SESSION 4 SUMMARY

Victor Miguel Terrón Macias

21/1/2021

SESSION 4. ALGUNAS DISTRIBUCIONES, TEOREMA CENTRAL DEL LÍMITE Y CONTRASTE DE HIPOTESIS

CONCEPTOS DE ESTADÍSTICA INFERENCIAL BÁSICOS

DISTRIBUCIÓN BINOMIAL

Se define como un experimento con las siguientes características: * Consiste en un número fijo, n, de pruebas idénticas. * Cada prueba resulta en uno de dos resultados: éxito S o fracaso F * La probabilidad de éxito en una sola prueba es igual a algún valor p y es la misma de una prueba a otra. La probabilidad de fracaso es igual a $q = 1 - p$ * Las pruebas son independientes * La variable aleatoria (v.a.discreta) de interés es Y, el número de éxitos observado durante las n pruebas.

Se dice que una variable aleatoria Y tiene una distribución binomial basada en n pruebas con probabilidad p de éxito, si y solo si.

$$P_x = \binom{n}{x} p^x q^{m-x}$$

De donde P es probabilidad binomial, de donde x es el numero de veces para obtener un resultado específico en n ensayos, de donde $\binom{n}{x}$, de donde p es la probabilidad de exito en un solo ensayo,q es probabilidad de fallo en un solo ensayo y n es el numero de ensayos. Y donde P que es la posibilidad tiene que cumplir con $0 \leq P \leq 1$

EJEMPLO DE APLICACION DE DISTRIBUCION BINOMIAL

La última novela de un autor ha tenido un gran éxito, hasta el punto de que el 80% de los lectores ya la han leído. Hallar la probabilidad de que en un grupo de 4 amigos que son aficionados a la lectura, 2 hayan leído la novela.

1. Hallar la probabilidad de que una persona haya leído el libro es de 0.8, por lo que la probabilidad de que no lo haya leído es de 0.2. De donde tenemos los siguientes datos: $n = 4, k = 2, p = 0.8, q = 0.2$
2. Hallar la probabilidad de que máximo 2 personas del grupo de 4 amigos hayan leído la novela: tenemos los siguientes datos:

$$P(x \leq 2) = P(x = 0) + P(x = 1) + P(x = 2)$$

Sustituyendo los datos en la fomrula de distribución binomial tenemos lo siguiente:

$$P_{x\leq 2} = \binom{4}{0}(0.8)^0(0.2)^{4-0} + \binom{4}{1}(0.8)^1(0.2)^{4-1} + \binom{4}{2}(0.8)^2(0.2)^{4-2} = 0.1808$$

DISTRIBUCIÓN NORMAL

A una distribución normal se le conoce como distribución Gaussiana o distribucion de Laplace Gauss. Se utiliza csolamente con variables continuas (variables que pueden tomar un numero infinito de valores entre dos valores cualesquiera de una caracteristica).

Su gráfica es en forma de campana y simétrica respecto de un determinado parametro estadístico. Éste tipo de distribución propicia el modelado de numerosos fenómenos naturales, sociales y psicológicos. El centro de la campana es el promedio. La desviación estandar nos indican que tan dispersos o separados están los datos conr especto la media.

Para calcular la probabilidad normal se divide la cantidad de casos favorables entre la cantidad de datos totales, para ello se tuliza la siguiente fórmula:

$$z = \frac{x-\mu}{\sigma}$$

De donde tenemos que x es el valor de la condición y μ es el promedio también conocido como media y σ es la desviación estandar. Manualmente podríamos calcularlo con las tablas de distribución normal.

Las funciones de densidad de probabilidad (de variables aleatorias continuas) cumplen con las siguientes propiedades:

- El área bajo la curva de la función de densidad de probabilidad es igual a 1
- La probabilidad de que X se encuentre en determinado intervalo (a, b) es igual al área bajo la curva entre los puntos a y b.
- $P(X = c) = 0$ para cualquier valor c para el que se encuentre definida la función de densidad.

La funcion de densidad de probabilidad de una variable aleatoria x que se distribuye como normal con media μ y desviacion estandar σ es:

$$f(x) = \frac{1}{\sigma\cdot\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{(2\sigma)^2}}$$

DISTRIBUCIÓN t DE STUDENT

La función de densidad t de Student se define como:

En el caso de la distribución t la media $\mu = 0$ y

$$\sigma^2 = \frac{v}{(v-2)}$$

para $v > 2$ respectivamente.

La apariencia general de la distribución t es similar a la de la distribución normal estándar: ambas son simétricas y unimodales, y el valor máximo de la ordenada se alcanza en la media $\mu = 0$. Sin embargo, la distribución t tiene colas más amplias que la normal; esto es, la probabilidad de las colas es mayor que en la distribución normal. A medida que el número de grados de libertad tiende a infinito, la forma límite de la distribución t es la distribución normal estándar.

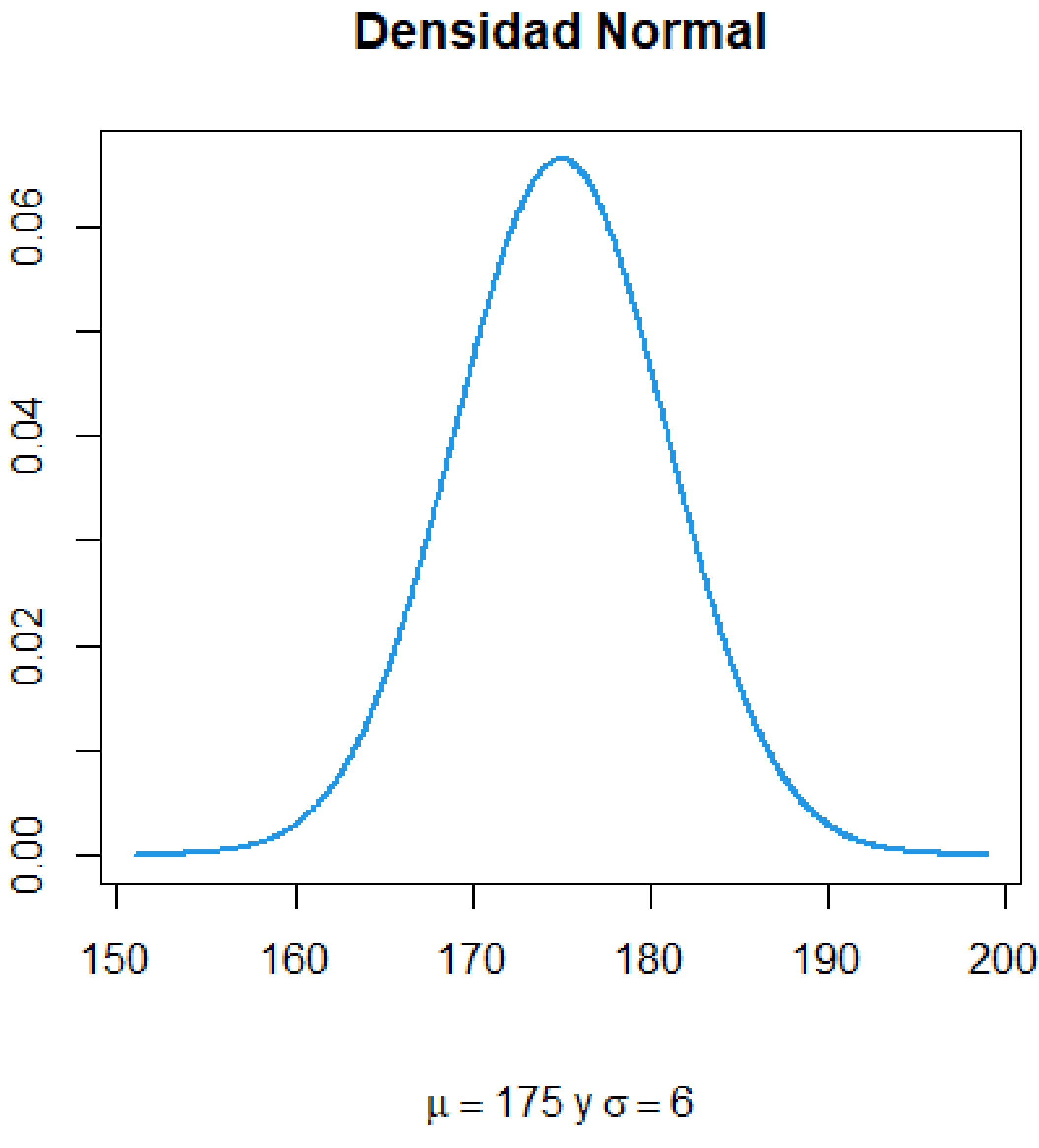


Figure 1: Distribución normal

PROPIEDADES DE LA DISTRIBUCIÓN t DE STUDENT

- Cada curva tiene forma de campana con centro en 0.
- Cada curva t está más dispersa que la curva normal estandar z
- A medida que v aumenta, la distribución de la curva t correspondiente disminuye
- A medida que $v \to \infty$ la secuencia de curvas t se aproxima a la curva normal estandar, por lo que la curva z recibe a veces el nombre de curva t con grados de libertad (gl)$gl = \infty$.

La distribución de la variable aleatoria t está dada por:

$$f(t) = \frac{\Gamma\left(\dfrac{v+1}{2}\right)}{\Gamma\left(\dfrac{v}{2}\right) \cdot \sqrt{\pi v \sigma}} \cdot \left(1 + \frac{1}{v}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\dfrac{v+1}{2}}$$

De donde debemos recordar que σ^2 corresponde a varianza y no desviación estandar. Y que v son los grados de libertad.

La formula de la varianza es:

$$\sigma^2 \cdot \frac{v}{v-2}$$

, la moda es $= \mu$

TEOREMA DEL LÍMITE CENTRAL

Sean y_1, y_2,. . . , y_n variables aleatorias independientes y distribuidas idénticamente con $E(y_i) = y \; vy_i = 2 < \infty$. Definamos:

$$U_n = \frac{\sum_{i=1}^{n} y_i - n\mu}{\sigma\sqrt{n}} = \frac{\bar{y} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$$

Entonces la función de distribución de U_n converge hacia la función de distribución normal estándar cuando $n \to \infty$. Esto es:

$$P(U_n) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} \cdot e^{\dfrac{-t^2}{2}} \, dt$$

Para toda u. Es decir que $\bar{y}$, está distribuida normalmente en forma asintótica con media μ y varianza

$$\frac{\sigma^2}{n}$$

.

El teorema central del límite se puede aplciar a una muestra aleatoria y_1,y_2,...,y_n para cualquier distribución mientras $E(y_i) = \mu$ y $V(y_i) = \sigma^2$ sean finitas y el tamaño muestral sea grande.

ESTIMADORES PUNTUALES INSESGADOS COMUNES

DEFINICIÓN Un estimador es una regla, a menudo expresada como una fórmula, que indica cómo calcular el valor de una estimación con base en las mediciones contenidas en una muestra.

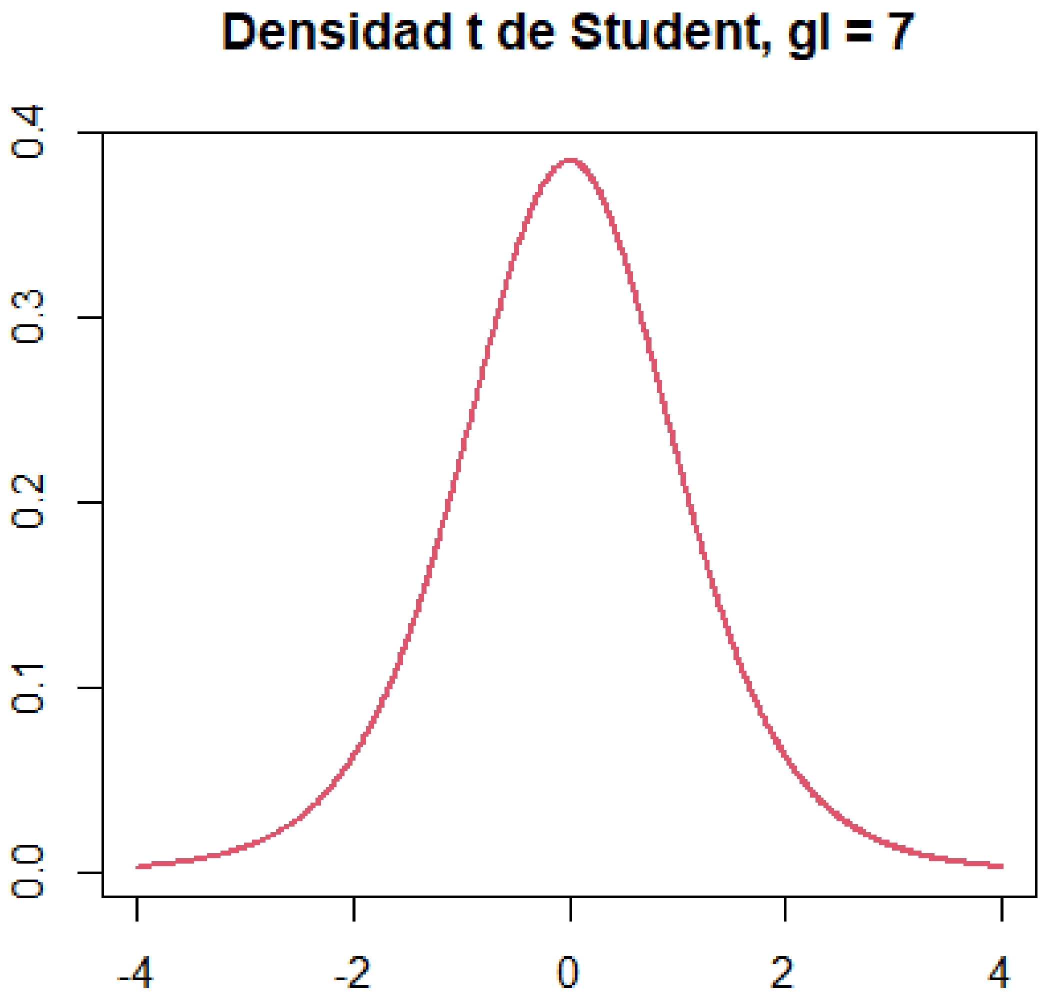


Figure 2: Densidad t de student con 7 GDL

Definición. Si $\hat{\theta}$ es un estimador puntual de un parámetro θ, entonces $\hat{\theta}$ es un estimador insesgado si $E\left(\hat{\theta}\right) = \theta$. Si $E\left(\hat{\theta}\right) \neq \theta$, se dice que $\hat{\theta}$ está sesgado.

Valores esperados y errores estándar de algunos estimadores puntuales comunes				
Parámetro objetivo θ	Tamaño(s) muestral(es)	Estimador puntual $\hat{\theta}$	$E(\hat{\theta})$	Error estándar $\sigma_{\hat{\theta}}$
μ	n	$\overline{Y}$	μ	$\frac{\sigma}{\sqrt{n}}$
p	n	$\hat{p} = \frac{Y}{n}$	p	$\sqrt{pq}$ [illegible]
$\mu_1 - \mu_2$	n_1 y n_2	$\overline{Y}_1 - \overline{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\sigma_1^2 \quad \sigma_2^2}$ [illegible]
$p_1 - p_2$	n_1 y n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{p_1q_1 + p_2q_2}$ [illegible]
σ_1^2 y σ_2^2 son las varianzas de las poblaciones 1 y 2, respectivamente.				
Se supone que las dos muestras son independientes.				

Figure 3: valores esperados y errores estándar de algunos estimadores puntuales comunes

CONTRASTE DE HIPOTESIS

Los elementos de un contraste de hipotesis son: * Hipótesis nula, H0 * Hipotesis alternativa, Ha * Estadístico de prueba * Región de rechazo

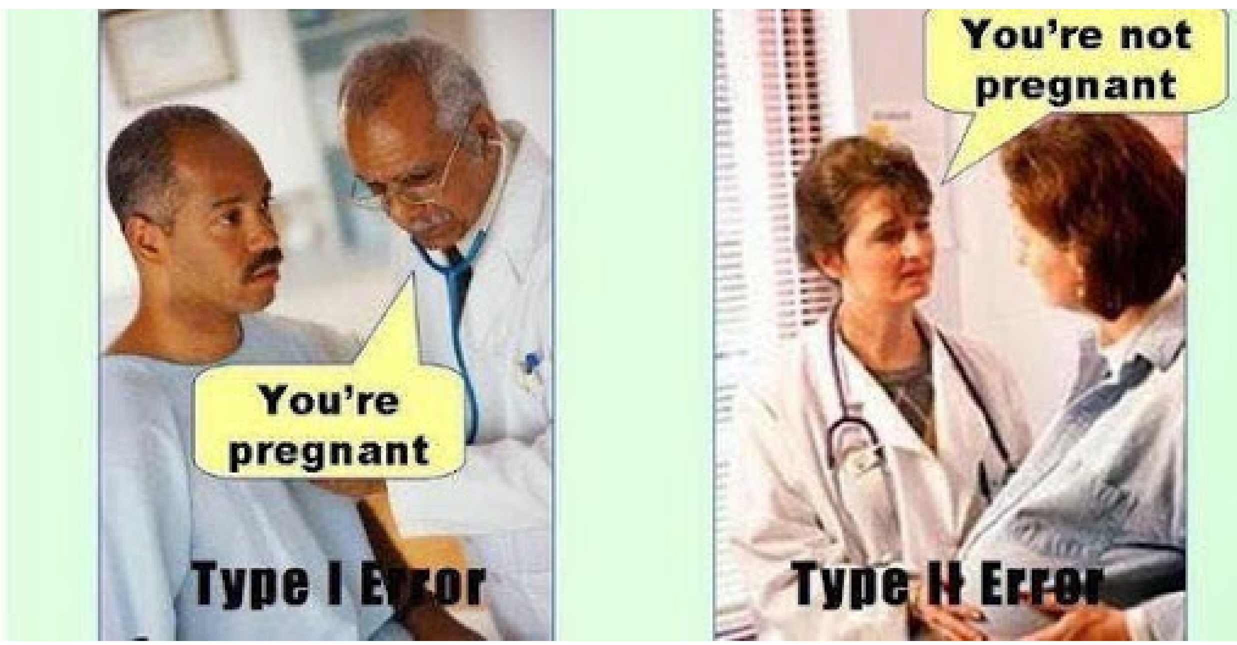
Nota: También llamaremos prueba de hipótesis a un contraste de hipótesis, sin caer en discusiones formales. Buscamos decidirnos por una de las hipótesis y no estamos exentos de cometer errores.

Definición. Se comete un *error tipo I* si H_0 es rechazada cuando H_0 es verdadera. La *probabilidad de un error tipo I* está denotada por α. El valor de α se denomina *nivel* de la prueba.

Se comente un *error tipo II* si H_0 es aceptada cuando H_a es verdadera. La probabilidad de un *error tipo II* está denotada por β.

Error tipo I y tipo II

H_0: No hay embarazo *vs* H_a: Hay embarazo



CONTRASTES COMUNES CON MUESTRAS GRANDES

Suponga que deseamos contrastar un conjunto de hipótesis respecto a un parámetro con base a una muestra aleatoria $Y_1, Y_2, \ldots, Y_n$. En esta sección desarrollaremos procedimientos de contrastes de hipótesis que están basados en un estimador que tiene una distribución muestral normal (aproximadamente) con media y error estándar .

Si 0 es un valor específico de, podemos probar H0: $\theta = 0$ contra Ha: $\theta > 0$. En este caso, las hipótesis nula y alternativa, el estadístico de prueba y la región de rechazo son como sigue:

$$H_0 : \theta = \theta_0 .$$

$$H_a : \theta > \theta_0 .$$

Estadístico de prueba: $\hat{\theta}$.

Región de rechazo: RR = { $\hat{\theta} > k$}, para alguna selección de k.

El valor real de k en la región de rechazo RR se determina al fijar la probabilidad α de error tipo I (el nivel de la prueba) y escoger k de conformidad. Si H_0 es verdadera, $\hat{\theta}$ tiene una distribución aproximadamente normal con media θ_0 y error estándar $\sigma_{\hat{\theta}}$.

Por tanto, si deseamos una prueba de nivel α ,

$$k = \theta_0 + z_\alpha \sigma_{\hat{\theta}}$$

Es la selección apropiada para k [si Z tiene una distribución normal estándar, entonces Z_α es tal que $P(Z > Z_\alpha) = \alpha$].

Figure 4: A

CONTRASTES DE HIPOTESIS DE NIVEL PARA MUESTRAS GRANDES

¿Cómo decidir cuál hipótesis alternativa usar para una prueba? La respuesta depende de la hipótesis que pretendemos apoyar. Si estamos interesados sólo en detectar un aumento en el porcentaje de piezas defectuosas, por ejemplo, debemos localizar la región de rechazo en la cola superior de la distribución normal estándar. Por otra parte, si deseamos detectar un cambio en p ya sea arriba o debajo de p=0.10, debemos localizar la región de rechazo en ambas colas de la distribución normal estándar y emplear una prueba de dos colas.

$H_0 : \theta = \theta_0$.

$H_a : \{\theta > \theta_0$ *(alternativa de cola superior)* . $\theta < \theta_0$ *(alternativa de cola inferior)* . $\theta \neq \theta_0$ *(alternativa de dos colas)* .

Estadístico de prueba:

$$Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

Región de rechazo:

$\{\{z > z_\alpha\}$ *(RR de cola superior)* . $\{z <- z_\alpha\}$ *(RR de cola inferior)* . $\left\{|z| > z_{\frac{\alpha}{2}}\right\}$ *(RR de dos colas)* .

Figure 5: B

Contraste de hipótesis con muestras pequeñas para μ y $\mu_1 - \mu_2$

Supongamos que $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de tamaño *n* de una distribución normal con media μ desconocida y varianza σ^2 desconocida. Si $\overline{Y}$ y S denotan la media muestral y la desviación estándar muestral, respectivamente, y si $H_0 : \mu = \mu_0$ es verdadera, entonces

$$T = \frac{\overline{Y} - \mu_0}{S/\sqrt{n}}$$

Tiene una distribución *t* con *n-1* grados de libertad.

Figure 6: C

Contraste de muestra pequeña para μ

Suposiciones: $Y_1, Y_2, \ldots, Y_n$ constituyen una muestra aleatoria de una distribución normal con $E(Y_i) = \mu$.

$H_0: \ \mu = \mu_0$.

$H_a: \ \{\mu > \mu_0 \ (alternativa\ de\ cola\ superior).\ \mu < \mu_0 \ (alternativa\ de\ cola\ inferior).\ \mu \neq \mu_0 \ (alternativa\ de\ dos\ colas).$

Estadístico de prueba: $T = \frac{\overline{Y} - \mu_0}{S/\sqrt{n}}$.

Región de rechazo: $\{t > t_\alpha \ \ (RR\ de\ cola\ superior).\ t < -t_\alpha \ (RR\ de\ cola\ inferior).\ |t| > t_{\alpha/2} \ \ (RR\ de\ dos\ colar).$

Figure 7: D

Contrastes con muestras pequeñas para comparar dos medias poblacionales

Suposiciones: muestras independientes de distribuciones normales con $\sigma_1^2 = \sigma_2^2$.

$H_0: \ \mu_1 - \mu_2 = D_0$. Donde D_0 es un número fijo.

