



ESTADÍSTICA COMPUTACIONAL - ILI-280

LABORATORIO ESTADÍSTICA COMPUTACIONAL 1

Estadística Descriptiva y Probabilidades

Autores:

Víctor Franchis Rojas (201.273.566-1) // Daniel Morales Miguel
(201.273.596-3)

Abril de 2015

1 Poblaciones versus Muestras

1.1 Datos: Estadística en Mundiales de Fútbol

1.1.1

Utilizando el total de goles, es decir, los goles convertidos por ambos equipos en un partido, se construirán tres tablas de frecuencia con un número de clases igual a 5, 15 y 30. Para ello se utilizar la librería *L^AT_EXtable*.

Total de Goles - K = 5

	Limites	Marca	Mediana	Frec. Absoluta
K_1	[0, 2.4]	1	1	354
K_2	(2.4 , 4.8]	3	3	274
K_3	(4.8 , 7.2]	5	5	131
K_4	(7.2 , 9.6]	8	8	21
K_4	(9.6 , 12]	11	11	5

Total de Goles - K = 15

	Limites	Marca	Mediana	Frec. Absoluta
K_1	[0, 0.8]	0	0	54
K_2	(0.8 , 1.6]	1	1	149
K_3	(1.6 , 2.4]	2	2	151
K_4	(2.4 , 3.2]	3	3	163
K_5	(3.2 , 4]	4	4	111
K_6	(4 , 4.8]	NA	NA	0
K_7	(4.8 , 5.6]	5	5	74
K_8	(5.6 , 6.4]	6	6	30
K_9	(6.4 , 7.2]	7	7	27
K_{10}	(7.2 , 8]	8	8	12
K_{11}	(8 , 8.8]	NA	NA	0
K_{12}	(8.8 , 9.6]	9	9	9
K_{13}	(9.6 , 10.4]	10	10	1
K_{14}	(10.4 , 11.2]	11	11	3
K_{15}	(11.2 , 12]	12	12	1

Se puede observar en todos los casos que los datos están fuertemente concentrados en las primeras clases, es decir, la mayoría de los partidos finalizaron con 0, 1, 2 o 3 goles. Además, encontramos valores NA y frecuencias

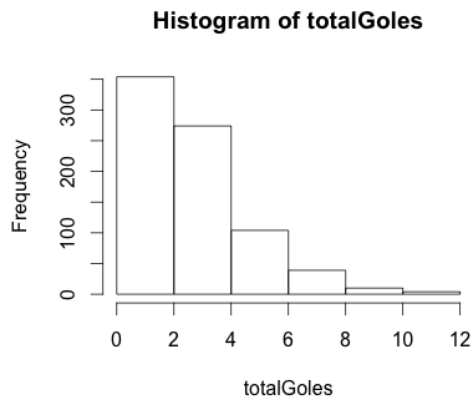
0 en varios de las clases debido a que no existen goles con números no naturales, por ejemplo, es absurdo hablar de 1.4 goles.

Finalmente, podemos concluir que de nuestra muestra de 785 partidos disputados, la cantidad de goles se encuentra generalmente entre los 0 y los 3 goles anotados.

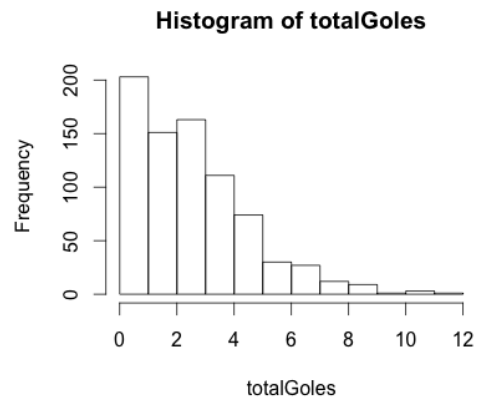
1.1.2

Se construirán 4 histogramas con número de clases iguales a 5, 10, 30, y 100 que mostrará la distribución de frecuencias para la variable en estudio. Esto se realizó utilizando al funció `hist` de R.

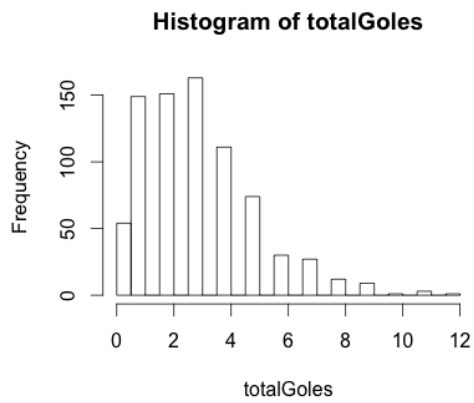
Así como nos fue posible observar mediante las tablas de la subsección anterior, la distribución de los goles, independiente de la cantidad de clases, está centrada en el comienzo del histograma. Además, notamos del primer histograma, que la mayoría de los goles se centran en las clases que van entre los valores 0 y 3, pero a medida que aumentamos el número de clases, es fácil apreciar que la moda está en los 3 goles. Nuevamente, notamos frecuencia 0 en las clases que sólo contienen valores no enteros.



