1	1					
## 18	86.89	1	2					
## 19	22.34	1	1					
## 20	74.77	2	0					
## 21	74.15	2	1					
## 22	31.62	3	1					
## 23	25.63	3	1					
## 24	68.10	3	2					
## 25	77.49	3	2					
## 26	49.73	2	1					
## 27	74.49	2	1					
## 28	37.16	3	2					
## 29	72.36	2	1					
## 30	32.64	2	1					
## 31	36.78	3	1					
## 32	23.85	3	0					
## 33	55.03	3	1					
## 34	24.10	3	0					
## 35	89.73	2	0					
## 36	42.04	2	1					
## 37	86.89	3	2					
## 38	22.34	2	1					
## 39	74.77	1	0					

```
## 40 22.50      2      1
```

## 1.6 Actividades Asumiendo probabilidad uniforme

¿Cuántos partidos se jugaron? Formule, explique y resuelva el problema considerando los conceptos de combinación/permutación.

Ya que el orde de los partidos si importa, ya que se juega uno de ida y uno de vuelta, es decir, 2 partidos por par de equipos, se utiliza el concepto de permutación, en donde utilizando la formula

$${}_n P_k = \frac{n!}{(n-k)!}$$

en donde N es la cantidad total de equipos (20) y n es la cantidad de partidos para dos equipos, se reflejará el resultado de el total de permutaciones del campeonato.

```
library(gtools)
n = 2 # Cantidad de partidos a jugar por par de equipos
N = 20 # Cantidad de equipos
permutaciones= nrow(permutations(N,n))
print(permutaciones)
```

```
## [1] 380
```

Analizando los resultados, la cantidad de partidos a jugar fueron 380, considerando que por cada partido existe uno de ida y uno de vuelta.

Cuál es la probabilidad de que su equipo al finalizar el torneo sea el ganador?

por distribución de bernoulli tenemos  $x$  = cantidad de veces que existe un ganador(1); ya que existen 20 equipos, el espacio muestral es de 20. Se considera éxito ganar, entonces la probabilidad es  $1/20$ . Se considera fracaso no ganar entonces  $q=1-P = 1-(1/20) = 19/20$ .

La probabilidad de que el equipo sea ganador es  $x = 1$  entonces  $P(x=1) = (1/20)^1 * (19/20)^{(1-1)} = 0.05$ , lo que nos indica que tenemos una probabilidad de 0.05% de ganar.