$H_a: \ \{\mu_1 - \mu_2 > D_0 \ (alternativa\ de\ cola\ superior).\ \mu_1 - \mu_2 < D_0 \ (alternativa\ de\ cola\ inferior).$

$\mu_1 - \mu_2 \neq D_0 \qquad (alternativa\ de\ dos\ colas)$

Estadístico de prueba:

$$T = \frac{\overline{Y}_1 - \overline{Y}_2 - D_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Donde *T* tiene una distribución t de Student con $n_1 + n_2 - 2$ grados de libertad y

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}}$$

Figure 8: E

Estadístico de prueba:

$$T = \frac{\overline{Y}_1 - \overline{Y}_2 - D_0}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

Donde T tiene una distribución t de Student con $n_1 + n_2 - 2$ grados de libertad y

$$S_p = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

Región de rechazo: $\{t > t_\alpha$ $(RR\ de\ cola\ superior)$. $t < -t_\alpha$ $(RR\ de\ cola\ inferior)$. $|t| > t_{\alpha/2}$ $(RR\ de\ dos\ colas)$.

Aquí, $P(T > t_\alpha) = \alpha$.

Figure 9: F

WORK

Estudiar algunas distribuciones de probabilidad muy comunes y útiles, obtener estimaciones puntuales con propiedades deseables utilizando algunos estimadores insesgados comunes. Llevar a cabo contrastes de hipótesis que ayuden a tomar decisiones.

En esta sesión estudiaremos temas relacionados con los siguientes puntos:

- Cálculo de probabilidades y cuantiles de las distribuciones binomial, normal y t de Student
- Generación de muestras aleatorias de las distribuciones estudiadas
- Estudio del teorema central del límite mediante simulaciones
- Propiedades de algunos estimadores puntuales insesgados comunes
- Contraste de hipótesis con muestras grandes y pequeñas

EJEMPLO 1 SESION 4. DISTRIBUCIONES BINOMIAL, NORMAL Y T DE STUDENT

OBJETIVO

- Aprender a obtener probabilidades, cuantiles y muestras aleatorias relacionadas con las distribuciones binomial, normal y t de Student
- Intepretar las probabilidades cuando se condieran las gráficas de las funciones de probabilidad y de densidad

REQUISITOS

- Tener `R` y Rstudio instalados
- Haber leído el pre-work

DESARROLLO

```r
library(ggplot2) # Utilizaremos estos paquetes para algunas gráficas
library(reshape2)
```

DISTRIBUCION BINOMIAL

En el caso de la **Distribución binomial** En `R` para calcular valores de las funciones de probabilidad, distribución o cuantiles de la distribución binomial (discreta), usamos las funciones *dbinom*, *pbinom* y *qbinom* respectivamente. Para generar muestras aleatorias de esta distribución utilizamos la función *rbinom*.

Consideremos un experimento binomial con $n = 30$ pruebas idénticas e independientes, en donde la probabilidad de éxito en cada prueba es $p = 0.2$ (parámetros $n = 30$ y $p = 0.2$) 1. Suponga que realiza un examen de opción múltiple con 30 preguntas, en donde cada pregunta tiene 5 posibles respuestas, pero solo una es correcta siempre. Si elige la respuesta al azar en cada pregunta, y estamos interesados en el número de respuestas correctas obtenidas al final ¿Podemos decir que estamos ante un experimento binomial?

FUNCIÓN DE PROBABILIDAD

Para obtener P(X = 20), es decir, la probabilidad de observar 20 éxitos exactamente, en R ejecutamos:

```r
dbinom(x = 20, size = 30, prob = 0.2)
```

```
[1] 3.382768e-08
```

Para obtener P($x \leq 20$), es decir, la probabilidad de observar a lo mucho 20 exitos o menos ejecutamos:

```r
pbinom(x<=20, size = 30, prob = 0.2)
```

La diferencia entre *dbinom*() y *pbino*() es que *dbinom*() me dice cuál es la probabilidad de que $Pr(X = x)$ mientras que *pbinom*() te calcula la probabilidad de que $Pr(X \leq x)$

Para encontrar el valor más pequeño b tal que P(X <= b) >= 0.35, es decir, el cuantil de orden 0.35, usamos:

CUANTILES

```r
qbinom(p = 0.35, size = 30, prob = 0.2) # b = 5

pbinom(q = 4, size = 30, prob = 0.2) # P(X <= 4) = 0.2552 < 0.35
pbinom(q = 5, size = 30, prob = 0.2) # P(X <= 5) = 0.4275 >= 0.35
pbinom(q = 6, size = 30, prob = 0.2) # P(X <= 6) = 0.6070 >= 0.35
```

```
[1] 5
[1] 0.2552333
[1] 0.4275124
[1] 0.6069699
```

MUESTRAS ALEATORIAS

Para obtener una muestra aleatoria de tamaño $n = 1000$, de la distribución binomial con parámetros como especificamos, hacemos

```
set.seed(4857) # Establecemos una semilla,
# para poder reproducir la muestra en el futuro
muestra <- rbinom(n = 1000, size = 30, prob = 0.2)
length(muestra); muestra[1:3]
```

```
[1] 1000
[1] 4 7 8
```

Podemos observar las frecuencias absolutas de los distintos valores obtenidos

```
as.data.frame(table(muestra))
```

```
   muestra Freq
1        0    1
2        1    8
3        2   31
4        3   77
5        4  132
6        5  161
7        6  196
8        7  146
9        8  126
10       9   65
11      10   30
12      11   13
13      12   10
14      13    2
15      15    2
```

También podemos observar las frecuencias relativas:

```
(df1 <- as.data.frame(table(muestra)/length(muestra)))

valg <- as.character(df1$muestra) # distintos valores generados por rbinom
(valg <- as.numeric(valg)) # Convertimos a números
```

```
   muestra  Freq
1        0 0.001
2        1 0.008
3        2 0.031
4        3 0.077
5        4 0.132
6        5 0.161
7        6 0.196
8        7 0.146
9        8 0.126
10       9 0.065
```

```
11       10 0.030
12       11 0.013
13       12 0.010
14       13 0.002
15       15 0.002
 [1]  0  1  2  3  4  5  6  7  8  9 10 11 12 13 15
```

Las frecuencias relativas son muy parecidas a las siguientes probabilidades

```r
(v1 <- round(sapply(valg, dbinom, size = 30, p = 0.2), 3))
```

```
 [1] 0.001 0.009 0.034 0.079 0.133 0.172 0.179 0.154 0.111 0.068 0.035 0.016
[13] 0.006 0.002 0.000
```

Combinamos y unimos en un solo dataframe $df1$ y $v1$

```r
(df2 <- cbind(df1, v1))
(names(df2) <- c("Exitos", "FR", "Prob"))

(df2 <- melt(df2)) # función del paquete reshape2
```

```
Using Exitos as id variables
```

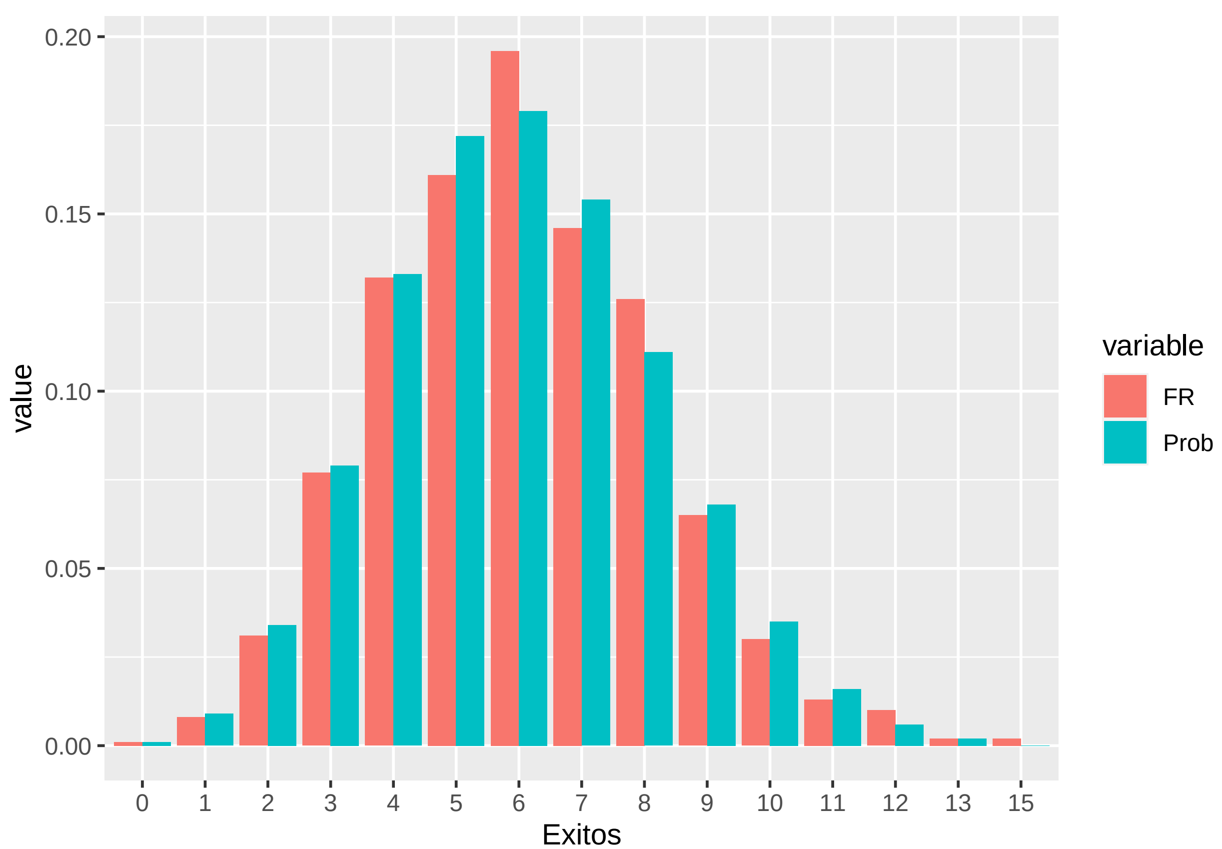
```
   muestra  Freq    v1
1        0 0.001 0.001
2        1 0.008 0.009
3        2 0.031 0.034
4        3 0.077 0.079
5        4 0.132 0.133
6        5 0.161 0.172
7        6 0.196 0.179
8        7 0.146 0.154
9        8 0.126 0.111
10       9 0.065 0.068
11      10 0.030 0.035
12      11 0.013 0.016
13      12 0.010 0.006
14      13 0.002 0.002
15      15 0.002 0.000
[1] "Exitos" "FR"     "Prob"
   Exitos variable value
1       0       FR 0.001
2       1       FR 0.008
3       2       FR 0.031
4       3       FR 0.077
5       4       FR 0.132
6       5       FR 0.161
7       6       FR 0.196
8       7       FR 0.146
9       8       FR 0.126
10      9       FR 0.065
11     10       FR 0.030
```

```
12      11        FR 0.013
13      12        FR 0.010
14      13        FR 0.002
15      15        FR 0.002
16       0      Prob 0.001
17       1      Prob 0.009
18       2      Prob 0.034
19       3      Prob 0.079
20       4      Prob 0.133
21       5      Prob 0.172
22       6      Prob 0.179
23       7      Prob 0.154
24       8      Prob 0.111
25       9      Prob 0.068
26      10      Prob 0.035
27      11      Prob 0.016
28      12      Prob 0.006
29      13      Prob 0.002
30      15      Prob 0.000
```

Melt en cierta forma une los dataframes en uno solo pero basandose en ciertas condiciones

Las frecuencias relativas son muy parecidas a las probabilidades:

```
ggplot(df2, aes(x = Exitos, y = value, fill = variable)) +
  geom_bar (stat="identity", position = "dodge")
```



```r
# Funciones del paquete ggplot2
```

DISTRIBUCIÓN NORMAL

En R para calcular valores de las funciones de densidad, distribución o cuantiles de la distribución normal (continua), usamos las funciones *dnorm*, *pnorm* y *qnorm* respectivamente. Para generar muestras aleatorias de esta distribución utilizamos la función *rnorm*.

Consideremos una variable aleatoria (v.a.) X que se distribuye como normal con media 175 y desviación estándar 6 (parámetros mu = 175 y sigma = 6).

FUNCIÓN DE DENSIDAD

La función de densidad sirve para caracterizar el comportamiento probable de una población en tanto especifíca la posibilidad relativa de que una variable aleatoria continuas X tome un valor cercano a x.

Densidad significa que la suma de todas las areas del histograma deben ser igual a 1. La funcion de densidad es el contorno. Es una línea continua que representa la distribucion de densidad de TODA LA POBLACIÓN.

FUNCION DE DISTRIBUCIÓN

La función de distribución asocia a cada valor de la variable alewatoria la probabilidad acumulada hasta ese valor, por ejemplo: Calcular la función de distribución de probabilidad de las puntuaciones obtenidas al lanzar un dado.

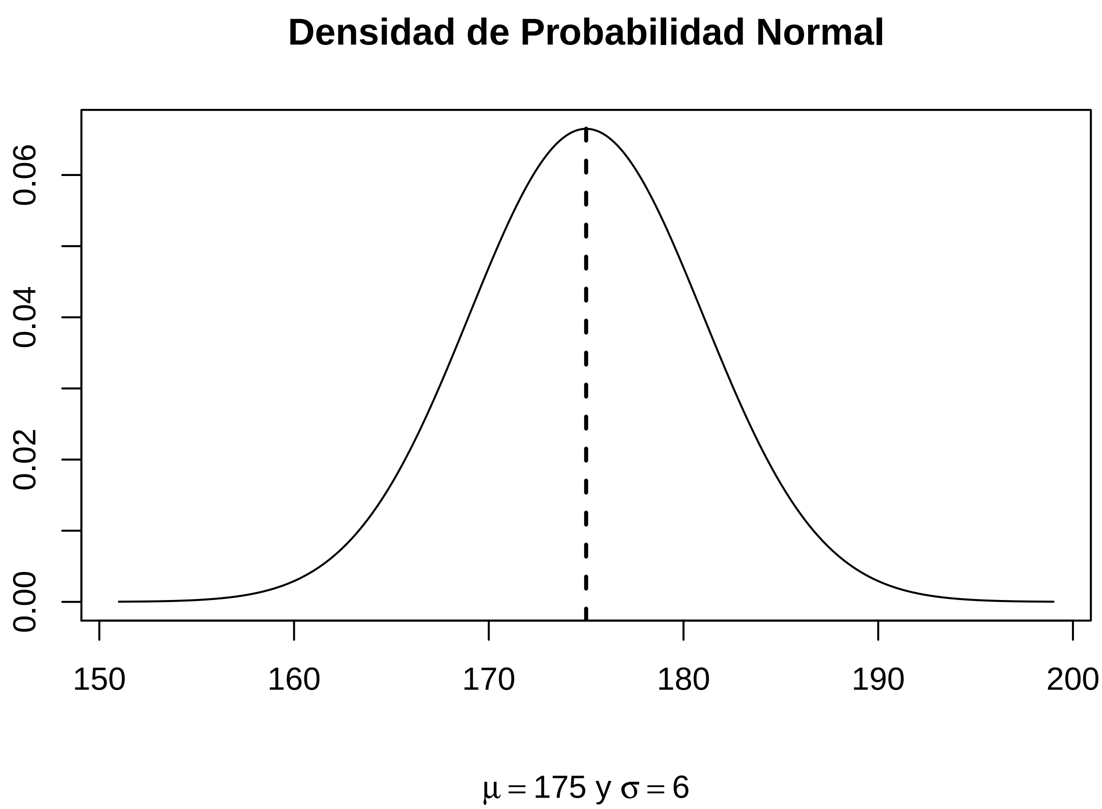
X	P_i
$x < 1$	0
$1 \leq x < 2$	$\frac{1}{6}$
$2 \leq x < 3$	$\frac{2}{6}$
$3 \leq x < 4$	$\frac{3}{6}$
$4 \leq x < 5$	$\frac{4}{6}$
$5 \leq x < 6$	$\frac{5}{6}$
$x \leq 6$	1

Siguiendo con el ejemplo disponible en el work tenemos lo siguiente:

Para obtener $P(x \leq 180)$, es decir, la probabilidad de que X tome un valor menor o igual a 180, ejecutamos:

```r
x <- seq(-4, 4, 0.01)*6 + 175 # Algunos posibles
# valores que puede tomar la v.a.
# X (mínimo: mu-4sigma, máximo: mu+4sigma)
```

```r
y <- dnorm(x, mean = 175, sd = 6) # Valores
# correspondientes de la función de
# densidad de probabilidad
plot(x, y, type = "l", xlab = "", ylab = "")#Te grafica
# en forma de campana los valores de x y y
title(main = "Densidad de Probabilidad Normal",
      sub = expression(paste(mu == 175, " y ", sigma == 6)))#Agrega leyendas
#a gráfico
abline(v = 175, lwd = 2, lty = 2) # La media es 175 grafica la línea
```




```r
pnorm(q = 180, mean = 175, sd = 6)
```

```
[1] 0.7976716
```

```r
par(mfrow = c(2, 2))#ESTABLECE EL LIENZO EN EL PLOT PARA GRAFICAR
#VARIAS EN UN MISMO ESPACIO
plot(x, y, type = "l",
     xlab = "",
     ylab = "")
#AGREGA LEYENDAS AL GRÁFICO
title(main = "Densidad de Probabilidad Normal",
      sub = expression(paste(mu == 175,
                             " y ",
                             sigma == 6)))
```

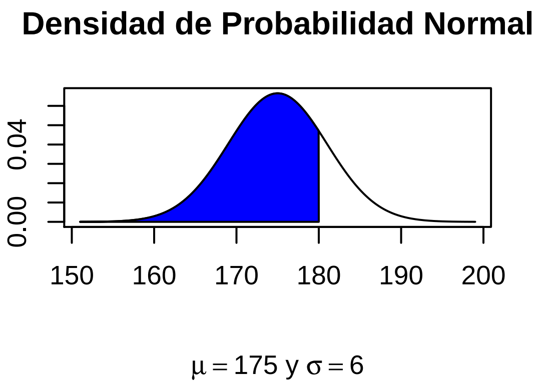
```r
polygon(c(min(x),
          x[x<=180],
          180),
        c(0, y[x<=180], 0),
        col="blue")#GRAFICA DENTRO DE LA DISTRIBUCIÓN
#LA PROBABILIDAD DE QUE X TOME VALORES MENORES A 180


#para obtener la probabilidad de que x P(x<=165), es decir, la probabilidad de
#que X tome un valor menor o igual a 165, ejecutamos:

pnorm(q = 165, mean = 175, sd = 6)
```

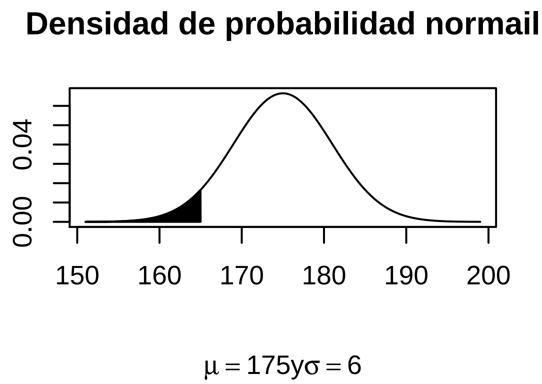
```
[1] 0.04779035
```

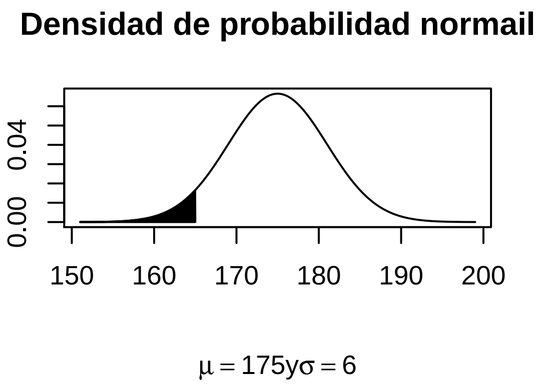
```r
plot(x,y,type="l",
     xlab = "",
     ylab = "")
title(main = "Densidad de probabilidad normail",
      sub = expression(paste(mu==175,"y",sigma==6)))
polygon(c(min(x),
          x[x<=165],165),
        c(0,
          y[x<=165],0),
        col = "black")

#para obtener la probabilidad de que x P(x<=165), es decir, la probabilidad de
#que X tome un valor menor o igual a 165, ejecutamos:

pnorm(q = 165, mean = 175, sd = 6)
```

```
[1] 0.04779035
```

```r
plot(x,y,type="l",
     xlab = "",
     ylab = "")
title(main = "Densidad de probabilidad normail",
      sub = expression(paste(mu==175,"y",sigma==6)))
polygon(c(min(x),
          x[x<=165],165),
        c(0,
          y[x<=165],0),
        col = "black")
```





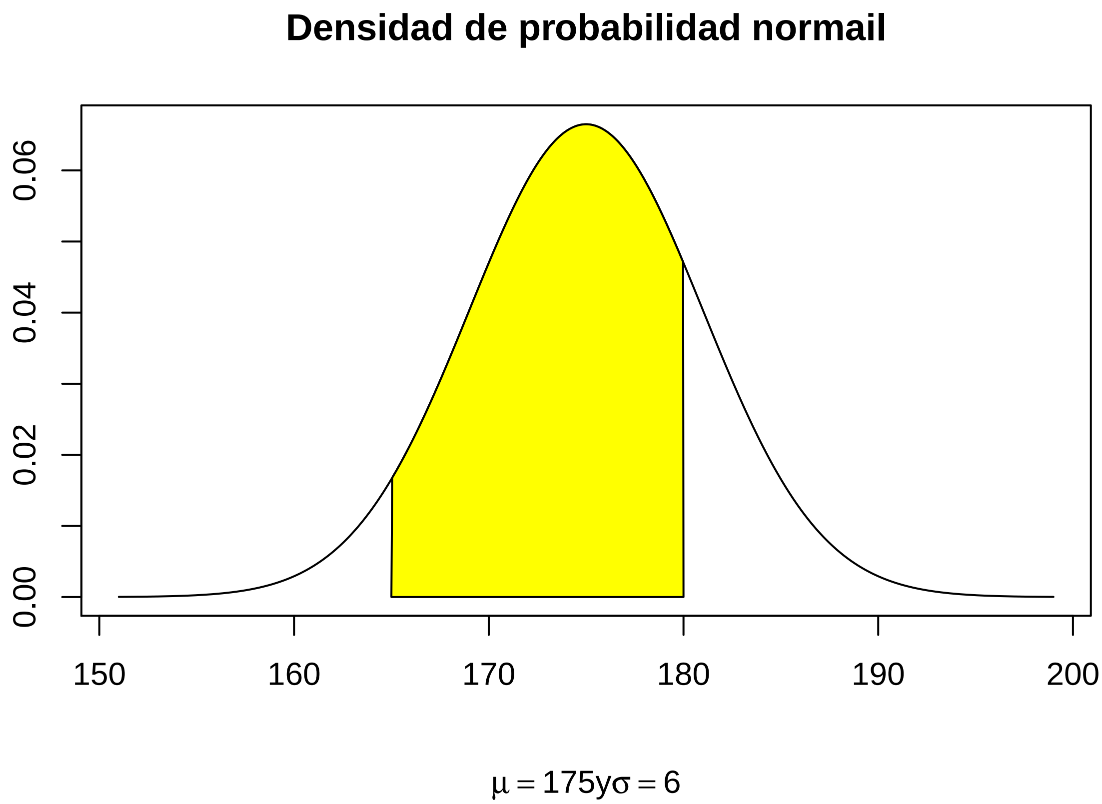


Para obtener la probabilidad de que $P(165 \leq X \leq 180)$ ejecutamos el siguiente comando:

```
#PARA OBTENER P(165 <= X <= 180), es decir, la probabilidad de que X
#tome un valor mayor o igual a 165 y menor o igual a 180, debemos correr

pnorm(q = 180, mean = 175, sd = 6) - pnorm(q = 165, mean = 175, sd = 6)
```

```
[1] 0.7498813
```

```
plot(x,y,type="l",
     xlab = "",
     ylab = "")
title(main = "Densidad de probabilidad normail",
      sub = expression(paste(mu==175,"y",sigma==6)))
polygon(c(165,
          x[x>=165&x<=180],180),
        c(0,
          y[x>=165&x<=180],0),
        col = "yellow")
```