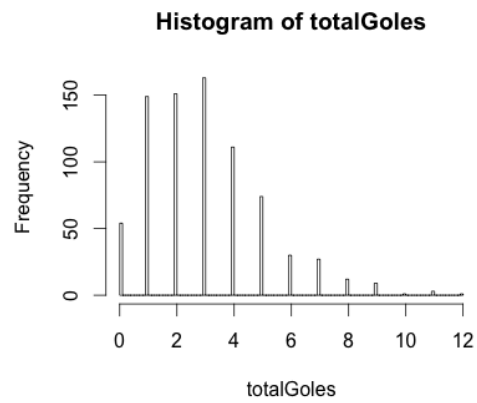
(a) Número de Clases: 5



(b) Número de Clases: 10



(a) Número de Clases: 30



(b) Número de Clases: 100

1.1.3

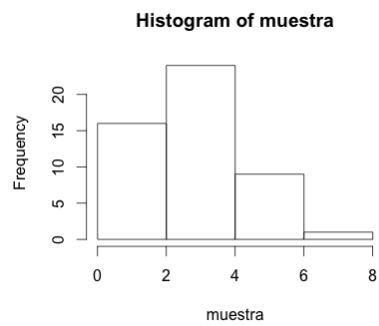
Para esta última subsección, realizaremos los histogramas y divisiones de clases utilizando la regla de Sturges y utilizando muestras en vez del total de nuestra población de datos, la cual irá variando según la siguiente instrucción: entre 10 y 100 de 10 en 10; luego entre 100 y 700 de 100 en 100. Para esta ocasión, utilizaremos el comando **sample** para conseguir una muestra aleatoria de nuestra población y **hist** para construir histogramas. Por simplicidad, se mostrarán sólo 5 de los 16 histogramas, de cualquier forma, los demás se encontrarán en la sección de anexos.

Resulta sencillo concluir que los datos siguen centralizados en los primeros valores del histograma, a pesar de que la muestra es aleatoria. Por otra parte, notemos que los datos sobre 6 van apareciendo a medida que aumentamos el tamaño de nuestra muestra, esto debido a que la cantidad de estos datos es bastante pequeña y la probabilidad de que aparezcan es también lo es.

Tamaño muestral: 10



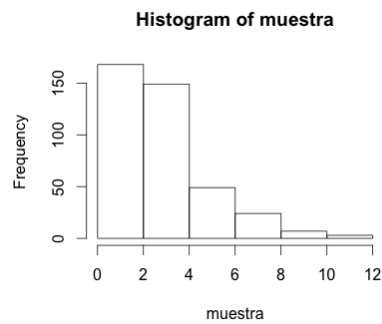
Tamaño muestral: 50



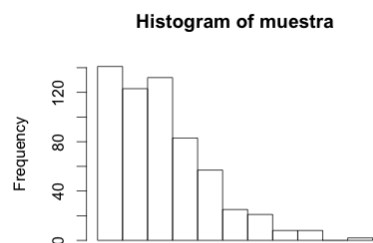
Tamaño muestral: 100



Tamaño muestral: 300



Tamaño muestral: 700



1.2 Medidas de Tendencia y Dispersión

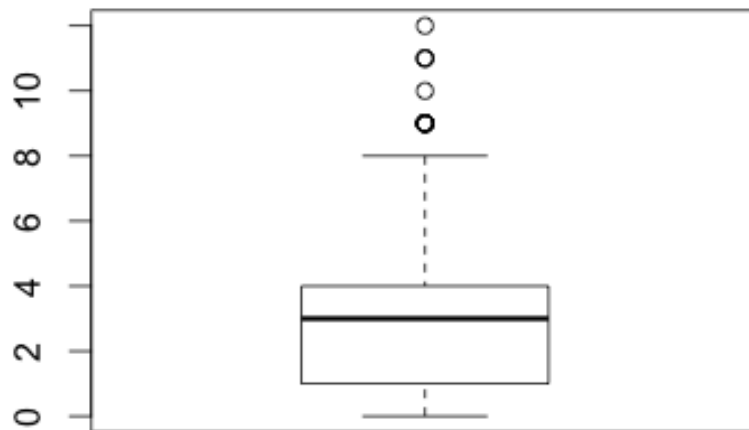
1.2.1

Para nuestra variable de estudio, calcularemos la media, moda, mediana y varianza.