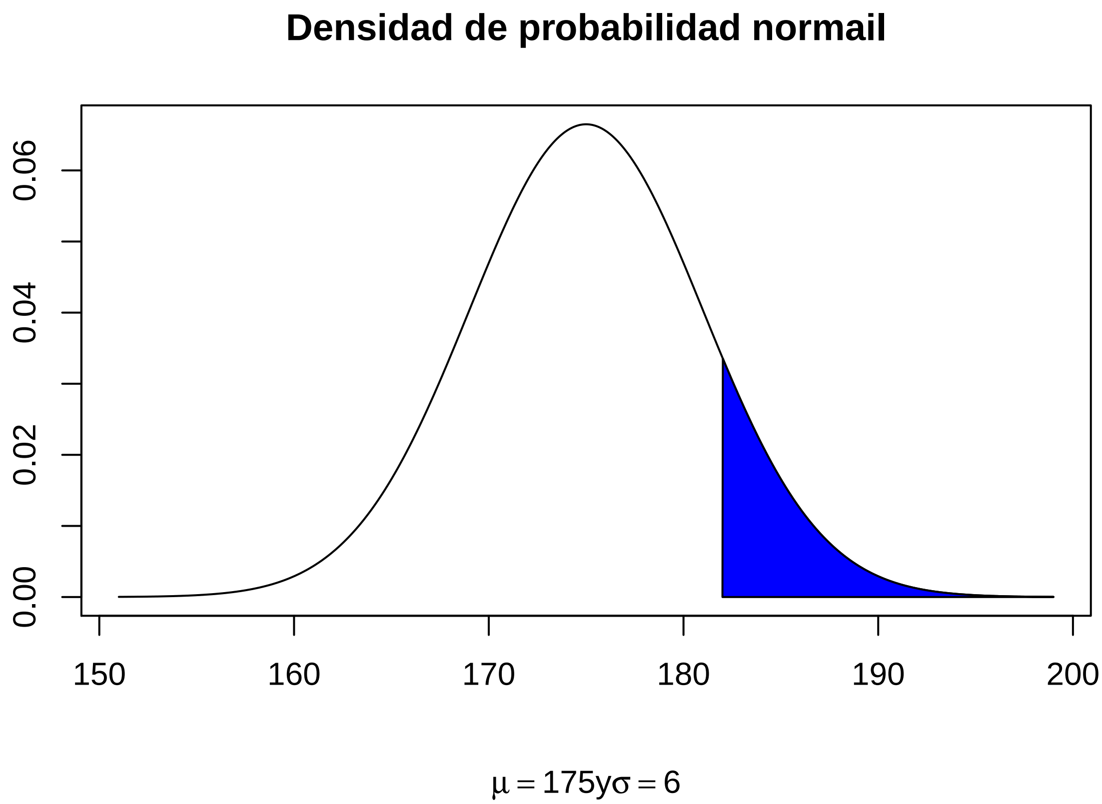


Para obtener $P(X \geq 182)$, es decir, la probabilidad de que X tome un valor mayor o igual a 182, una alternativa es:

```
pnorm(q = 182, mean = 175, sd = 6, lower.tail = FALSE)
```

```
[1] 0.1216725
```

```
#graficando
plot(x,y,type="l",
     xlab = "",
     ylab = "")
title(main = "Densidad de probabilidad normail",
      sub = expression(paste(mu==175,"y",sigma==6)))

polygon(c(182,
          x[x>=182], max(x)),
        c(0,
          y[x>=182], 0),
        col="blue")
```




```
dev.off() # Para mostrar solo una gráfica
```

```
null device
          1
```

CUANTILES

Recordemos que los cuantiles ayudan a determinar porcentajes que superan o no un determinado valor de la variable. Para encontrar el número b, tal que $P(X \leq b) = 0.75$, es decir, el cuantil de orden 0.75, ejecutamos:

```
(b <- qnorm(p = 0.75, mean = 175, sd = 6))
```

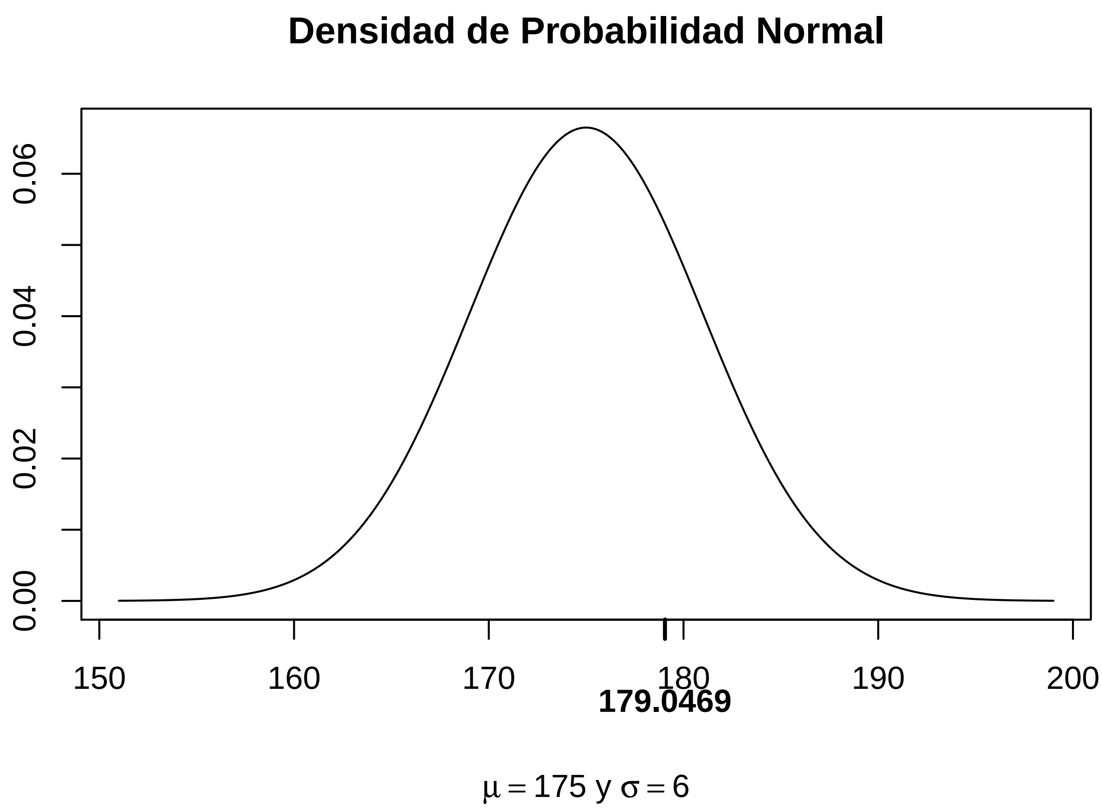
```
[1] 179.0469
```

```
#COMPROBANDO DE FORMA COMUN
pnorm(b, 175, 6)
```

```
[1] 0.75
```

```
# El cuantil se encuentra en el eje de medición (eje horizontal)
plot(x, y,
     type = "l",
```

```
     xlab="",
     ylab="")
title(main = "Densidad de Probabilidad Normal",
      sub = expression(paste(mu == 175, " y ", sigma == 6)))
#COLOCA UNA LÍNEA DONDE SE ENCUENTRA EL CUANTIL
axis(side = 1, at = b, font = 2, padj = 1, lwd = 2)
```




```
# SIDE un entero que especifica donde se colocará la linea
#1 abajo 2 izquierda 3 arriva 4 derecha
#AT es donde se colcoará la linea
#padj ajusta a cada linea, 0 es arriba o derecha, 1 es izq o inferior
```

MUESTRAS ALEATORIAS

Para generar una muestra aleatoria de tamaño n = 1000 de la v.a. X corremos la siguiente instrucción

```
set.seed(7563) # Para poder reproducir la muestra en el futuro
muestra <- rnorm(n = 1000, mean = 175, sd = 6)
length(muestra); mdf <- as.data.frame(muestra)
```

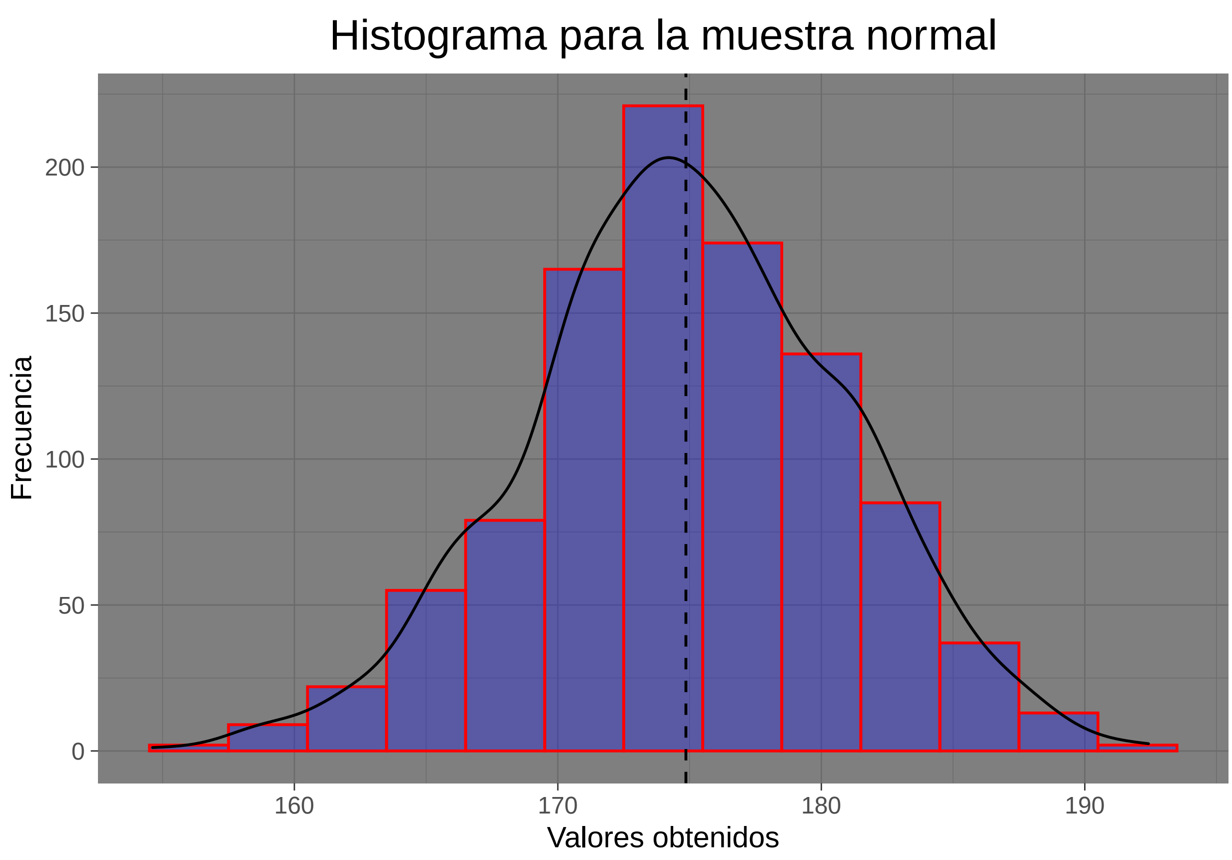
```
[1] 1000
```

```r
tail(mdf)
```

```
     muestra
995  164.7160
996  177.3317
997  156.9474
998  177.1626
999  171.3045
1000 172.4188
```

Observamos que el histograma de la muestra generada tiene forma de campana similar a la densidad de una normal

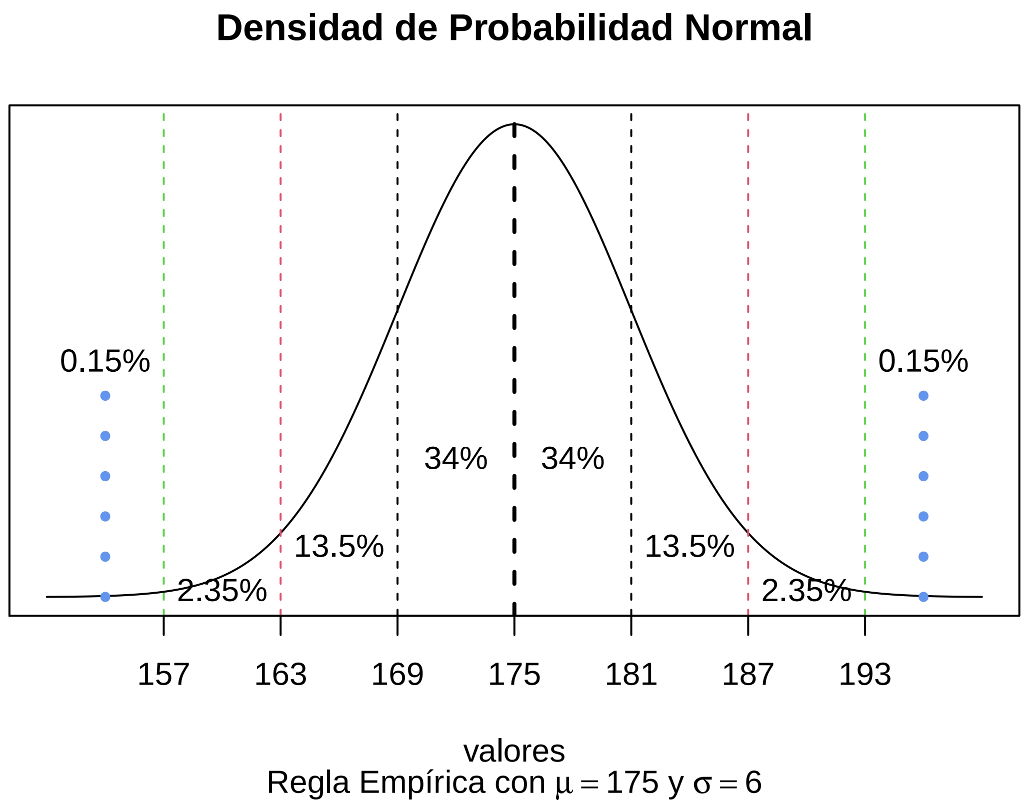
```r
ggplot(mdf, aes(muestra)) +
  geom_histogram(colour = 'red',
                 fill = 'blue',
                 alpha = 0.3, # Intensidad del color fill
                 binwidth = 3) +
  geom_density(aes(y = 3*..count..))+
  geom_vline(xintercept = mean(mdf$muestra), linetype="dashed", color = "black") +
  ggtitle('Histograma para la muestra normal') +
  labs(x = 'Valores obtenidos', y = 'Frecuencia')+
  theme_dark() +
  theme(plot.title = element_text(hjust = 0.5, size = 16))
```



REGLA EMPÍRICA

La regla empírica es una abreviatura utilizada para recordar el porcentaje de valores que se encuentran dentro de una banda alrededor de la media en una distribución normal con un ancho de dos, cuatro y seis veces la desviación típica, respectivamente.

```r
mean <- 175; sd <- 6
x <- seq(mean-4*sd, mean+4*sd, 0.01)
y <- dnorm(x, mean, sd)
plot(x, y, type = "l",
     xlab="valores",
     ylab = "",
     xaxt = "n",
     yaxt = "n")
title(main = "Densidad de Probabilidad Normal",
      sub = expression(paste("Regla Empírica con ",
                             mu == 175,
                             " y ",
                             sigma == 6)))
abline(v=mean, lty = 2, lwd = 2)
for(k in c(-3, -2, -1, 1, 2, 3)) abline(v = mean+k*sd, lty = 2, col = abs(k))
ps <- c(mean - 3*sd, mean - 2*sd,
        mean - sd, mean,
        mean + sd,
        mean + 2*sd,
        mean + 3*sd)
axis(side = 1, at = ps)
x0 <- NULL
for(i in 1:length(ps)-1) x0 <- c(x0, (ps[i]+ps[i+1])/2)
y0 <- dnorm(x0, mean, sd)*1/3
text(x = x0, y = y0, labels = c("2.35%", "13.5%", "34%", "34%", "13.5%", "2.35%"))
x1 <- (x[1]+ps[1])/2; y1 <- dnorm(mean, mean, sd)*1/2
xf <- (x[length(x)]+ps[length(ps)])/2; yf <- dnorm(mean, mean, sd)*1/2
text(x = c(x1, xf), y = c(y1, yf), labels = c("0.15%", "0.15%"))
segments(x0 = x1, y0 = 0, x1 = x1, y1 = y1,               # Draw one line as in Example 1
         col = "cornflowerblue",                               # Color of line
         lwd = 5,                                              # Thickness of line
         lty = "dotted")
segments(x0 = xf, y0 = 0, x1 = xf, y1 = yf,
         col = "cornflowerblue",
         lwd = 5,
         lty = "dotted")
```



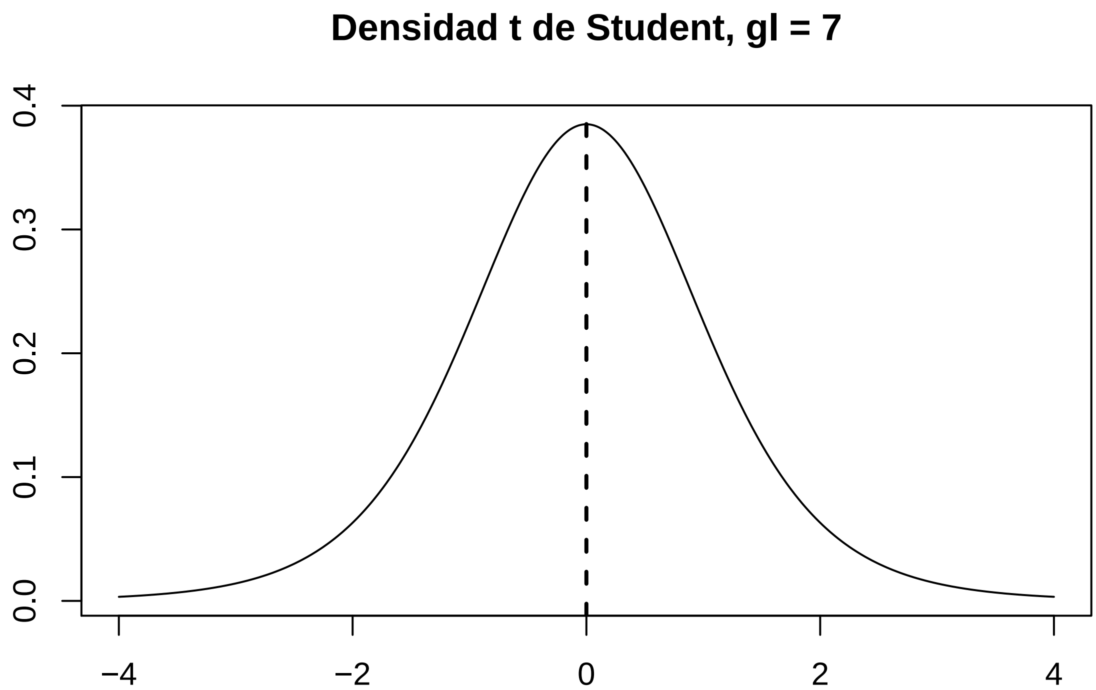
DISTRIBUCIÓN t DE STUDENT

En R para calcular valores de las funciones de densidad, distribución o cuantiles de la distribución t de Student (continua), usamos las funciones dt, pt y qt respectivamente. Para generar muestras aleatorias de esta distribución utilizamos la función rt.

Consideremos una variable aleatoria (v.a.) T que se distribuye como t de Student con 7 grados de libertad (gl) (parámetro gl = 7)

FUNCIÓN DE DENSIDAD

```
x <- seq(-4, 4, 0.01) # Algunos valores que puede tomar la v.a.
# T con 7 gl
y <- dt(x, df = 7) # Valores correspondientes de
# la densidad t de Student con 7 gl
plot(x, y, type = "l", main = "Densidad t de Student, gl = 7", xlab="", ylab="")
abline(v = 0, lwd=2, lty=2)
```



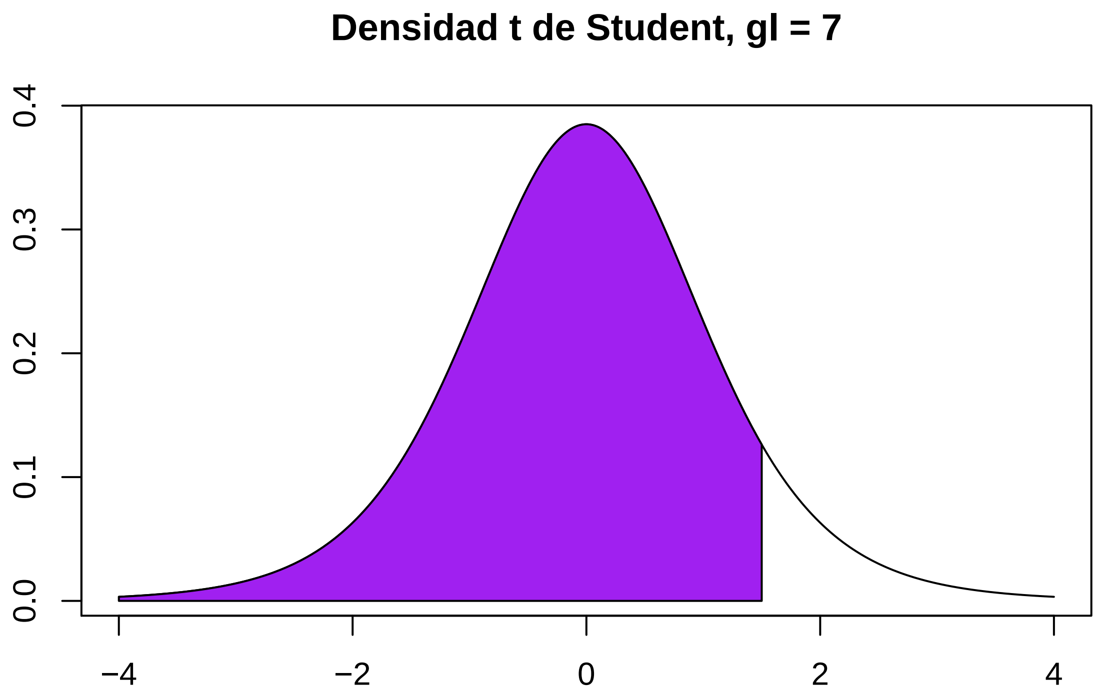
Para encontrar P(T <= 1.5), ejecutamos la siguiente instrucción

```
pt(q = 1.5, df = 7)
```

```
[1] 0.9113508
```

Observemos la región que corresponde a esta probabilidad en la siguiente gráfica

```
plot(x, y,
     type = "l",
     main = "Densidad t de Student, gl = 7",
     xlab="",
     ylab="")
polygon(c(min(x),x[x<=1.5], 1.5),
        c(0, y[x<=1.5], 0),
        col="purple")
```



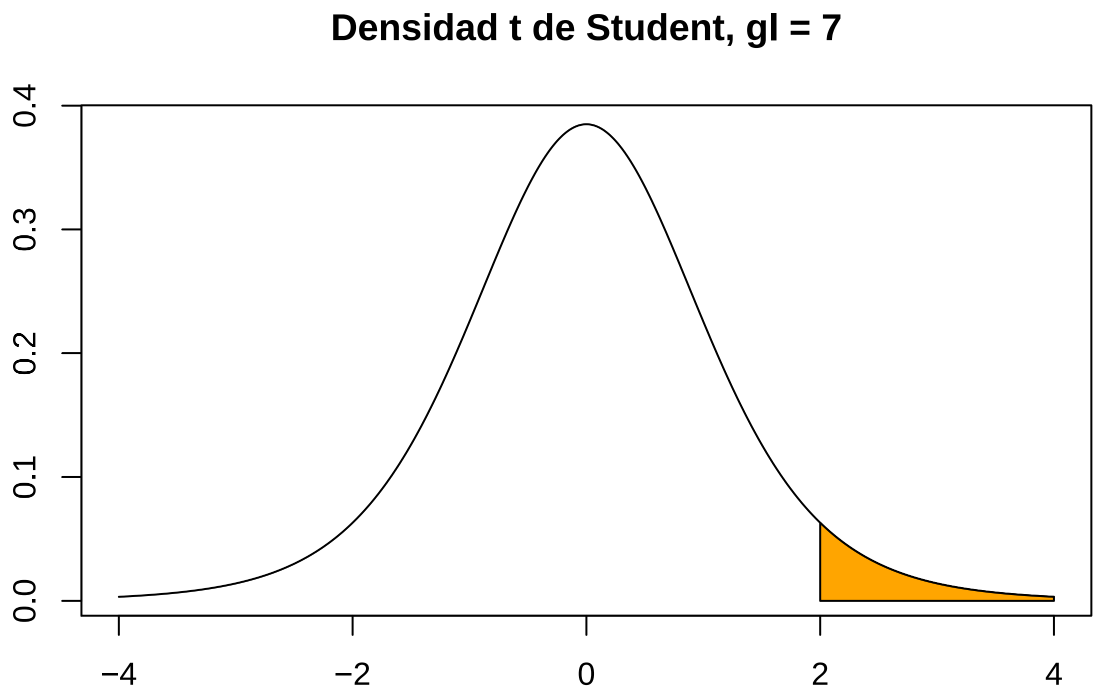
Para encontrar P(T >= 2), ejecutamos

```
pt(q = 2, df = 7, lower.tail = FALSE)
```

```
[1] 0.04280966
```

Observemos la región que corresponde a esta probabilidad en la siguiente gráfica

```
plot(x, y, type = "l",
     main = "Densidad t de Student, gl = 7",
     xlab="", ylab="")
polygon(c(2, x[x>=2], max(x)),
        c(0, y[x>=2], 0),
        col="orange")
```



CUANTILES

Para encontrar el número d tal que P(T <= d) = 0.025, es decir, el cuantil de orden 0.025, corremos la siguiente instrucción

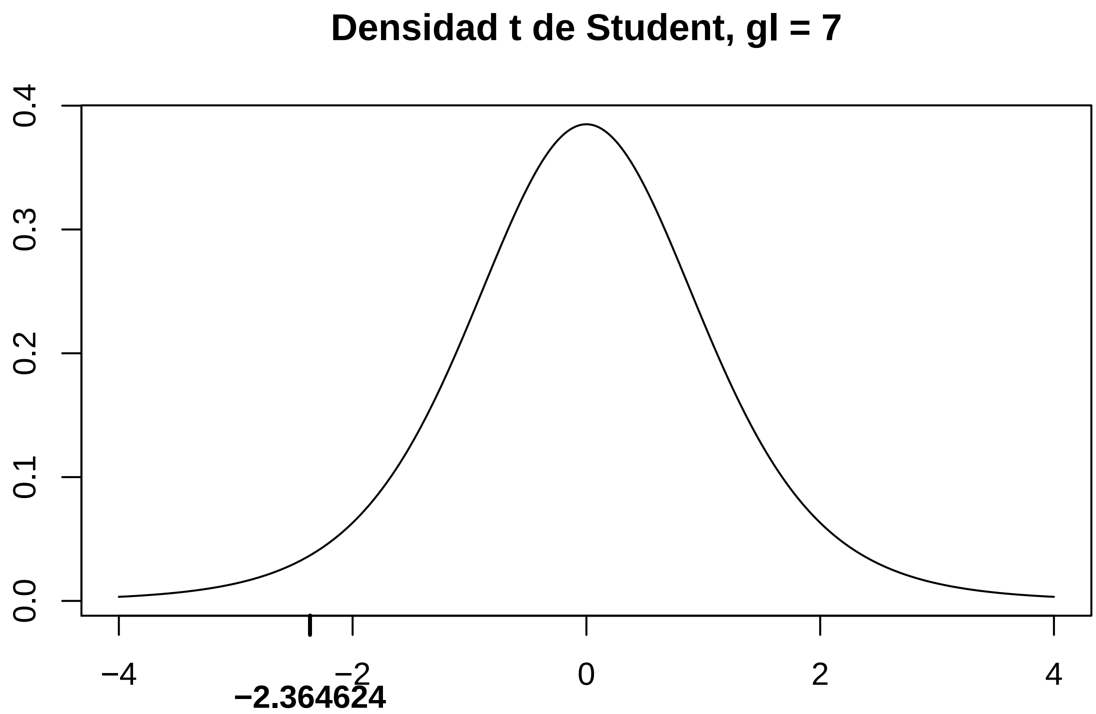
```
(d <- qt(p = 0.025, df = 7))
```

```
[1] -2.364624
```

```
#comprobando
pt(q = d, df = 7)
```

```
[1] 0.025
```

```
# Mostramos el cuantil encontrado en el eje de medición (eje horizontal)
plot(x, y,
     type = "l",
     main = "Densidad t de Student, gl = 7",
     xlab="",
     ylab="")
axis(side = 1,
     at = d,
     font = 2,
     padj = 1,
     lwd = 2)
```



MUESTRAS ALEATORIAS

Para generar una muestra aleatoria de tamaño n = 1000 de la v.a. T corremos la siguiente instrucción

```
set.seed(777) # Para poder reproducir la muestra en el futuro
muestra <- rt(n = 1000, df = 7)
length(muestra); mdf <- as.data.frame(muestra)
```

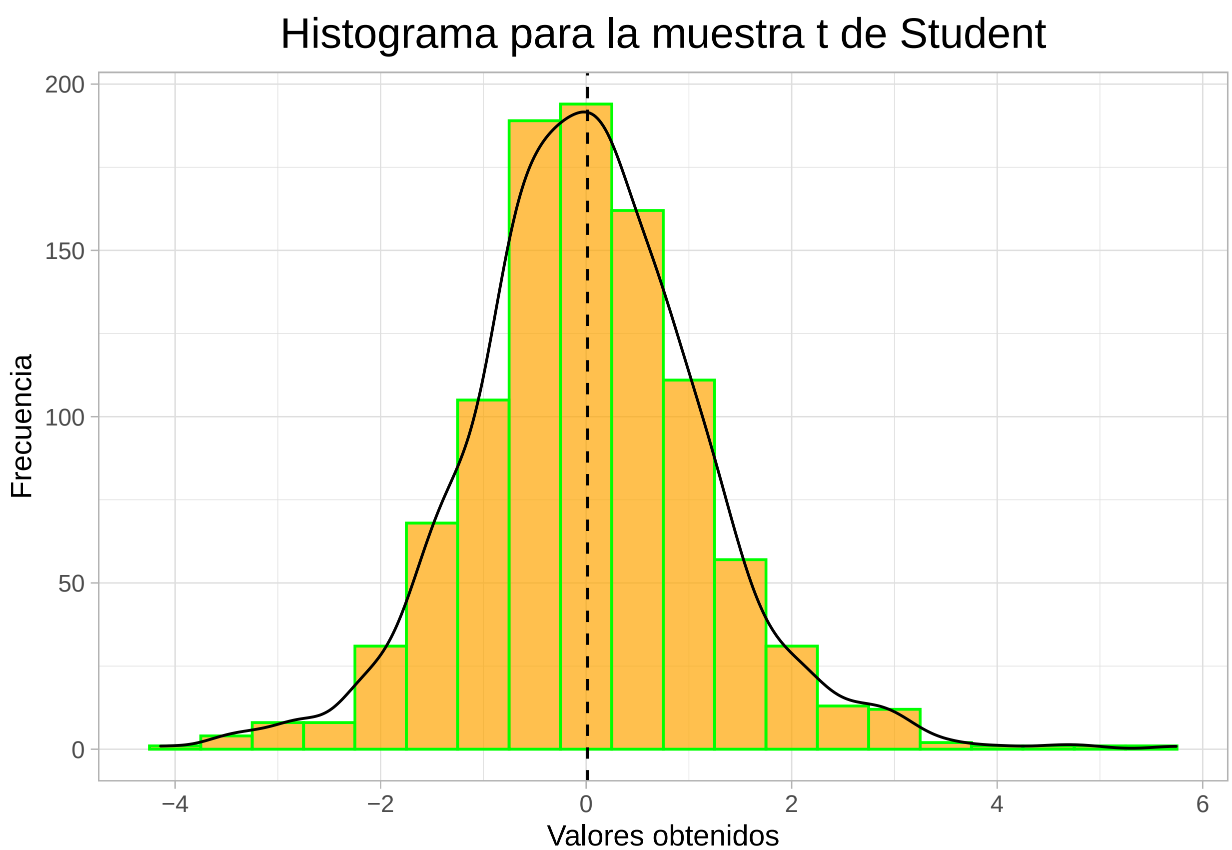
```
[1] 1000
```

```
tail(mdf)
```

```
           muestra
995   0.6257672958
996  -0.2794513280
997  -0.3037697310
998   0.4514007565
999  -0.0003503659
1000 -0.1808370397
```

Observamos que el histograma de la muestra generada tiene forma de campana similar a la densidad t de Student

```r
ggplot(mdf, aes(muestra)) +
  geom_histogram(colour = 'green',
                 fill = 'orange',
                 alpha = 0.7, # Intensidad del color fill
                 binwidth = 0.5) +
  geom_density(aes(y = 0.5*..count..))+
  geom_vline(xintercept = mean(mdf$muestra), linetype="dashed", color = "black") +
  ggtitle('Histograma para la muestra t de Student') +
  labs(x = 'Valores obtenidos', y = 'Frecuencia')+
  theme_light() +
  theme(plot.title = element_text(hjust = 0.5, size = 16))
```



ATENUAR COLORES

Para atenuar colores es necesario utilzar $alpha = 0.8$ o la cantidad porcentual para atenuarse.

RETO 1. DISTRIBUCIONES BINOMIAL, NORMAL Y T DE STUDENT

OBJETIVO

- Calcular probabilidades y cuantiles relacionadas con algunas distribuciones de probabilidad útiles y comunes
- Generar muestras aleatorias que provengan de las distribuciones estudiadas

REQUISITOS

- Haber trabajado con el Prework y el Work

DESARROLLO

DISTRIBUCIÓN BINOMIAL

Consideremos un experimento binomial con $n = 35$ pruebas idénticas e independientes, en donde la probabilidad de éxito en cada prueba es $p = 0.51$. Encuentre lo siguiente:

1. La probabilidad de observar exactamente 10 éxitos
2. La probabilidad de observar 10 o más exitos
3. El cuantil de orden 0.5
4. Genere una muestra aleatoria de tamaño 1000 de esta distribución, construya una tabla de frecuencias relativas con los resultados y realice el gráfico de barras de los resultados que muestre las frecuencias relativas.

DISTRIBUCIÓN NORMAL

Considere una variable aleatoria normal con media 110 y desviación estándar 7. Realice lo siguiente:

1. Grafique la función de densidad de probabilidad
2. Encuentre la probabilidad de que la v.a. sea mayor o igual a 140
3. Encuentre el cuantil de orden 0.95
4. Genere una muestra aleatoria de tamaño 1000 y realice el histograma de frecuencias relativas para esta muestra

```
#RETO 1 SESSION 4


##distribucion binomial
# Consideremos un experimento binomial con $n = 35$ pruebas idénticas e
# independientes, en donde la probabilidad de éxito en cada prueba es
# $p = 0.51$. Encuentre lo siguiente:
#
# 1. La probabilidad de observar exactamente 10 éxitos
# 2. La probabilidad de observar 10 o más exitos
# 3. El cuantil de orden 0.5
# 4. Genere una muestra aleatoria de tamaño 1000 de esta
# distribución, construya una tabla de frecuencias relativas con los
```

```r
# resultados y realice el gráfico de barras de los resultados que
# muestre las frecuencias relativas.

library(ggplot2)
#PUNTO 1
dbinom(x=10,size = 35,prob = 0.51)
```

```
[1] 0.003930318
```

```r
#PUNTO 2
pbinom(q=9,size = 35,prob = 0.51,
       lower.tail = FALSE)
```

```
[1] 0.9979185
```

```r
#PUNTO 3
qbinom(p = 0.5,size = 35,prob = 0.51)
```

```
[1] 18
```

```r
# PUNTO 4
set.seed(123)
ale1000<- rbinom(n = 1000,size = 35,prob = 0.51)
class(ale1000)
```

```
[1] "integer"
```

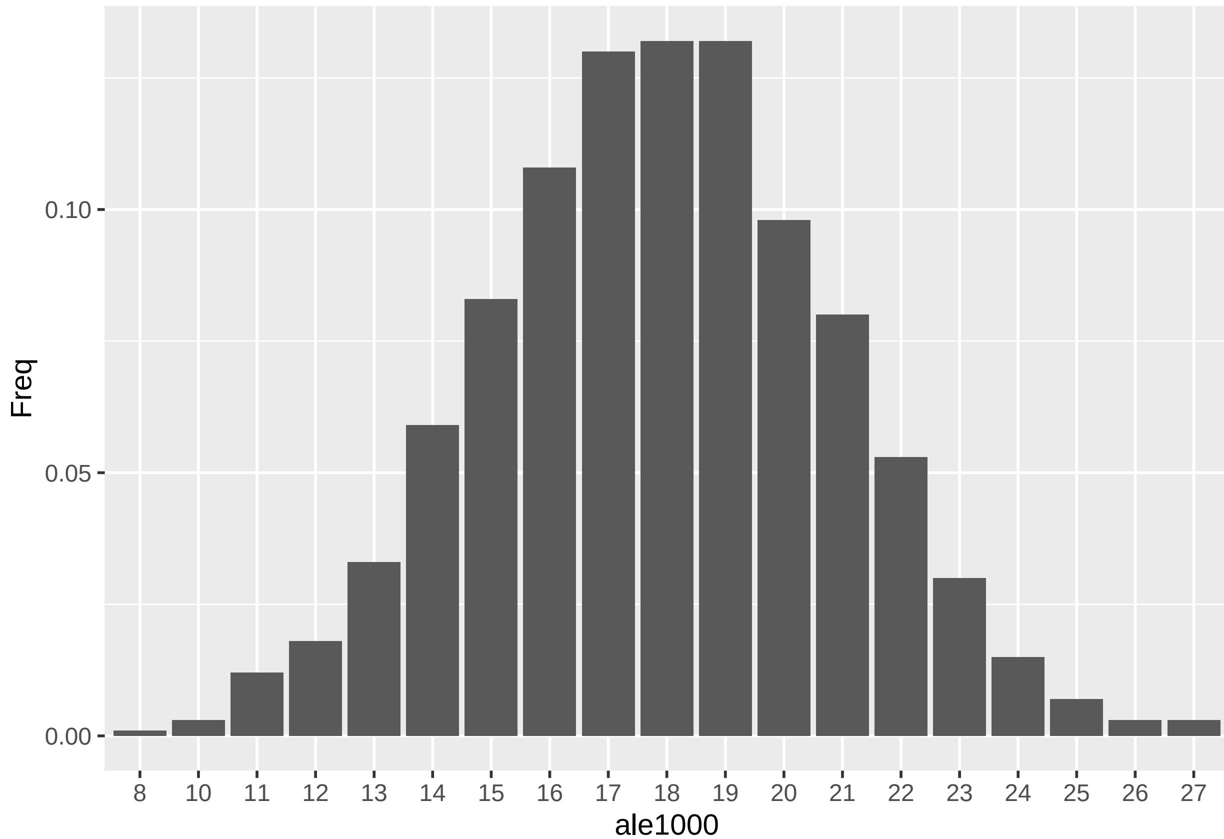
```r
ale1000dataaframe<-as.data.frame(table(ale1000)/length(ale1000))
head(ale1000dataaframe)
```

```
  ale1000  Freq
1       8 0.001
2      10 0.003
3      11 0.012
4      12 0.018
5      13 0.033
6      14 0.059
```

```r
tail(ale1000dataaframe)
```

```
   ale1000  Freq
14      22 0.053
15      23 0.030
16      24 0.015
17      25 0.007
18      26 0.003
19      27 0.003
```

```r
ggplot(ale1000dataaframe,
       aes(x = ale1000,y = Freq))+geom_bar(stat = "identity")
```




```r
##DISTRIBUCION NORMAL
# Considere una variable aleatoria normal con media 110
#y desviación estándar 7. Realice lo siguiente:
# 1. Grafique la función de densidad de probabilidad
# 2. Encuentre la probabilidad de que la v.a. sea mayor o igual a 140
# 3. Encuentre el cuantil de orden 0.95
# 4. Genere una muestra aleatoria de tamaño 1000 y realice el histograma
# de frecuencias relativas para esta muestra

#PUNTO 1

#GENERANDO LOS VALORES ALEATORIOS
vala1<- seq(10,100,by = 0.1)
View(vala1)
vala2<- dnorm(x = vala1,mean = 110,sd = 7)
class(vala1)
```

```
[1] "numeric"
```

```r
class(vala2)
```

```
[1] "numeric"
```

```r
df<-data.frame(vala1,vala2)
class(df)
```

```
[1] "data.frame"
```

```r
View(df)
grafico<-ggplot(df,aes(vala1,vala2))+geom_line()


#PUNTO 2 PROBABILIDAD DE QUE VARIABLE ALEATORIA SEA MAYOR O IGUAL A 140
#LOWR TAIL SI ES VERDADERO LA PROBABILIDAD SERÁ P(X<=x) SI ES FALSO ES P(X>x)
pnorm(q = 140,mean = 110,sd = 7,lower.tail = FALSE)
```

```
[1] 9.107649e-06
```

```r
#PUNTO 3. CUARTIL DE ORDEN 0.95
qnorm(p = 0.95,mean = 110,sd = 7)
```

```
[1] 121.514
```

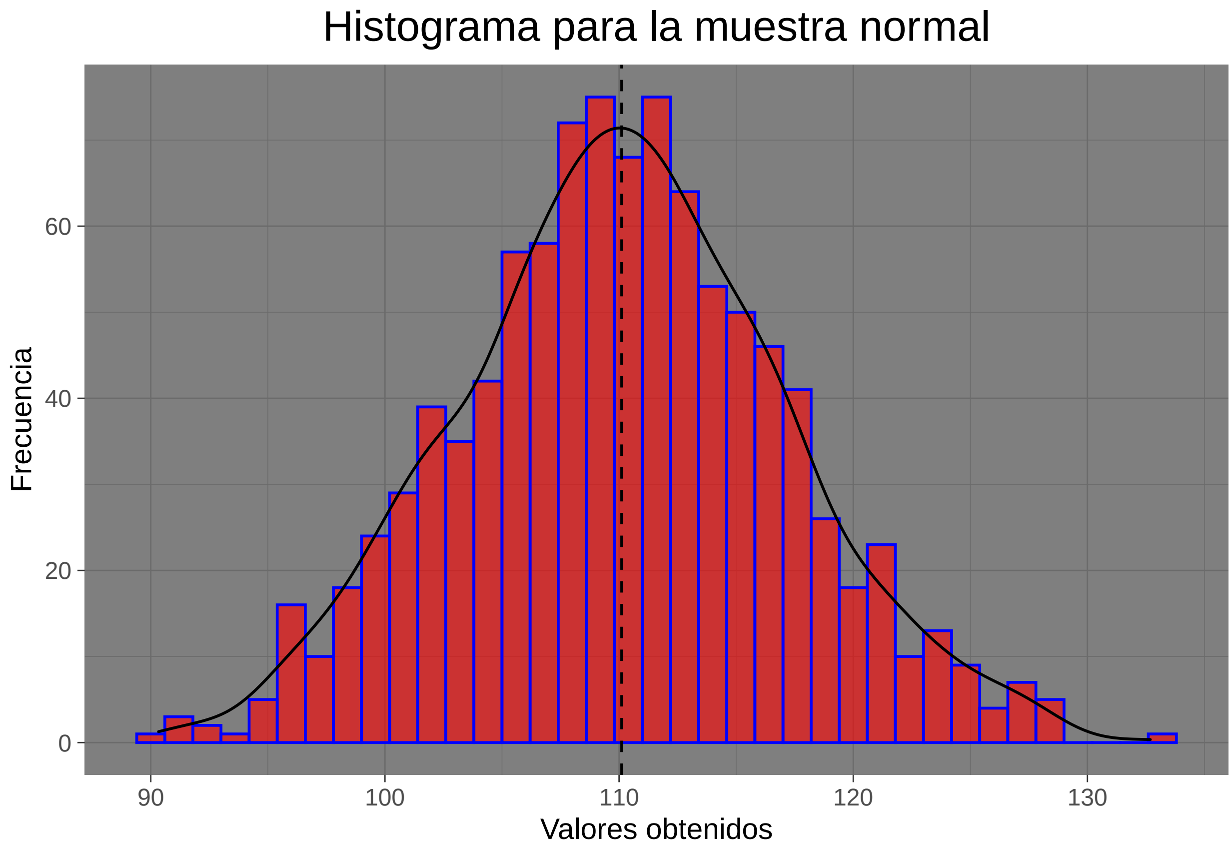
```r
#PUNTO 4.MUESTRA ALEATORIA DE TAMAÑO 1000 Y GENERE HISTOGRAMA DE FRECUENCIAS
set.seed(123)
datara<-rnorm(n = 1000,mean = 110,sd = 7)
dataradf<-as.data.frame(x = datara)
class(datara)
```

```
[1] "numeric"
```

```r
class(dataradf)
```

```
[1] "data.frame"
```

```r
ggplot(data = dataradf,aes(datara))+
  geom_histogram(colour="blue",
                 fill="red",
                 alpha=0.6,
                 binwidth = 1.2)+
geom_density(aes(y=1.2*..count..))+
  geom_vline(xintercept = mean(datara),
             linetype="dashed", color = "black")+
  ggtitle(label = 'Histograma para la muestra normal')+
  labs(x='Valores obtenidos',y='Frecuencia')+
  theme_dark()+
  theme(plot.title = element_text(hjust = 0.5, size = 16))
```



TEOREMA CENTRAL DEL LÍMITE

OBJETIVO

- Comprender lo que afirma el teorema central del límite obteniendo muestras aleatorias de diferentes tamaños en R y observando la manera en la que se distribuyen las medias de las muestras generadas

REQUISITOS

- Tener R y RStudio instalados
- Haber estudiado el Prework

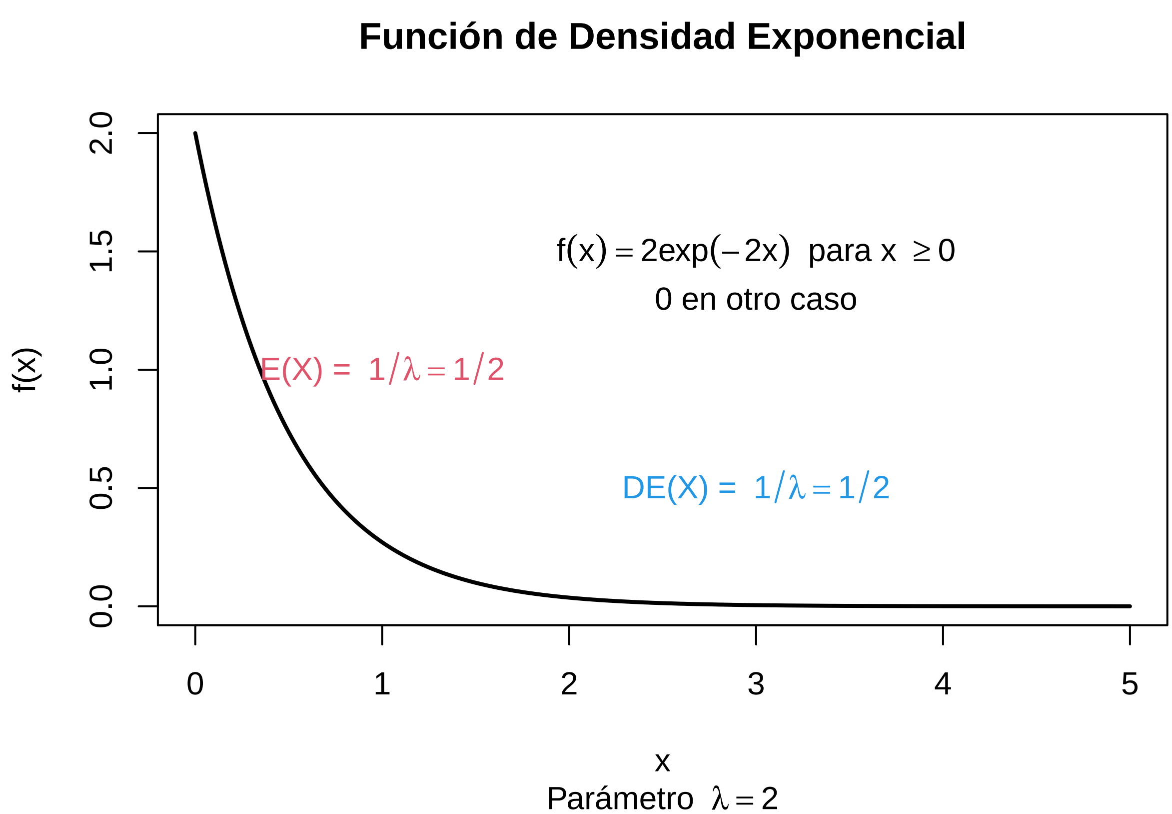
##DESARROLLO

El teorema central del límite (TCL) es una teoría estadística que establece que, dada una muestra suficientemente grande de la población, la distribución de las medias muestrales seguirá una distribución normal.