Media	3
Moda	3
Mediana	3
Varianza	4.193878

1.2.2

Se realizará la confección de un boxplot para nuestra variable Total de Goles utilizando el comando `boxplot` de R:



Se puede observar la existencia de outliers, estos van desde los 8 (sin considerar este valor) hasta los 12 goles por partido. A pesar de la existencia

de outliers, y como un buen conocedor de este deporte sabrá, una cantidad mayor a 8 goles por partido es rara vez apreciada, nuestra media no se ve afecta a ello, situación que en pocas ocasiones sucede. Al contrario de los valores de media y mediana, la varianza si se ve afecta a estos outliers, por lo cual el valor de 4.193878 lo confirma. Gracias a estos valores, podemos afirmar que los partidos de mundiales la mayoría de los partidos jugados presenta una cantidad de 3 goles en total, de los cuales un equipo siempre saldrá victorioso.

1.2.3

Media	3
Moda	3
Mediana	3
Varianza	4.193878

100.2919 69.22969

1.2.4

100.2919 69.22969 El coeficiente de variación indica la dispersión de una muestra porcentualmente comparado con su media aritmética, y se calcula dividiendo la desviación estándar por el valor absoluto de la media. Para los goles de Chile, se obtuvo un coeficiente de variación de 100,29% que representa un alto nivel de dispersión en los datos, en comparación al coeficiente de variación obtenido para los goles de Brasil, que es de 69,22%. Esto significa que para el caso de Chile hay una mayor heterogeneidad en los datos, y tiene sentido ya que al observarlos, se aprecia que son pocos datos y hay un par de valores que se alejan bastante de las cifras más repetidas. Al haber pocos datos, con algunos de ellos que se alejan bastante de la media, y con esto se produce un aumento de la variación de dispersión. Por otro lado, en el caso de Brasil, hay una gran cantidad de datos, y la mayoría de ellos se concentran dentro de un rango (1-3 goles) y también cuenta con valores que se alejan de esta media. Sin embargo, al haber una gran cantidad de datos concentrados, los datos más alejados no influyen mucho en el coeficiente de variación, y es por eso que este valor resulta menor que el obtenido en el caso de Chile.

1.2.5

Muestra 1: Moda: 2 Media: 1.918033 Mediana: 2 Varianza: 1.643169 Muestra 2: Moda: 2 Media: 2.071429 Mediana: 2 Varianza: 2.322078 Muestra 3: Moda: 0 Media: 1.685714 Mediana: 2 Varianza: 2.280672 Muestra 4: Moda: 3 Media: 1.941176 Mediana: 2

1.2.6

1.2.7

Anexos

Anexo 1

Total de Goles - K = 30

	Limites	Marca	Mediana	Frec. Absoluta
K_1	[0, 0.4]	0	0	54
K_2	(0.4 , 0.8]	NA	NA	0
K_3	(0.8 , 1.2]	1	1	149
K_4	(1.2 , 1.6]	NA	NA	0
K_5	(1.6 , 2]	2	2	151
K_6	(2 , 2.4]	NA	NA	0
K_7	(2.4 , 2.8]	NA	NA	0
K_8	(2.8 , 3.2]	3	3	163
K_9	(3.2 , 3.6]	NA	NA	0
K_{10}	(3.6 , 4]	4	4	111
K_{11}	(4 , 4.4]	NA	NA	0
K_{12}	(4.4 , 4.8]	NA	NA	0
K_{13}	(4.8 , 5.2]	5	5	74
K_{14}	(5.2 , 5.6]	NA	NA	0
K_{15}	(5.6 , 6]	6	6	30
K_{16}	(6 , 6.4]	NA	NA	0
K_{17}	(6.4 , 6.8]	NA	NA	0
K_{18}	(6.8 , 7.2]	7	7	27
K_{19}	(7.2 , 7.6]	NA	NA	0
K_{20}	(7.6 , 8]	8	8	12
K_{21}	(8 , 8.4]	NA	NA	0
K_{22}	(8.4 , 8.8]	NA	NA	0
K_{23}	(8.8 , 9.2]	9	9	9
K_{24}	(9.2 , 9.6]	NA	NA	0
K_{25}	(9.6 , 10]	10	10	1
K_{26}	(10 , 10.4]	NA	NA	0
K_{27}	(10.4 , 10.8]	NA	NA	0
K_{28}	(10.8 , 11.2]	11	11	3
K_{29}	(11.2 , 11.6]	NA	NA	0
K_{30}	(11.6 , 12]	12	12	1