Además, el TCL afirma que a medida que el tamaño de la muestra se incrementa, la media muestral se acercará a la media de la población. Por tanto, mediante el TCL podemos definir la distribución de la media muestral de una determinada población con una varianza conocida. De manera que la distribución seguirá una distribución normal si el tamaño de la muestra es lo suficientemente grande.

Cargamos el paquete ggplot2 para hacer algunas gráficas

```
#TEOREMA CENTRAL DEL LIMITE


library(ggplot2)
# Consideremos una variable aleatoria (v.a.) X
# con distribución exponencial y parámetro lambda = 2
x <- seq(0, 5, 0.02)
plot(x, dexp(x, rate = 2), type = "l", lwd = 2, ylab = "")
title(main = "Función de Densidad Exponencial", ylab = "f(x)",
      sub = expression("Parámetro " ~ lambda == 2))
text(x = 3, y = 1.5, labels = expression(f(x)==2*exp(-2*x) ~ " para x "  >= 0))
text(x = 3, y = 1.3, labels = paste("0 en otro caso"))
text(x = 1, y = 1, labels = expression("E(X) = " ~ 1/lambda == 1/2), col = 2)
text(x = 3, y = 0.5, labels = expression("DE(X) = " ~ 1/lambda == 1/2), col = 4)
```




```
print("Ahora obtenemos una muestra aleatoria de tamaño n = 4 de la distribución exponencial considerada
```

```
[1] "Ahora obtenemos una muestra aleatoria de tamaño n = 4 de la distribución exponencial considerada"
```

```
set.seed(10) # Para reproducir posteriormente la muestra
(m1.4 <- rexp(n = 4, rate = 2))
```

```
[1] 0.007478203 0.460110602 0.376079469 0.787520925
```

```r
print("Obtenemos la media de la muestra generada")
```

```
[1] "Obtenemos la media de la muestra generada"
```

```r
mean(m1.4)
```

```
[1] 0.4077973
```

```r
print("Ahora obtenemos 5 muestras de tamaño 3")
```

```
[1] "Ahora obtenemos 5 muestras de tamaño 3"
```

```r
set.seed(64) # Para reproducir las muestras en el futuro
(m5.3 <- sapply(X = rep(3, 5), FUN = rexp, 2))
```

```
          [,1]        [,2]       [,3]      [,4]       [,5]
[1,] 1.5874413 0.031138339 0.08650897 0.3297026 0.82773066
[2,] 0.3272063 0.120168796 0.59122902 0.1333955 0.30709894
[3,] 1.5801568 0.006056407 1.08174334 0.5965121 0.02107677
```

```r
print("Obtenemos las medias de las 5 muestras")
```

```
[1] "Obtenemos las medias de las 5 muestras"
```

```r
(media5.3 <- apply(m5.3, 2, mean))
```

```
[1] 1.16493482 0.05245451 0.58649378 0.35320341 0.38530212
```

```r
print("Ahora obtenemos 1000 muestras de tamaño 7 y las 1000 medias correspondientes a las muestras")
```

```
[1] "Ahora obtenemos 1000 muestras de tamaño 7 y las 1000 medias correspondientes a las muestras"
```

```r
set.seed(465) # Para reproducir las muestras en el futuro
m1000.7 <- sapply(X = rep(7, 1000), FUN = rexp, 2)
media1000.7 <- apply(m1000.7, 2, mean)
mdf <- as.data.frame(media1000.7)
tail(mdf)
```

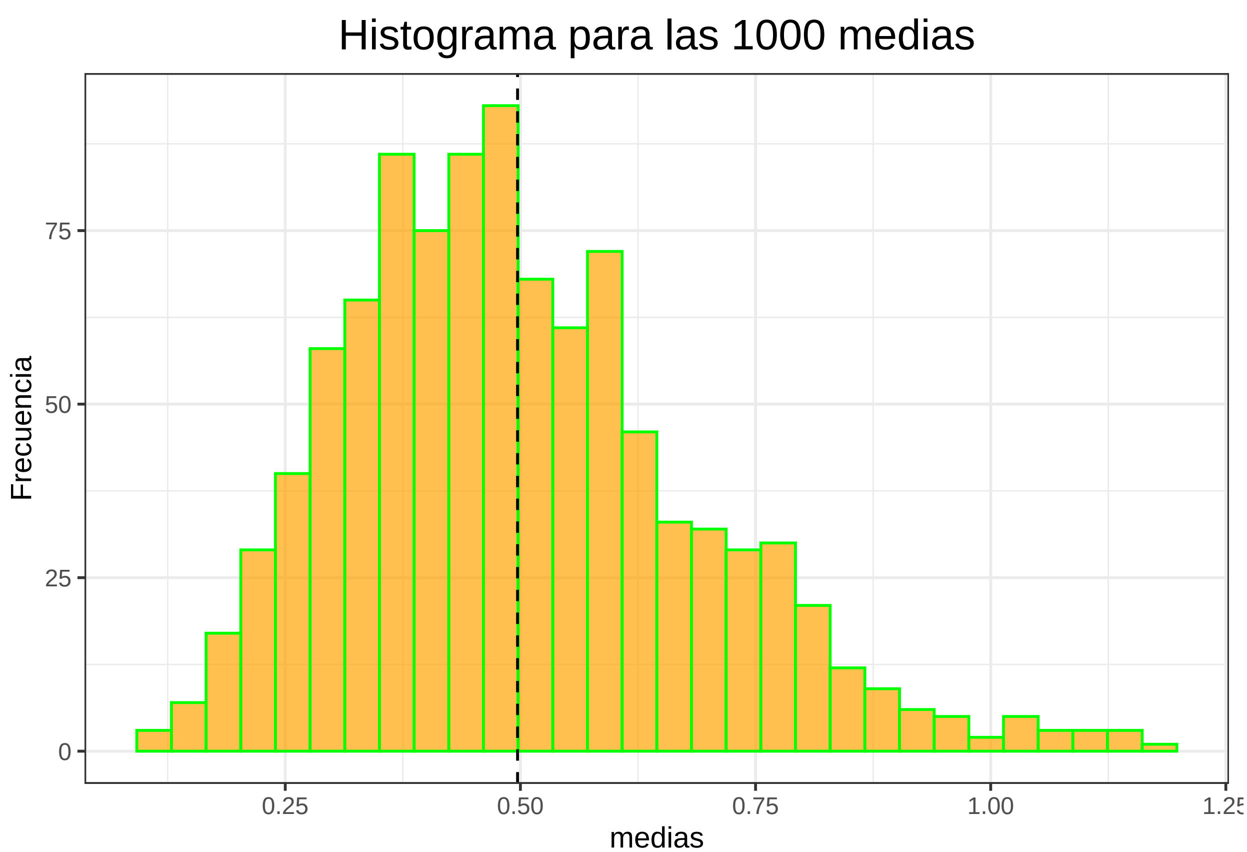
```
     media1000.7
995    0.5444073
996    0.3751581
997    0.5589332
998    0.6565002
999    0.2337858
1000   0.3259742
```

```r
print("Observamos que el histograma de las medias tiene forma de campana")
```

```
[1] "Observamos que el histograma de las medias tiene forma de campana"
```

```r
ggplot(mdf, aes(media1000.7)) +
  geom_histogram(colour = 'green',
                 fill = 'orange',
                 alpha = 0.7) + # Intensidad del color fill
  geom_vline(xintercept = mean(media1000.7), linetype="dashed", color = "black") +
  ggtitle('Histograma para las 1000 medias') +
  labs(x = 'medias', y = 'Frecuencia')+
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5, size = 16))
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```r
mean(media1000.7); 1/2 # Media de las 1000 medias y media de
```

```
[1] 0.4969754
```

```
[1] 0.5
```

```r
# la población de la cual vienen las 1000 muestras
sd(media1000.7); (1/2)/sqrt(7) # DE de las 1000 medias y DE
```

```
[1] 0.1854891
```

```
[1] 0.1889822
```

```r
# de la población de la cual vienen las 1000 muestras dividida
# por la raíz del tamaño de la muestra

print("Ahora obtenemos 1000 muestras de tamaño 33 y las 1000 medias correspondientes a las muestras")
```

```
[1] "Ahora obtenemos 1000 muestras de tamaño 33 y las 1000 medias correspondientes a las muestras"
```

```r
set.seed(4465) # Para reproducir las muestras en el futuro
m1000.33 <- sapply(X = rep(33, 1000), FUN = rexp, 2)
media1000.33 <- apply(m1000.33, 2, mean)
mdf <- as.data.frame(media1000.33)
tail(mdf)
```

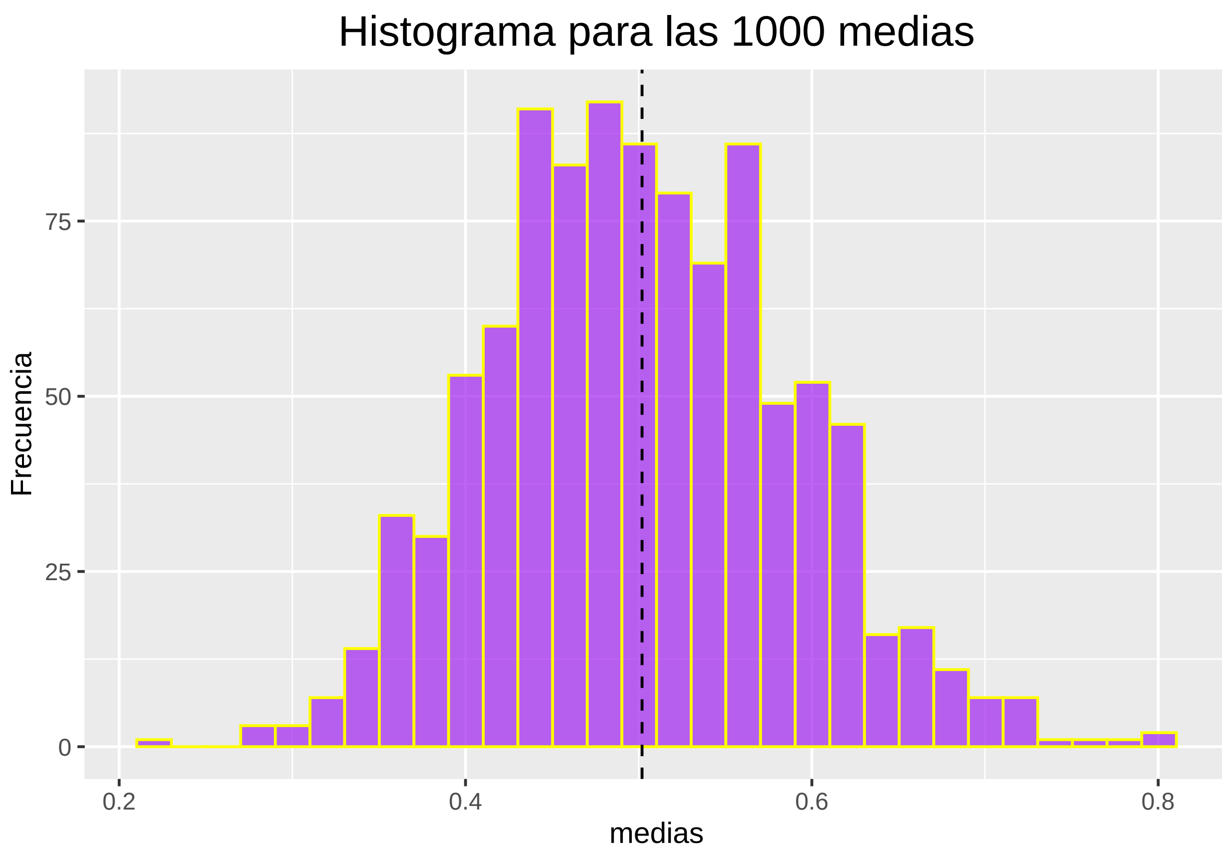
```
     media1000.33
995     0.3818621
996     0.3609060
997     0.5153507
998     0.5261520
999     0.5053655
1000    0.4573147
```

```r
print("Observamos que el histograma de las medias es más parecida todavía a una campana")
```

```
[1] "Observamos que el histograma de las medias es más parecida todavía a una campana"
```

```r
ggplot(mdf, aes(media1000.33)) +
  geom_histogram(colour = 'yellow',
                 fill = 'purple',
                 alpha = 0.7) + # Intensidad del color fill
  geom_vline(xintercept = mean(media1000.33), linetype="dashed", color = "black") +
  ggtitle('Histograma para las 1000 medias') +
  labs(x = 'medias', y = 'Frecuencia')+
  theme_get() +
  theme(plot.title = element_text(hjust = 0.5, size = 16))
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```r
mean(media1000.33); 1/2 # Media de las 1000 medias y media de la
```

```
[1] 0.501952
```

```
[1] 0.5
```

```r
# población de la cual vienen las 1000 muestras
sd(media1000.33); (1/2)/sqrt(33) # DE de las 1000 medias
```

```
[1] 0.08624136
```

```
[1] 0.08703883
```

```r
# y DE de la población de la cual vienen las 1000 muestras dividida
# por la raíz del tamaño de la muestra

print("Ahora obtenemos 1000 muestras de tamaño 400 y las 1000 medias correspondientes a las muestras")
```

```
[1] "Ahora obtenemos 1000 muestras de tamaño 400 y las 1000 medias correspondientes a las muestras"
```

```r
set.seed(543465) # Para reproducir las muestras en el futuro
m1000.400 <- sapply(X = rep(400, 1000), FUN = rexp, 2)
media1000.400 <- apply(m1000.400, 2, mean)
mdf <- as.data.frame(media1000.400)
tail(mdf)
```

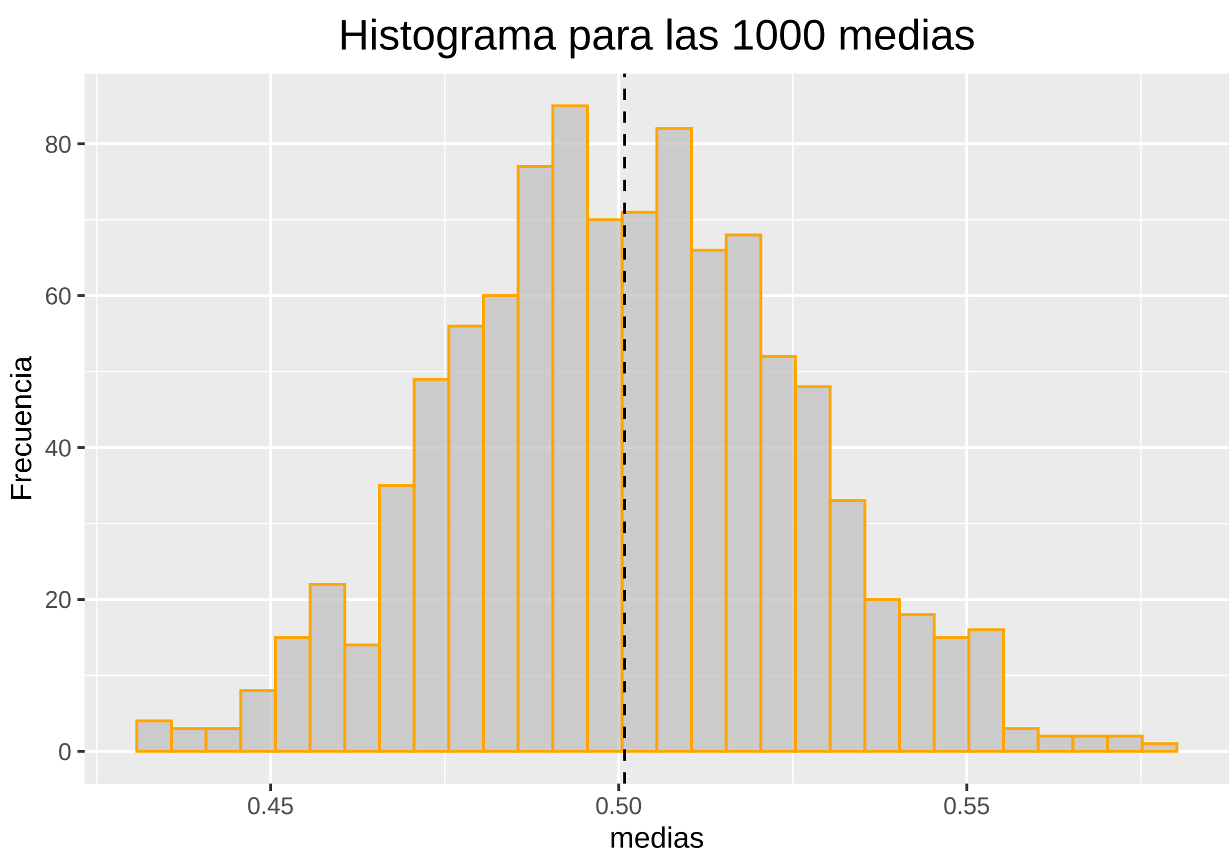
```
     media1000.400
995      0.4656883
996      0.4779040
997      0.4803765
998      0.4944309
999      0.4988992
1000     0.5106146
```

```r
print("Observamos que el histograma de las medias es más parecida todavía a una campana")
```

```
[1] "Observamos que el histograma de las medias es más parecida todavía a una campana"
```

```r
ggplot(mdf, aes(media1000.400)) +
  geom_histogram(colour = 'orange',
                 fill = 'gray',
                 alpha = 0.7) + # Intensidad del color fill
  geom_vline(xintercept = mean(media1000.400), linetype="dashed", color = "black") +
  ggtitle('Histograma para las 1000 medias') +
  labs(x = 'medias', y = 'Frecuencia')+
  theme_gray() +
  theme(plot.title = element_text(hjust = 0.5, size = 16))
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```r
mean(media1000.400); 1/2 # Media de las 1000 medias y media
```

```
[1] 0.5008446
```

```
[1] 0.5
```

```r
# de la población de la cual vienen las 1000 muestras
sd(media1000.400); (1/2)/sqrt(400) # DE de las 1000 medias
```

```
[1] 0.02453529
```

```
[1] 0.025
```

```r
# y DE de la población de la cual vienen las 1000 muestras
# dividida por la raíz del tamaño de la muestra
```

EJEMPLO 3. ALGUNOS ESTIMADORES PUNTUALES INSESGADOS COMUNES

OBJETIVO

- Entender la idea de estimador insesgado de un parámetro

REQUISITOS

- Tener `R` y RStudio instalados
- Hbaer leido el prework

DESARROLLO

Un estimador insesgado es aquel cuya esperanza matemática coincide con el valor del parámetro que sea desea estimar. En caso de no coincidir se dice que el estimador tiene sesgo. La razón de buscar un estimador insesgado es que el parámetro que deseamos estimar esté bien estimado. Es decir, si queremos estimar la media de goles por partido de determinado jugador de fútbol, hemos de utilizar una fórmula que nos proporcione un valor lo más aproximado posible al valor real.

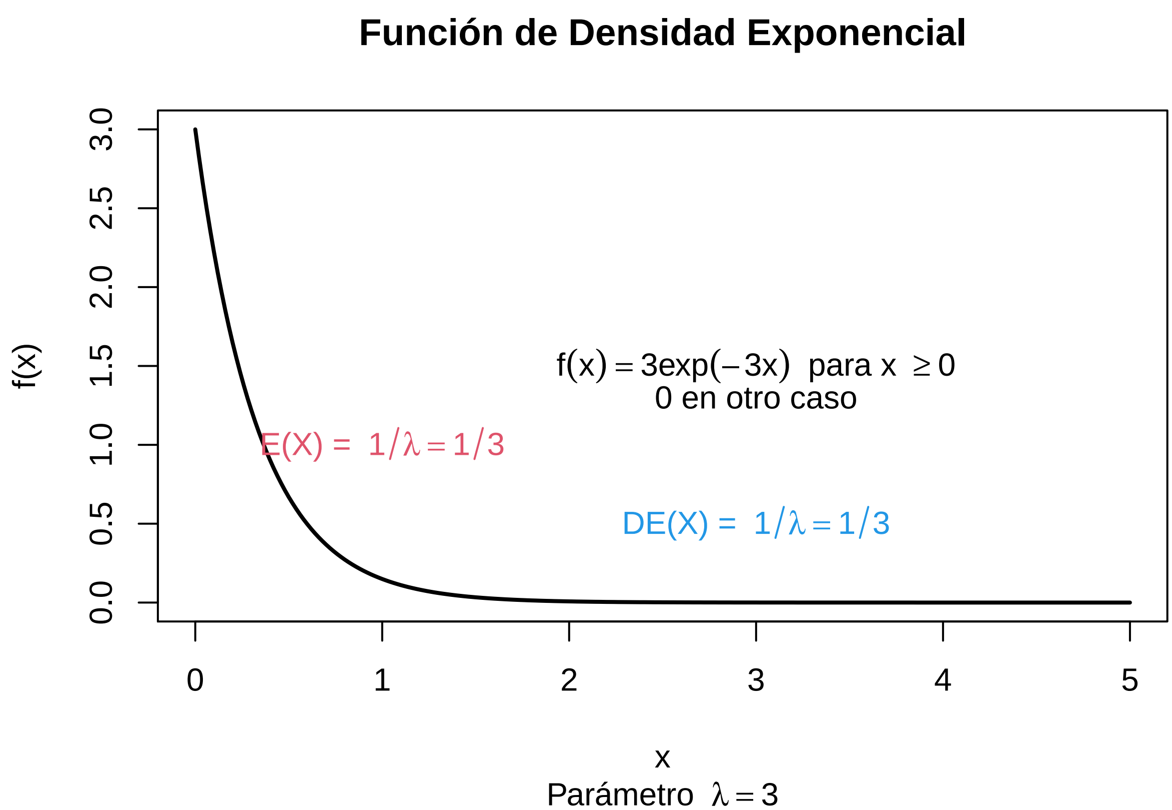
```r
#EJEMPLO 3

print("vargamos el paquete ggplot2 para hacer algunas gráficas")
```

```
[1] "vargamos el paquete ggplot2 para hacer algunas gráficas"
```

```r
library(ggplot2)

print("Consideremos una variable aleatoria X con distribución exponencial y parametro lambda=3")
```

```
[1] "Consideremos una variable aleatoria X con distribución exponencial y parametro lambda=3"
```

```r
x <- seq(0, 5, 0.02)
plot(x, dexp(x, rate = 3), type = "l", lwd = 2, ylab = "")
title(main = "Función de Densidad Exponencial", ylab = "f(x)",
      sub = expression("Parámetro " ~ lambda == 3))
text(x = 3, y = 1.5, labels = expression(f(x)==3*exp(-3*x) ~ " para x "  >= 0))
text(x = 3, y = 1.3, labels = paste("0 en otro caso"))
text(x = 1, y = 1, labels = expression("E(X) = " ~ 1/lambda == 1/3), col = 2)
text(x = 3, y = 0.5, labels = expression("DE(X) = " ~ 1/lambda == 1/3), col = 4)
```




```r
print("Obtenemos 1200 muestras aleatorias de tamaño 350 y las 1200 medias correspondientes a las muestra
```

```
[1] "Obtenemos 1200 muestras aleatorias de tamaño 350 y las 1200 medias correspondientes a las muestras
```

```r
set.seed(65) # Para reproducir las muestras en el futuro
m1200.350 <- sapply(X = rep(350, 1200), FUN = rexp, rate = 3)
media1200.350 <- apply(m1200.350, 2, mean)
mdf <- as.data.frame(media1200.350)
tail(mdf)
```

```
     media1200.350
1195     0.3757215
1196     0.3415659
1197     0.3043333
1198     0.3233880
```

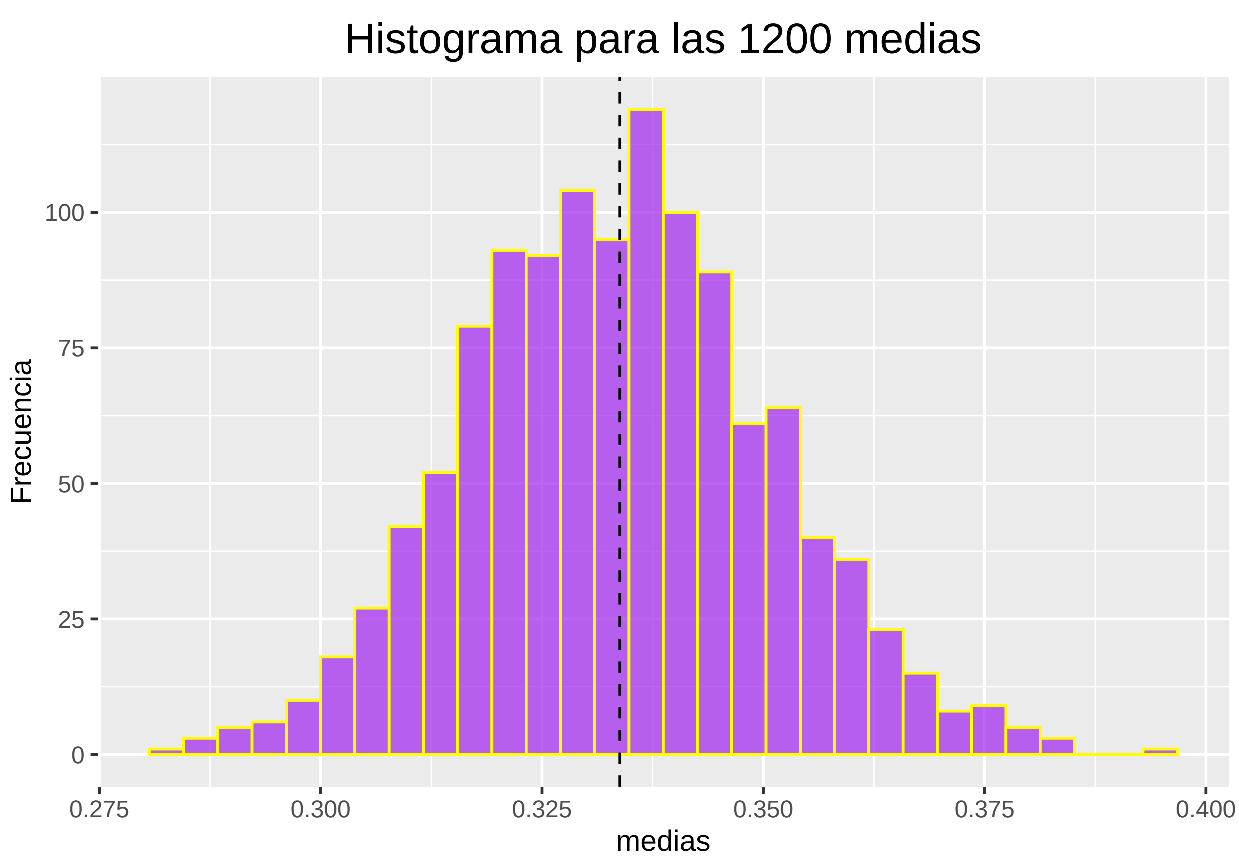
```
1199      0.3240008
1200      0.3344396
```

```
print("Observamos que el histograma de las medias tiene forma de campana")
```

```
[1] "Observamos que el histograma de las medias tiene forma de campana"
```

```
ggplot(mdf, aes(media1200.350)) +
  geom_histogram(colour = 'yellow',
                 fill = 'purple',
                 alpha = 0.7) + # Intensidad del color fill
  geom_vline(xintercept = mean(media1200.350),
             linetype="dashed",
             color = "black") +
  ggtitle('Histograma para las 1200 medias') +
  labs(x = 'medias',
       y = 'Frecuencia')+
  theme_get() +
  theme(plot.title = element_text(hjust = 0.5, size = 16))
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
mean(media1200.350); 1/3 # Media de las 1200 medias y media de
```

```
[1] 0.3337942
```

```
[1] 0.3333333
```

```
# la población de la cual provienen las 1200 muestras
sd(media1200.350); (1/3)/sqrt(350) # DE de las 1200 medias y DE
```

```
[1] 0.01714254
```

```
[1] 0.01781742
```

```
# de la población de la cual provienen las 1200 muestras dividida
# por la raíz del tamaño de las muestras




print("ENSAYO BERNOULLI Con las siguientes instrucciones obtenemos un solo valor, en donde el 0 (fracaso
```

```
[1] "ENSAYO BERNOULLI Con las siguientes instrucciones obtenemos un solo valor, en donde el 0 (fracaso)
```

```
set.seed(345)
sample(x = c(0, 1), size = 1, prob = c(0.45, 0.55))
```

```
[1] 1
```

```
rbinom(n = 1, size = 1, prob = 0.55)
```

```
[1] 1
```

```
print("Obtenemos 1000 muestras de tamaño 31 de una v.a. Bernoulli con p = 0.55")
```

```
[1] "Obtenemos 1000 muestras de tamaño 31 de una v.a. Bernoulli con p = 0.55"
```

```
set.seed(5434) # Para reproducir las muestras en el futuro
m1000.31 <- sapply(X = rep(31, 1000), FUN = function(n) sample(x = c(0, 1), size = n, replace = TRUE, pr
media1000.31 <- apply(m1000.31, 2, mean)
mdf <- as.data.frame(media1000.31)
tail(mdf)
```

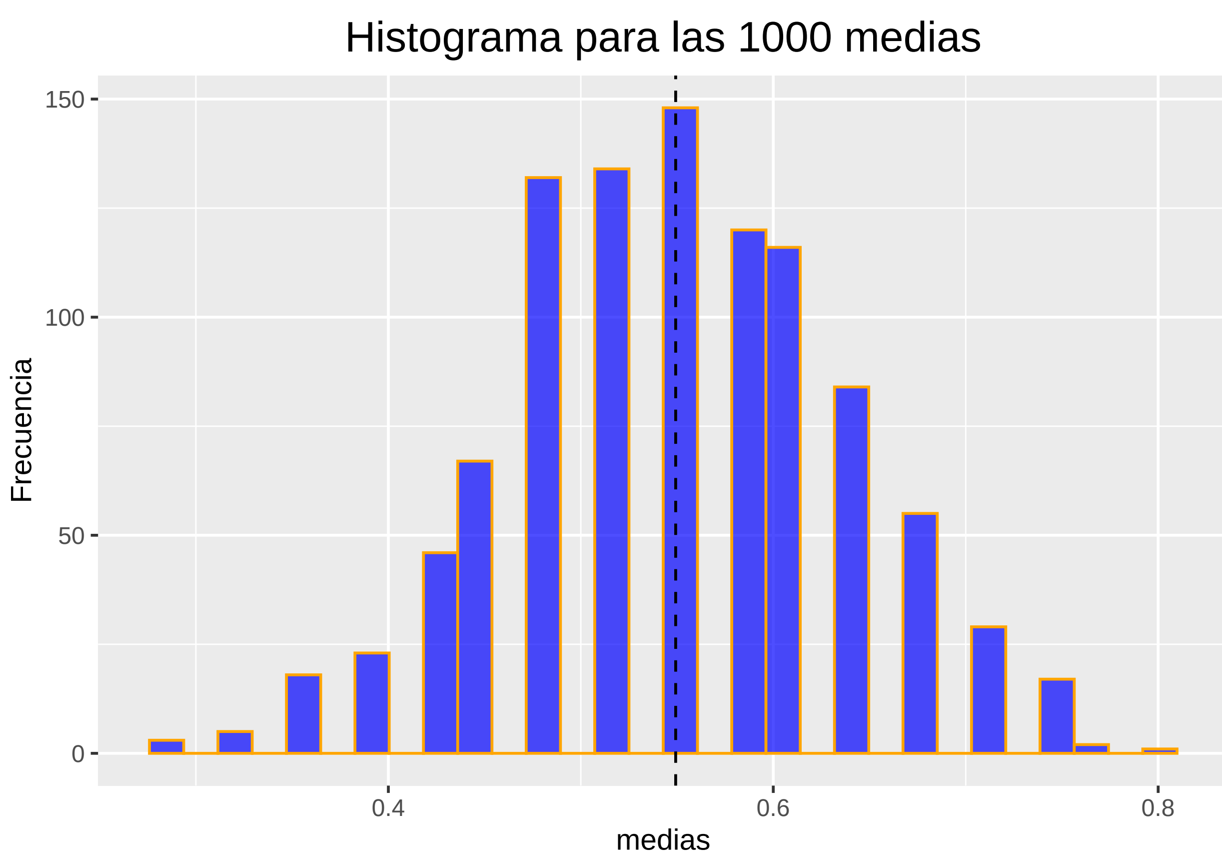
```
     media1000.31
995     0.6451613
996     0.6129032
997     0.6774194
998     0.5161290
999     0.5806452
1000    0.5161290
```

```r
print("Observamos que el histograma de las medias es parecida a una campana")
```

```
[1] "Observamos que el histograma de las medias es parecida a una campana"
```

```r
ggplot(mdf, aes(media1000.31)) +
  geom_histogram(colour = 'orange',
                 fill = 'blue',
                 alpha = 0.7) + # Intensidad del color fill
  geom_vline(xintercept = mean(media1000.31),
             linetype="dashed",
             color = "black") +
  ggtitle('Histograma para las 1000 medias') +
  labs(x = 'medias',
       y = 'Frecuencia')+
  theme_grey() +
  theme(plot.title = element_text(hjust = 0.5, size = 16))
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```r
mean(media1000.31); 0.55 # Media de las 1000 medias y media de la población de la cual provienen las 10
```

```
[1] 0.5492903
```

```
[1] 0.55
```

```r
sd(media1000.31); sqrt(0.55*0.45)/sqrt(31) # DE de las 1000 medias y DE
```

```
[1] 0.08767195
```

```
[1] 0.08935251
```

```r
# de la población de la  cual provienen las 1000 muestras dividida
# por la raíz del tamaño de la muestra


print("Obtenemos 1150 muestras aleatorias de tamaño n1 = 54 de una distribución exponencial con parámet
```

```
[1] "Obtenemos 1150 muestras aleatorias de tamaño n1 = 54 de una distribución exponencial con parámetro
```

```r
set.seed(65) # Para reproducir las muestras en el futuro
m1150.54 <- sapply(X = rep(54, 1150), FUN = rexp, rate = 3.2)
media1150.54 <- apply(m1150.54, 2, mean)
m1150.41 <- sapply(X = rep(41, 1150), FUN = rexp, rate = 1.5)
media1150.41 <- apply(m1150.41, 2, mean)
dif.medias <- media1150.54 - media1150.41 # Diferencia de medias
dmdf <- as.data.frame(dif.medias)
tail(dmdf)
```

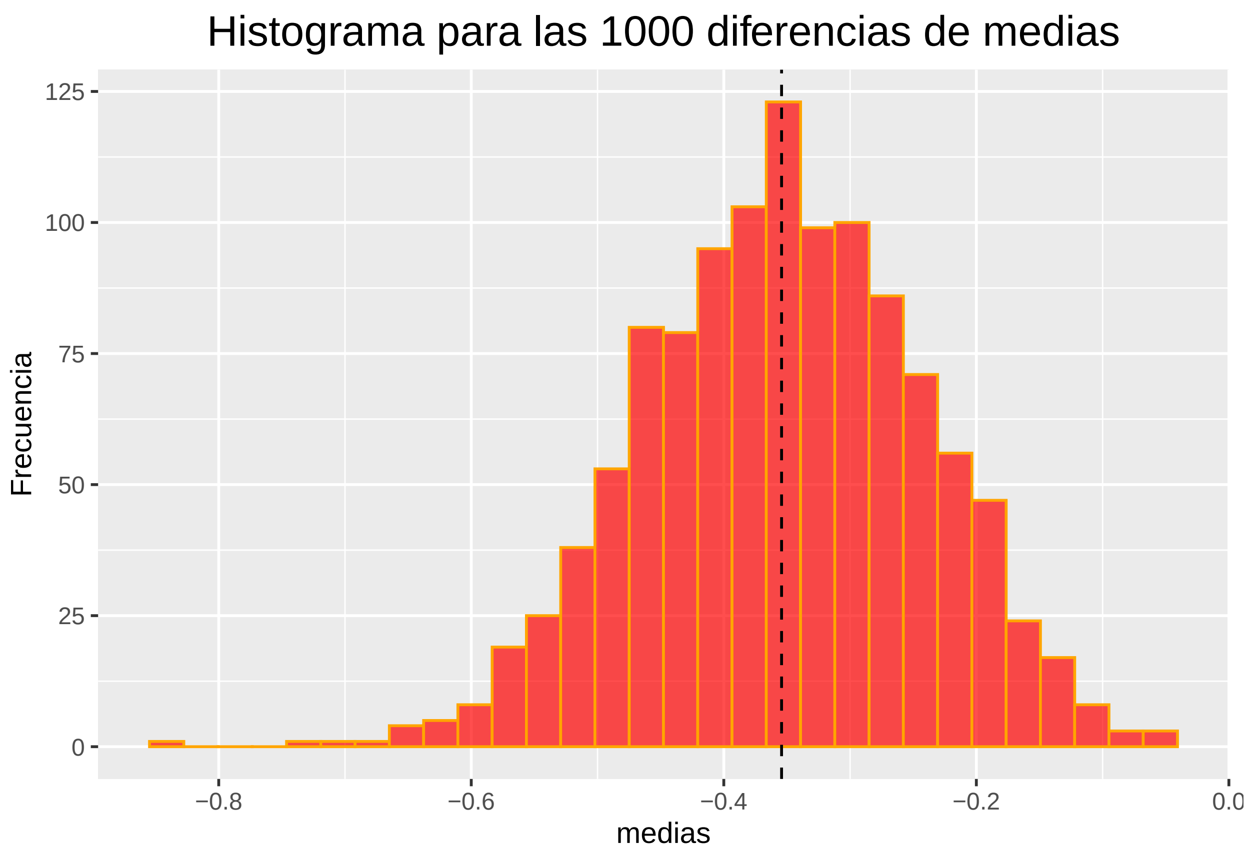
```
     dif.medias
1145 -0.3322635
1146 -0.4433986
1147 -0.2908058
1148 -0.3414530
1149 -0.1469445
1150 -0.4128351
```

```r
print("Observamos que el histograma de las diferencias de medias es parecida a una campana")
```

```
[1] "Observamos que el histograma de las diferencias de medias es parecida a una campana"
```

```r
ggplot(dmdf, aes(dif.medias)) +
  geom_histogram(colour = 'orange',
                 fill = 'red',
                 alpha = 0.7) + # Intensidad del color fill
  geom_vline(xintercept = mean(dif.medias), linetype="dashed", color = "black") +
  ggtitle('Histograma para las 1000 diferencias de medias') +
  labs(x = 'medias', y = 'Frecuencia')+
  theme_grey() +
  theme(plot.title = element_text(hjust = 0.5, size = 16))
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
mean(dif.medias); 1/3.2-1/1.5 # Media de las 1150 diferencias de medias y
```

```
[1] -0.3542439
```

```
[1] -0.3541667
```

```
# diferencia de medias de las poblaciones de las cuales provienen
# las 11500 muestras
sd(dif.medias); sqrt((1/3.2^2)/54 + (1/1.5^2)/41) # DE de las 1150
```

```
[1] 0.110714
```

```
[1] 0.1124658
```

```
# diferencias de medias y DE dada en literatura

print("Obtenemos 1100 muestras de tamaño n1 = 48 de una v.a. Bernoulli con p1 = 0.65 y otras 1100 muest
```

```
[1] "Obtenemos 1100 muestras de tamaño n1 = 48 de una v.a. Bernoulli con p1 = 0.65 y otras 1100 muestras
```

```
set.seed(7434) # Para reproducir las muestras en el futuro
m1100.48 <- sapply(X = rep(48, 1100), FUN = function(n) sample(x = c(0, 1), size = n, replace = TRUE, p
m1100.35 <- sapply(X = rep(35, 1100), FUN = function(n) sample(x = c(0, 1), size = n, replace = TRUE, p
```

```r
media1100.48 <- apply(m1100.48, 2, mean)
media1100.35 <- apply(m1100.35, 2, mean)
dif.medias <- media1100.48 - media1100.35
dmdf <- as.data.frame(dif.medias)
tail(dmdf)
```

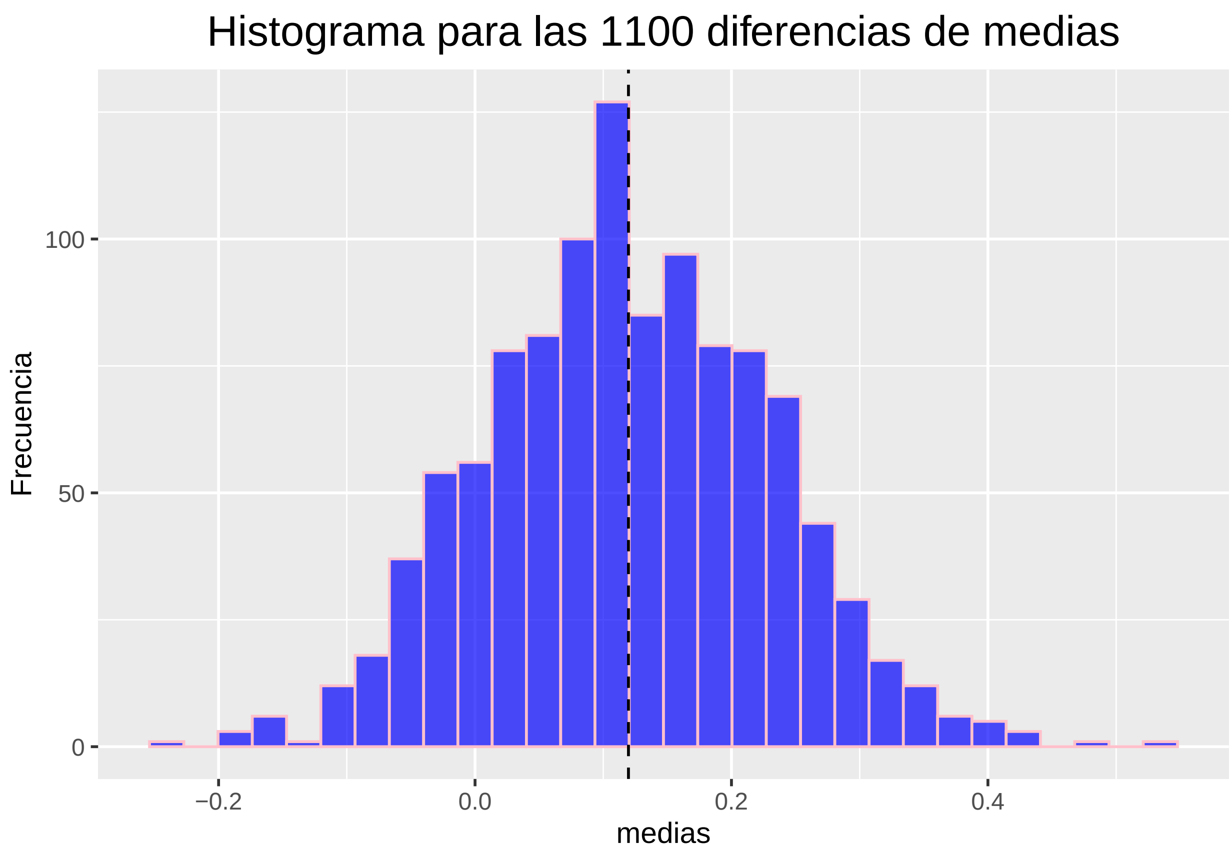
```
     dif.medias
1095 0.05357143
1096 0.07202381
1097 0.37321429
1098 0.20952381
1099 0.25119048
1100 0.09761905
```

```r
print("Observamos que el histograma de las diferencias de medias es parecida a una campana")
```

```
[1] "Observamos que el histograma de las diferencias de medias es parecida a una campana"
```

```r
ggplot(dmdf, aes(dif.medias)) +
  geom_histogram(colour = 'pink',
                 fill = 'blue',
                 alpha = 0.7) + # Intensidad del color fill
  geom_vline(xintercept = mean(dif.medias),
             linetype="dashed",
             color = "black") +
  ggtitle('Histograma para las 1100 diferencias de medias') +
  labs(x = 'medias',
       y = 'Frecuencia')+
  theme_grey() +
  theme(plot.title = element_text(hjust = 0.5, size = 16))
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
mean(dif.medias); 0.65 - 0.53 # Media de las 1100
```

```
[1] 0.1196558
```

```
[1] 0.12
```

```
#diferencias de medias y diferencia de medias de las poblaciones
# de las cuales provienen las muestras
sd(dif.medias); sqrt((0.65*0.35)/48 + (0.53*0.47)/35) # DE de las 1100
```

```
[1] 0.110132
```

```
[1] 0.1088886
```

```
# diferencias de medias y DE dada en literatura
```

RETO 2 SESION 4. ALGUNOS ESTIMADORES PUNTUALES INSESGADOS COMUNES

OBJETIVO

- Entender la idea del estimador puntual insesgado

REQUISITOS

- Haber estudiado el prework y el work

DESARROLLO

1. Genere 1500 muestras de tamaño 67 de la distribución exponencial con parámetro 5
2. Obtenga las 1500 medias correspondientes a cada una de las muestras
3. Realice el histograma de frecuencias de las 1500 medias
4. Encuentre la media muestral y desviación estandar muestral de las 1500 medias
5. Compare la media muestral encontrada en el paso anterior con la media real (1/5) de la población de la cual provienen las muestras
6. Compare la desviación estándar muestral encontrada con la desviación estándar real (1/5) de la población de la cual provienen las muestras pero dividida por 67 (el tamaño de las muestras).

```
#RETO 2 SESION 4
# 1. Genere 1500 muestras de tamaño 67 de la distribución exponencial con
# parámetro 5
# 2. Obtenga las 1500 medias correspondientes a cada una de las muestras
# 3. Realice el histograma de frecuencias de las 1500 medias
# 4. Encuentre la media muestral y desviación estándar muestral de las
# 1500 medias
# 5. Compare la media muestral encontrada en el paso anterior con
# la media real (1/5) de la población de la cual provienen las muestras
# 6. Compare la desviación estándar muestral encontrada con la desviación
# estándar real (1/5) de la población de la cual provienen las muestras
# pero dividida por 67 (el tamaño de las muestras)

library(ggplot2)
set.seed(123)#PARA QUE TODO EL TEAM GENERE LOS MISMOS VALORES
#PUNTO1
d1 <- sapply(X = rep(67, 1500), FUN = rexp, rate = 5)
class(d1)
```

```
[1] "matrix" "array"
```

```
#PUNTO 2
medd1<-apply(d1,2,mean)
#SE HACE LA CONVERSION A DATA FRAME
dfmedd1<-as.data.frame(medd1)
class(dfmedd1)
```

```
[1] "data.frame"
```

```
#PUNTO 3 HISTOGRAMA
graficop3<-ggplot(dfmedd1,aes(medd1))+
  geom_histogram(colour='blue',
                 fill='red',
                 alpha=0.5)+
  geom_vline(xintercept=mean(medd1),
             linetype="dashed",
             color="black")+
```

```r
  ggtitle("Histograma 1500 medias")+
  labs(x="Media",y="Frecuencia")+
  theme(plot.title = element_text(hjust = 0.5, size = 16))

#PUNTO 4
mean(medd1)
```

```
[1] 0.1995011
```

```r
sd(medd1)
```

```
[1] 0.02395705
```

```r
#PUNTO 5
mean(medd1); 1/5
```

```
[1] 0.1995011
```

```
[1] 0.2
```

```r
#PUNTO 6
sd(medd1);(1/5)/sqrt(67)
```

```
[1] 0.02395705
```

```
[1] 0.02443389
```

EJEMPLO 4. CONTRASTE DE HIPOTESIS

OBJETIVO

Llevar a cabo contrastes de hipotesis que ayuden a tomar decisiones

DESARROLLO

Un contraste de hipótesis es un procedimiento para juzgar si una propiedad que se supone en una población estadística es compatible con lo observado en una muestra de dicha población. Fue iniciada por Ronald Fisher y fundamentada posteriormente por Jerzy Neyman y Karl Pearson.

```r
# EJEMPLO 4. Contraste de hipótesis

# Contrastes comunes con muestras grandes
# Contraste de cola superior
# Dada una muestra aleatoria de tamaño n = 40

set.seed(104)
muestra <- rexp(n = 40, rate = 6) # media = 1/6 aprox 0.1667
# (media de la población)
tail(as.data.frame(muestra))
```

```
       muestra
35 0.125213495
36 0.166680130
37 0.128717925
38 0.003860131
39 0.045212421
40 0.086816614
```

```r
# estamos interesados en contrastar las hipótesis H0: mu = 0.1 vs
# H1: mu > 0.1 (contraste de cola superior)

# El valor observado del estadístico de prueba en este caso está dado por

z0 <- (mean(muestra)-0.1)/(sd(muestra)/sqrt(40))
z0
```

```
[1] 3.41015
```

```r
# que proviene de una distribución normal estándar aproximadamente.

# Supongamos que estamos interesados en encontrar la región de
# rechazo (de cola superior) con un nivel de significancia alpha = 0.05,
# debemos encontrar el valor z_{0.05} que satisface P(Z > z_{0.05}) = 0.05.

(z.05 <- qnorm(p = 0.05, lower.tail = FALSE))
```

```
[1] 1.644854
```

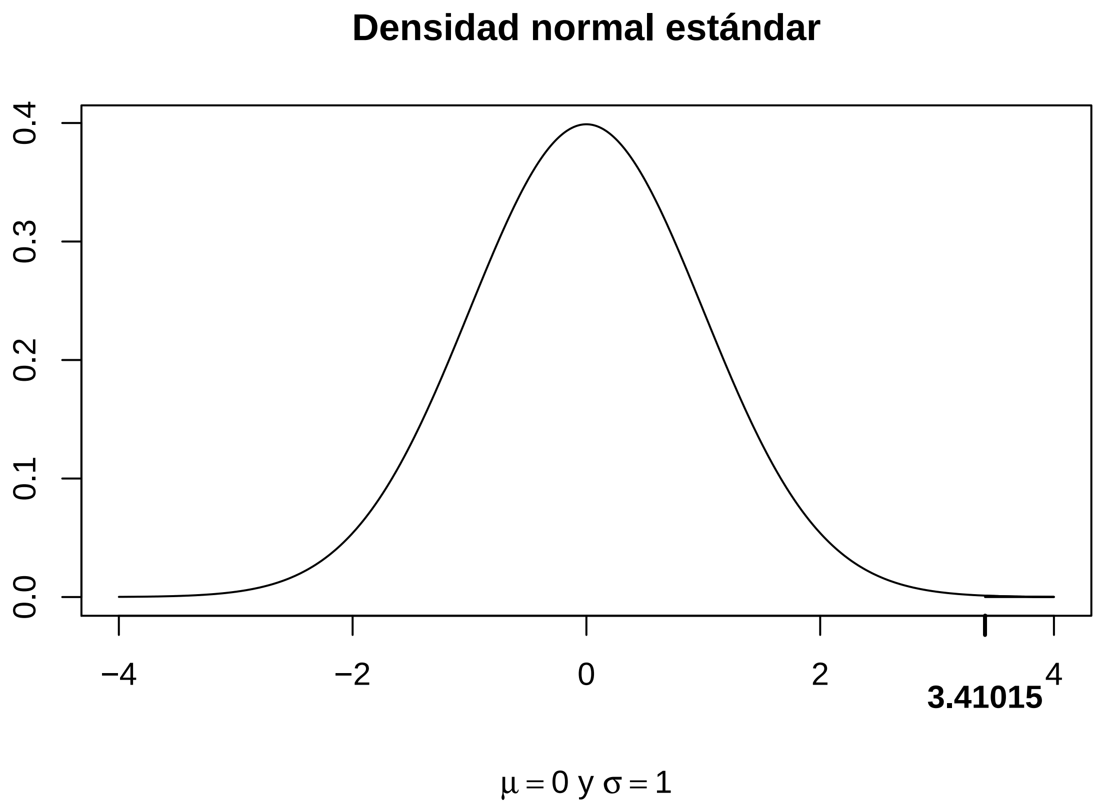
```r
# Como

z0 > z.05
```

```
[1] TRUE
```

```r
# rechazamos la hipótesis nula

# p-value El p-value lo podemos calcular como

(pvalue <- pnorm(z0, lower.tail = FALSE))
```

```
[1] 0.0003246356
```

```r
x <- seq(-4, 4, 0.01)
y <- dnorm(x)
plot(x, y, type = "l", xlab="", ylab="")
title(main = "Densidad normal estándar",
      sub = expression(paste(mu == 0, " y ", sigma == 1)))

polygon(c(z0, x[x>=z0], max(x)), c(0, y[x>=z0], 0), col="red")
axis(side = 1,
     at = z0,
     font = 2,
     padj = 1,
     lwd = 2)
```




```
# Contraste de cola inferior
# Dada una muestra aleatoria de tamaño n = 45

set.seed(174)
muestra <- sample(x = c(1, 0),
                  size = 45,
                  replace = TRUE,
                  prob = c(0.67, 0.33)); 0.67 # media real de la población
```

```
[1] 0.67
```

```
tail(as.data.frame(muestra))
```

```
   muestra
40       1
41       1
42       0
43       1
44       1
45       0
```

```
# estamos interesados en contrastar las hipótesis
# H0: p = 0.9 vs H1: p < 0.9 (contraste de cola inferior)
```

```r
# El valor observado del estadístico de prueba en este caso está dado por

z0 <- (mean(muestra)-0.9)/sqrt((0.9*(1-0.9))/45)
z0
```

```
[1] -7.205108
```

```r
# que proviene de una distribución normal estándar aproximadamente.

# Supongamos que estamos interesados en encontrar la
# región de rechazo (de cola inferior) con un nivel de
# significancia alpha = 0.05, debemos encontrar el valor
# z_{0.05} que satisface P(Z < z_{0.05}) = 0.05.

(z.05 <- qnorm(p = 0.05))
```

```
[1] -1.644854
```

```r
# Como

z0 < z.05
```

```
[1] TRUE
```

```r
# rechazamos la hipótesis nula.

# p-value El p-value lo podemos calcular como

(pvalue <- pnorm(z0)) # p-value muy pequeño
```

```
[1] 2.899895e-13
```

```r
# Contraste de dos colas
# Dada dos muestras aleatorias de tamaños n1 = 56 y n2 = 63

set.seed(174376)
m1 <- rexp(n = 56, rate = 4.1); 1/4.1 # media real de la población
```

```
[1] 0.2439024
```

```r
tail(as.data.frame(m1))
```

```
            m1
51 0.141795401
52 0.159477276
53 0.191133777
54 0.004537536
55 0.287620197
56 0.100260806
```

```r
m2 <- rexp(n = 63, rate = 3.4); 1/3.4 # media real de la población
```

```
[1] 0.2941176
```

```r
tail(as.data.frame(m2))
```

```
           m2
58 0.23177951
59 0.05689371
60 0.24539693
61 0.15520638
62 0.44972595
63 0.30483678
```

```r
1/4.1-1/3.4 # diferencia de medias real
```

```
[1] -0.05021521
```

```r
# estamos interesados en contrastar las
# hipótesis H0: mu1-mu2 = 0 vs H1: mu1-mu2 diferente de 0
# (contraste de dos colas)

# El valor observado del estadístico de prueba en este caso está dado por

z0 <- (mean(m1)-mean(m2)-0)/sqrt(var(m1)/56 + var(m2)/63)
z0
```

```
[1] 0.6932319
```

```r
# que proviene de una distribución normal estándar aproximadamente.

# Supongamos que estamos interesados en encontrar la
# región de rechazo (de dos colas) con un nivel de
# significancia alpha = 0.05, debemos encontrar el valor
# z_{0.025} que satisface P(Z > z_{0.025}) = 0.025.

(z.025 <- qnorm(p = 0.025, lower.tail = FALSE))
```

```
[1] 1.959964
```

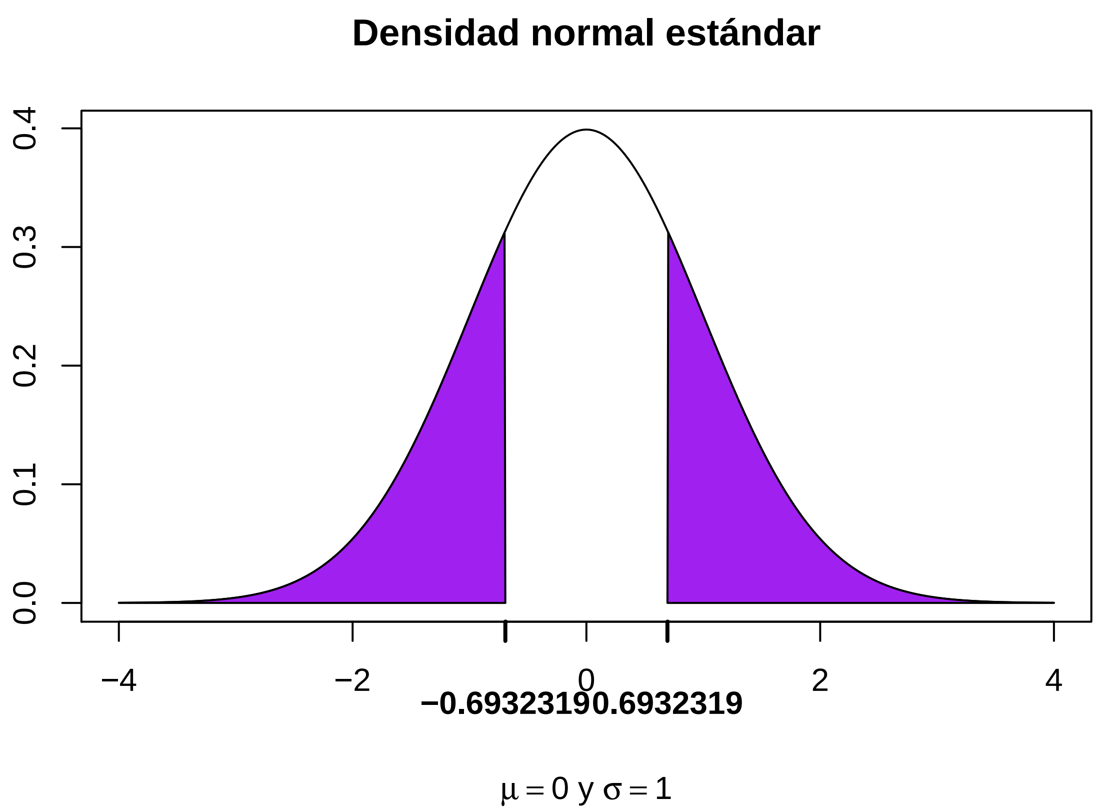
```r
# Como

(z0 < -z.025) | (z0 > z.025)
```

```
[1] FALSE
```

```r
# fallamos en rechazar la hipótesis nula.

# p-value El p-value lo podemos calcular como

(pvalue <- 2*pnorm(z0, lower.tail = FALSE))
```

```
[1] 0.488164
```

```
x <- seq(-4, 4, 0.01)
y <- dnorm(x)
plot(x, y, type = "l", xlab="", ylab="")
title(main = "Densidad normal estándar",
      sub = expression(paste(mu == 0, " y ", sigma == 1)))

polygon(c(min(x), x[x<=-z0], -z0), c(0, y[x<=-z0], 0), col="purple")
axis(side = 1, at = -z0, font = 2, padj = 1, lwd = 2)

polygon(c(z0, x[x>=z0], max(x)), c(0, y[x>=z0], 0), col="purple")
axis(side = 1, at = z0, font = 2, padj = 1, lwd = 2)
```




```
# Contraste de hipótesis con muestras pequeñas para mu y mu1 - mu2
# Contraste de cola superior
# Dada una muestra aleatoria de tamaño n = 15

set.seed(124)
muestra <- rnorm(n = 15, mean = 175, sd = 6)
tail(as.data.frame(muestra))
```

```
    muestra
10 182.2429
11 176.9100
```

```
12 166.4572
13 172.5695
14 180.9723
15 180.7529
```

```r
# estamos interesados en contrastar las
# hipótesis H0: mu = 170 vs H1: mu > 170 (contraste de cola superior)

# El valor observado del estadístico de prueba en este caso está dado por

t0 <- (mean(muestra)-170)/(sd(muestra)/sqrt(15))
t0
```

```
[1] 4.410437
```

```r
# que proviene de una distribución t de Student
# con n-1 = 14 grados de libertad (gl).

# Supongamos que estamos interesados en encontrar
# la región de rechazo (de cola superior) con un nivel
# de significancia alpha = 0.05, debemos
# encontrar el valor t_{0.05} que satisface
# $P(T > t_{0.05}) = 0.05$, donde $T$ se distribuye
# como t de Student con n-1 = 14 gl.

(t.05 <- qt(p = 0.05, df = 14, lower.tail = FALSE))
```

```
[1] 1.76131
```

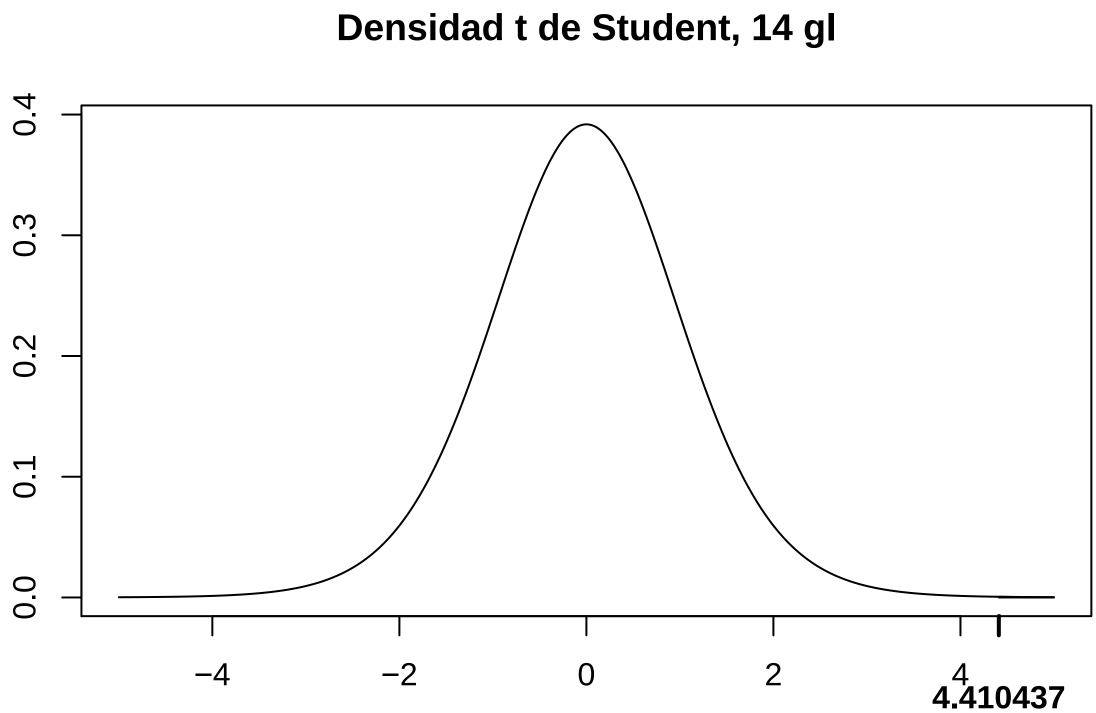
```r
# Como

t0 > t.05
```

```
[1] TRUE
```

```r
# rechazamos la hipótesis nula

# p-value El p-value lo podemos calcular como

(pvalue <- pt(t0, df = 14, lower.tail = FALSE))
```

```
[1] 0.000296395
```

```r
x <- seq(-5, 5, 0.01)
y <- dt(x, df = 14)
plot(x, y, type = "l", xlab="", ylab="")
title(main = "Densidad t de Student, 14 gl")

polygon(c(t0, x[x>=t0], max(x)), c(0, y[x>=t0], 0), col="red")
axis(side = 1, at = t0, font = 2, padj = 1, lwd = 2)
```




```
t.test(x = muestra,
       alternative = "greater",
       mu = 170)
```

```
    One Sample t-test

data:  muestra
t = 4.4104, df = 14, p-value = 0.0002964
alternative hypothesis: true mean is greater than 170
95 percent confidence interval:
 173.6258      Inf
sample estimates:
mean of x
 176.0365
```

```
# Contraste de dos colas
# Dada dos muestras aleatorias de tamaños n1 = 23 y n2 = 20

set.seed(1776)
m1 <- rnorm(n = 23, mean = 175, sd = 3)
tail(as.data.frame(m1))
```

```
         m1
18 173.9126
```

```
19 175.7334
20 171.3733
21 170.9339
22 178.3138
23 172.2944
```

```r
m2 <- rnorm(n = 20, mean = 160, sd = 3)
tail(as.data.frame(m2))
```

```
         m2
15 164.4266
16 158.2315
17 166.0989
18 162.0823
19 161.2146
20 161.9395
```

```r
175-160 # diferencia de medias real
```

```
[1] 15
```

```r
# estamos interesados en contrastar las
# hipótesis H0: mu1-mu2 = 0 vs H1: mu1-mu2
# diferente de 0 (contraste de dos colas)

# El valor observado del estadístico de prueba en este caso está dado por

t0 <- (mean(m1)-mean(m2)-0)/(sqrt((22*var(m1)+19*var(m2))/(23+20-2))*sqrt(1/23+1/20))
t0
```

```
[1] 13.44122
```

```r
# que proviene de una distribución t de Student con 23 + 20 - 2 = 41 gl

# Supongamos que estamos interesados en encontrar
# la región de rechazo (de dos colas) con un nivel de
# significancia alpha = 0.05, debemos encontrar el valor
# t_{0.025} que satisface P(T > t_{0.025}) = 0.025.

(t.025 <- qt(p = 0.025, df= 41, lower.tail = FALSE))
```

```
[1] 2.019541
```

```r
# Como

(t0 < -t.025) | (t0 > t.025)
```

```
[1] TRUE
```

```r
# rechazamos la hipótesis nula.

# p-value El p-value lo podemos calcular como

(pvalue <- 2*pt(t0, df = 41, lower.tail = FALSE))
```

```
[1] 1.28906e-16
```

```r
t.test(x = m1, y = m2,
       alternative = "two.sided",
       mu = 0, paired = FALSE, var.equal = TRUE)
```

```

    Two Sample t-test

data:  m1 and m2
t = 13.441, df = 41, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 12.00835 16.25489
sample estimates:
mean of x mean of y
 175.4867  161.3551
```

RETO 3. CONTRASTE DE HIPOTESIS

OBJETIVO

- Llevar a cabo un contraste de hiptoesis acerca de la media de una población normal cuando el tamaño muestral es pequeño

REQUISITOS

- Haber trabajado con el prework y el work

DESARROLLO

Dada una muestra aleatoria de tamaño n = 10 de una distribución normal

```r
set.seed(124)
muestra <- rnorm(n = 10, mean = 110, sd = 7) # muestra pequeña
tail(as.data.frame(muestra))
```

```
    muestra
5  119.9788
6  115.2114
7  114.9016
8  108.3945
9  111.3797
10 118.4501
```

Estamos interesados en contrastar las hipótesis H0: mu = 120 vs H1: mu < 120 (contraste de cola inferior).

1. Decida si rechazar o no la hipótesis nula si el nivel de significancia es alpha = 0.05
2. Obtenga el p-value de la prueba
3. Lleve a cabo la prueba con la función *t.test*

```r
#RETO 3 SESION 4
# Reto 3. Contraste de hipótesis

# Dada una muestra aleatoria de tamaño n = 10 de la distribución normal

set.seed(124)
muestra <- rnorm(n = 10, mean = 110, sd = 7)
tail(as.data.frame(muestra))
```

```
    muestra
5  119.9788
6  115.2114
7  114.9016
8  108.3945
9  111.3797
10 118.4501
```

```r
# estamos interesados en contrastar las hipótesis H0:
# mu = 120 vs H1: mu < 120 (contraste de cola inferior).

# 1. Decida si rechazar o no la hipótesis nula si el
# nivel de significancia es alpha = 0.05
# 2. Obtenga el p-value de la prueba
# 3. Lleve a cabo la prueba con la función t.test


#### Contraste de cola inferior

# El valor observado del estadístico de prueba en este caso está dado por

t0 <- (mean(muestra)-120)/(sd(muestra)/sqrt(10))
t0
```

```
[1] -4.444938
```

```r
# que proviene de una distribución t de Student con
# n - 1 = 9 grados de libertad (gl).

# estamos interesados en encontrar la región
# de rechazo (de cola inferior) con un nivel de
# significancia alpha = 0.05, debemos encontrar el valor
# t0.05 que satisface P(T < t0.05) = 0.05, donde T se
# distribuye como t de Student con n-1 = 9 gl.

(t.05 <- qt(p = 0.05, df = 9))
```

```
[1] -1.833113
```

```r
# Como

t0 < t.05
```

```
[1] TRUE
```

```r
# rechazamos la hipótesis nula

# **p-value** El p-value lo podemos calcular como

(pvalue <- pt(t0, df = 9)) # 2.
```

```
[1] 0.0008059723
```

```r
t.test(x = muestra, # 3.
       alternative = "less",
       mu = 120)
```

```

    One Sample t-test

data:  muestra
t = -4.4449, df = 9, p-value = 0.000806
alternative hypothesis: true mean is less than 120
95 percent confidence interval:
     -Inf 115.0074
sample estimates:
mean of x
 111.5034
```

POSTWORK

OBJETIVO

Investigar la dependencia o independecia de las variables aleatorias X y Y, el número de goles anotados por el equipo de casa y el número de goles anotados por el equipo visitante.

REQUISITOS

- R, RStudio
- Haber trabajado con el Prework y el Work

DESARROLLO

Ahora investigarás la dependencia o independencia del número de goles anotados por el equipo de casa y el número de goles anotados por el equipo visitante mediante un procedimiento denominado bootstrap, revisa bibliografía en internet para que tengas nociones de este desarrollo.

Ya hemos estimado las probabilidades conjuntas de que el equipo de casa anote X=x goles (x=0,1,. . . ,8), y el equipo visitante anote Y=y goles (y=0,1,. . . ,6), en un partido. Obtén una tabla de cocientes al dividir estas probabilidades conjuntas por el producto de las probabilidades marginales correspondientes.

Mediante un procedimiento de boostrap, obtén más cocientes similares a los obtenidos en la tabla del punto anterior. Esto para tener una idea de las distribuciones de la cual vienen los cocientes en la tabla anterior. Menciona en cuáles casos le parece razonable suponer que los cocientes de la tabla en el punto 1, son iguales a 1 (en tal caso tendríamos independencia de las variables aleatorias X y Y).

```
#Solución del Postwork Sesión 4

#Lo primero que haremos es cargar los paquetes que usaremos más adelante. Usamos las
#funciones suppressWarnings y supperssMessages para que no se impriman mensajes ni
#advertencias al cargar el paquete.

suppressWarnings(suppressMessages(library(dplyr)))
suppressWarnings(suppressMessages(library(reshape2)))
suppressWarnings(suppressMessages(library(ggplot2)))

#Comenzamos importando los datos que se encuentran en archivos csv a R

url1718 <- "https://www.football-data.co.uk/mmz4281/1718/SP1.csv"
url1819 <- "https://www.football-data.co.uk/mmz4281/1819/SP1.csv"
url1920 <- "https://www.football-data.co.uk/mmz4281/1920/SP1.csv"

d1718 <- read.csv(file = url1718) # Importación de los datos a R
d1819 <- read.csv(file = url1819)
d1920 <- read.csv(file = url1920)

#Obtenemos una mejor idea de los datos que se encuentran en los data frames
# con las funciones str, head, View y summary

str(d1718); str(d1819); str(d1920)
```

```
'data.frame':   380 obs. of  64 variables:
 $ Div       : chr  "SP1" "SP1" "SP1" "SP1" ...
 $ Date      : chr  "18/08/17" "18/08/17" "19/08/17" "19/08/17" ...
 $ HomeTeam  : chr  "Leganes" "Valencia" "Celta" "Girona" ...
 $ AwayTeam  : chr  "Alaves" "Las Palmas" "Sociedad" "Ath Madrid" ...
 $ FTHG      : int  1 1 2 2 1 0 2 0 1 0 ...
 $ FTAG      : int  0 0 3 2 1 0 0 3 0 1 ...
 $ FTR       : chr  "H" "H" "A" "D" ...
 $ HTHG      : int  1 1 1 2 1 0 2 0 0 0 ...
 $ HTAG      : int  0 0 1 0 1 0 0 2 0 0 ...
 $ HTR       : chr  "H" "H" "D" "H" ...
 $ HS        : int  16 22 16 13 9 12 15 12 14 10 ...
 $ AS        : int  6 5 13 9 9 8 3 16 9 13 ...
 $ HST       : int  9 6 5 6 4 2 2 6 3 4 ...
 $ AST       : int  3 4 6 3 6 2 0 8 1 6 ...
 $ HF        : int  14 25 12 15 14 16 16 16 18 16 ...
 $ AF        : int  18 13 11 15 12 15 15 12 14 15 ...
 $ HC        : int  4 5 5 6 7 7 8 4 11 3 ...
 $ AC        : int  2 2 4 0 3 6 0 4 6 7 ...
 $ HY        : int  0 3 3 2 2 1 2 5 1 2 ...
```

```
 $ AY        : int  1 3 1 4 4 3 1 1 3 3 ...
 $ HR        : int  0 0 0 0 1 0 0 0 0 0 ...
 $ AR        : int  0 1 0 1 0 1 0 1 0 0 ...
 $ B365H     : num  2.05 1.75 2.38 8 1.62 1.5 1.17 9.5 3.25 2.1 ...
 $ B365D     : num  3.2 3.8 3.25 4.33 4 4 8 5.75 3.25 3.3 ...
 $ B365A     : num  4.1 4.5 3.2 1.45 5.5 7.5 15 1.3 2.3 3.7 ...
 $ BWH       : num  2.05 1.75 2.4 7.5 1.62 1.48 1.18 9.25 3.25 2.15 ...
 $ BWD       : num  3.1 3.9 3.3 4.33 3.9 4.25 7.5 5.75 3.2 3.3 ...
 $ BWA       : num  4.1 4.6 3 1.45 5.75 7 14.5 1.3 2.3 3.5 ...
 $ IWH       : num  2.1 1.75 2.5 7.2 1.55 1.5 1.17 7.5 3.3 2.1 ...
 $ IWD       : num  3.4 3.6 3.3 4.4 4 4.2 7.5 5.5 3.35 3.4 ...
 $ IWA       : num  3.5 4.8 2.85 1.45 6.2 6.5 15 1.35 2.2 3.5 ...
 $ LBH       : num  2.05 1.75 2.35 7.5 1.6 1.5 1.2 9.5 3.25 2.1 ...
 $ LBD       : num  3 3.8 3.25 4 3.9 4 6.5 5.25 3.1 3.1 ...
 $ LBA       : num  4.2 4.33 3 1.5 5.5 7 15 1.3 2.3 3.4 ...
 $ PSH       : num  2.03 1.78 2.44 8.36 1.62 ...
 $ PSD       : num  3.25 4.01 3.4 4.38 4.17 4.37 7.35 5.79 3.24 3.36 ...
 $ PSA       : num  4.52 4.83 3.16 1.49 6.18 7.31 15.5 1.33 2.36 3.49 ...
 $ WHH       : num  2.05 1.8 2.4 8 1.67 1.5 1.22 11 3.1 2.2 ...
 $ WHD       : num  3.1 3.75 3.4 4.2 3.6 4 6 4.5 3.1 3.3 ...
 $ WHA       : num  4 4.2 2.9 1.44 5.5 7 13 1.33 2.4 3.3 ...
 $ VCH       : num  2.05 1.8 2.4 7.5 1.65 1.5 1.2 9.5 3.25 2.15 ...
 $ VCD       : num  3.2 4 3.4 4.3 4 4.2 7 5.75 3.25 3.3 ...
 $ VCA       : num  4.4 4.6 3.13 1.5 5.75 7 13 1.3 2.3 3.5 ...
 $ Bb1X2     : int  35 35 35 35 35 34 35 35 34 34 ...
 $ BbMxH     : num  2.12 1.83 2.5 8.36 1.69 ...
 $ BbAvH     : num  2.03 1.77 2.39 7.53 1.63 1.5 1.19 9.68 3.26 2.18 ...
 $ BbMxD     : num  3.4 4.04 3.5 4.4 4.17 4.4 8 5.86 3.35 3.4 ...
 $ BbAvD     : num  3.15 3.86 3.32 4.17 3.93 4.17 7.11 5.44 3.17 3.26 ...
 $ BbMxA     : num  4.52 4.83 3.2 1.51 6.2 7.5 17 1.35 2.4 3.7 ...
 $ BbAvA     : num  4.17 4.46 3.01 1.48 5.58 ...
 $ BbOU      : int  31 33 34 34 33 32 27 27 32 32 ...
 $ BbMx.2.5  : num  2.84 1.69 2.03 2.2 1.81 2.01 1.44 1.5 2.42 2.25 ...
 $ BbAv.2.5  : num  2.68 1.64 1.98 2.11 1.75 1.94 1.4 1.46 2.36 2.14 ...
 $ BbMx.2.5.1: num  1.53 2.4 1.9 1.8 2.14 1.96 3.1 2.95 1.63 1.76 ...
 $ BbAv.2.5.1: num  1.46 2.27 1.84 1.74 2.09 1.87 2.88 2.64 1.58 1.7 ...
 $ BbAH      : int  18 16 18 16 16 17 17 16 15 17 ...
 $ BbAHh     : num  -0.5 -0.75 -0.25 1.25 -1 -1 -2 1.5 0.25 -0.25 ...
 $ BbMxAHH   : num  2.07 2.05 2.08 1.77 2.12 1.9 2.05 2.03 1.93 1.92 ...
 $ BbAvAHH   : num  2.03 1.97 2.05 1.75 2.06 1.86 2 1.98 1.89 1.88 ...
 $ BbMxAHA   : num  1.9 1.96 1.87 2.25 1.86 2.05 1.91 1.95 2.03 2.04 ...
 $ BbAvAHA   : num  1.86 1.91 1.83 2.16 1.82 2.01 1.86 1.89 1.98 1.99 ...
 $ PSCH      : num  1.98 1.78 2.12 6.93 1.64 1.53 1.2 12.4 3.31 2.2 ...
 $ PSCD      : num  3.35 4.24 3.53 3.83 4.18 4.48 8.25 7 3.32 3.27 ...
 $ PSCA      : num  4.63 4.43 3.74 1.63 5.82 6.91 15.2 1.26 2.4 3.85 ...

'data.frame':   380 obs. of  61 variables:
 $ Div       : chr  "SP1" "SP1" "SP1" "SP1" ...
 $ Date      : chr  "17/08/2018" "17/08/2018" "18/08/2018" "18/08/2018" ...
 $ HomeTeam  : chr  "Betis" "Girona" "Barcelona" "Celta" ...
 $ AwayTeam  : chr  "Levante" "Valladolid" "Alaves" "Espanol" ...
 $ FTHG      : int  0 0 3 1 1 1 2 1 2 1 ...
 $ FTAG      : int  3 0 0 1 2 2 0 4 1 1 ...
 $ FTR       : chr  "A" "D" "H" "D" ...
```

```
$ HTHG       : int  0 0 0 0 1 0 1 0 1 0 ...
$ HTAG       : int  1 0 0 1 1 2 0 3 1 1 ...
$ HTR        : chr  "A" "D" "D" "A" ...
$ HS         : int  22 13 25 12 16 18 10 13 17 13 ...
$ AS         : int  6 2 3 14 8 8 4 17 12 9 ...
$ HST        : int  8 1 9 2 7 6 3 2 5 4 ...
$ AST        : int  4 1 0 5 4 6 1 8 2 3 ...
$ HF         : int  10 21 6 13 16 12 11 6 12 10 ...
$ AF         : int  10 20 13 14 10 13 27 15 13 15 ...
$ HC         : int  5 3 7 8 4 7 3 2 6 4 ...
$ AC         : int  3 2 1 7 6 0 0 6 2 10 ...
$ HY         : int  0 1 0 3 2 1 1 1 4 2 ...
$ AY         : int  2 1 2 2 3 1 7 0 5 3 ...
$ HR         : int  0 0 0 0 0 0 0 0 0 0 ...
$ AR         : int  0 0 0 0 0 0 0 0 0 0 ...
$ B365H      : num  1.66 1.75 1.11 1.85 2.04 1.66 1.2 3.25 1.75 3 ...
$ B365D      : num  4 3.6 10 3.5 3.4 3.75 7 3.6 3.3 3.2 ...
$ B365A      : num  5 5 21 4.5 3.8 5.5 13 2.14 5.5 2.5 ...
$ BWH        : num  1.7 1.75 1.11 1.91 2.05 1.7 1.18 3.5 1.78 2.85 ...
$ BWD        : num  3.7 3.5 10 3.4 3.3 3.7 7.25 3.5 3.5 3.25 ...
$ BWA        : num  5.25 5.25 20 4.25 3.9 5.25 16 2.1 5 2.55 ...
$ IWH        : num  1.75 1.8 1.12 1.9 2 1.7 1.2 3.5 1.85 2.85 ...
$ IWD        : num  3.6 3.6 9 3.5 3.4 3.75 6.5 3.4 3.5 3.2 ...
$ IWA        : num  4.9 4.5 20 4.1 3.8 5 15 2.1 4.4 2.55 ...
$ PSH        : num  1.69 1.8 1.11 1.93 2.06 1.72 1.2 3.46 1.79 3.12 ...
$ PSD        : num  4.19 3.7 11.27 3.64 3.51 ...
$ PSA        : num  5.11 4.99 25.4 4.27 3.91 ...
$ WHH        : num  1.67 1.75 1.08 1.91 2.05 1.73 1.22 3.3 1.8 3 ...
$ WHD        : num  3.9 3.6 9 3.5 3.3 3.6 6 3.7 3.4 3.2 ...
$ WHA        : num  4.75 4.6 29 4 3.6 4.75 13 2.05 4.75 2.4 ...
$ VCH        : num  1.67 1.8 1.1 1.93 2.05 1.7 1.2 3.4 1.8 3 ...
$ VCD        : num  4.2 3.7 10.5 3.5 3.5 3.8 7 3.6 3.4 3.2 ...
$ VCA        : num  5.2 4.8 34 4.4 3.9 5 13 2.1 5 2.45 ...
$ Bb1X2      : int  40 40 40 38 40 40 39 40 40 39 ...
$ BbMxH      : num  1.75 1.85 1.13 1.97 2.11 1.76 1.24 3.53 1.85 3.12 ...
$ BbAvH      : num  1.68 1.78 1.1 1.9 2.03 1.7 1.21 3.38 1.78 2.99 ...
$ BbMxD      : num  4.25 3.83 11.5 3.73 3.62 3.93 7.36 3.75 3.64 3.29 ...
$ BbAvD      : num  4 3.6 9.82 3.53 3.43 3.77 6.66 3.56 3.43 3.14 ...
$ BbMxA      : num  5.25 5.27 41 4.5 3.93 ...
$ BbAvA      : num  4.95 4.79 25.67 4.2 3.76 ...
$ BbOU       : int  38 38 32 36 37 37 33 37 36 36 ...
$ BbMx.2.5   : num  1.82 2.21 1.39 2.13 2.05 1.95 1.5 1.83 2.49 2.45 ...
$ BbAv.2.5   : num  1.76 2.13 1.34 2.06 1.99 1.88 1.45 1.76 2.35 2.33 ...
$ BbMx.2.5.1: num  2.15 1.78 3.4 1.84 1.88 1.98 2.75 2.13 1.64 1.65 ...
$ BbAv.2.5.1: num  2.06 1.71 3.18 1.76 1.81 1.91 2.66 2.04 1.58 1.59 ...
$ BbAH       : int  20 20 19 18 18 19 19 19 18 17 ...
$ BbAHh      : num  -0.75 -0.75 -2.5 -0.75 -0.25 -0.75 -1.75 0.25 -0.75 0.25 ...
$ BbMxAHH    : num  1.89 2.06 1.95 2.26 1.76 1.96 1.85 2.08 2.11 1.82 ...
$ BbAvAHH    : num  1.85 2.01 1.91 2.18 1.74 1.91 1.8 2.03 2.04 1.75 ...
$ BbMxAHA    : num  2.07 1.9 2 1.74 2.23 2.01 2.15 1.86 1.86 2.23 ...
$ BbAvAHA    : num  2 1.85 1.95 1.71 2.14 1.94 2.07 1.83 1.82 2.12 ...
$ PSCH       : num  1.59 1.76 1.1 2.18 2.32 1.77 1.19 4.57 1.69 3.55 ...
$ PSCD       : num  4.42 3.57 11.85 3.26 3.21 ...
$ PSCA       : num  5.89 5.62 32.17 3.85 3.53 ...
```

```
'data.frame':   380 obs. of  105 variables:
 $ Div        : chr  "SP1" "SP1" "SP1" "SP1" ...
 $ Date       : chr  "16/08/2019" "17/08/2019" "17/08/2019" "17/08/2019" ...
 $ Time       : chr  "20:00" "16:00" "18:00" "19:00" ...
 $ HomeTeam   : chr  "Ath Bilbao" "Celta" "Valencia" "Mallorca" ...
 $ AwayTeam   : chr  "Barcelona" "Real Madrid" "Sociedad" "Eibar" ...
 $ FTHG       : int  1 1 1 2 0 4 1 0 1 1 ...
 $ FTAG       : int  0 3 1 1 1 4 0 2 2 0 ...
 $ FTR        : chr  "H" "A" "D" "H" ...
 $ HTHG       : int  0 0 0 1 0 1 0 0 0 1 ...
 $ HTAG       : int  0 1 0 0 0 1 0 1 0 0 ...
 $ HTR        : chr  "D" "A" "D" "H" ...
 $ HS         : int  11 7 14 16 13 12 9 7 13 5 ...
 $ AS         : int  11 17 12 11 4 14 16 12 14 6 ...
 $ HST        : int  5 4 6 4 2 7 2 2 4 5 ...
 $ AST        : int  2 11 3 5 2 7 4 4 3 0 ...
 $ HF         : int  14 17 13 13 17 10 18 11 11 19 ...
 $ AF         : int  9 12 14 14 11 16 15 17 19 22 ...
 $ HC         : int  3 6 3 9 8 2 2 8 6 3 ...
 $ AC         : int  8 4 3 3 0 7 9 4 1 4 ...
 $ HY         : int  1 5 4 2 1 3 2 2 2 3 ...
 $ AY         : int  1 2 4 3 4 1 1 2 6 4 ...
 $ HR         : int  0 0 1 0 1 0 0 0 1 1 ...
 $ AR         : int  0 1 0 0 0 0 0 0 0 1 ...
 $ B365H      : num  5.25 4.75 1.66 2.8 2 1.6 2.15 3.2 1.66 1.44 ...
 $ B365D      : num  3.8 4.2 3.75 3.2 3.2 3.8 3.2 3.3 3.75 4.33 ...
 $ B365A      : num  1.65 1.65 5.5 2.6 4.2 6.5 3.6 2.3 5.5 8 ...
 $ BWH        : num  5.5 4.4 1.67 2.95 2.05 1.6 2.15 3.1 1.65 1.45 ...
 $ BWD        : num  3.8 4.2 3.75 3.1 3.25 3.8 3.3 3.4 3.75 4.33 ...
 $ BWA        : num  1.65 1.72 5.5 2.6 3.9 6.25 3.6 2.3 5.75 7.5 ...
 $ IWH        : num  5 5.3 1.67 2.9 2.05 1.63 2.2 3.1 1.63 1.45 ...
 $ IWD        : num  3.8 4.2 3.75 3.1 3.1 4 3.25 3.4 3.75 4.4 ...
 $ IWA        : num  1.7 1.6 5.3 2.6 4.05 5.5 3.4 2.3 5.7 7.2 ...
 $ PSH        : num  5.15 4.73 1.68 2.98 2.1 1.62 2.29 3.13 1.63 1.49 ...
 $ PSD        : num  3.84 4.18 3.94 3.14 3.21 3.99 3.31 3.56 3.81 4.34 ...
 $ PSA        : num  1.74 1.72 5.47 2.66 4.13 6.13 3.45 2.33 6.38 7.58 ...
 $ WHH        : num  5 5.25 1.67 2.9 2.05 1.6 2.25 3 1.62 1.47 ...
 $ WHD        : num  3.8 4.2 3.8 3.1 3.2 3.9 3.3 3.5 3.75 4.2 ...
 $ WHA        : num  1.7 1.6 5.25 2.62 4 5.8 3.3 2.3 6 8 ...
 $ VCH        : num  5 4.75 1.67 2.9 2.1 1.65 2.25 3 1.62 1.45 ...
 $ VCD        : num  3.8 4.2 3.9 3.13 3.2 4 3.3 3.5 3.8 4.2 ...
 $ VCA        : num  1.75 1.73 5.75 2.7 4.1 5.75 3.3 2.3 5.75 8 ...
 $ MaxH       : num  5.5 5.3 1.72 3.05 2.1 1.65 2.31 3.2 1.67 1.52 ...
 $ MaxD       : num  3.95 4.4 3.98 3.2 3.3 4.15 3.4 3.56 3.9 4.5 ...
 $ MaxA       : num  1.76 1.73 5.75 2.7 4.25 6.5 3.6 2.4 6.5 8.5 ...
 $ AvgH       : num  5.07 4.67 1.68 2.91 2.06 1.61 2.23 3.08 1.64 1.47 ...
 $ AvgD       : num  3.81 4.12 3.8 3.09 3.18 3.95 3.25 3.41 3.76 4.23 ...
 $ AvgA       : num  1.71 1.69 5.29 2.62 4.02 5.8 3.43 2.33 5.78 7.63 ...
 $ B365.2.5   : num  1.8 1.53 2 2.3 2.5 1.8 2.1 1.9 2.1 2.2 ...
 $ B365.2.5.1 : num  2 2.5 1.8 1.61 1.53 2 1.72 1.9 1.72 1.66 ...
 $ P.2.5      : num  1.81 1.52 2.08 2.45 2.72 1.88 2.16 1.95 2.16 2.3 ...
 $ P.2.5.1    : num  2.09 2.66 1.82 1.6 1.5 2.02 1.76 1.95 1.76 1.68 ...
 $ Max.2.5    : num  1.85 1.53 2.14 2.47 2.75 1.9 2.2 1.98 2.21 2.3 ...
 $ Max.2.5.1  : num  2.11 2.72 1.83 1.65 1.54 2.05 1.77 1.95 1.78 1.71 ...
```

```
 $ Avg.2.5    : num  1.79 1.49 2.07 2.34 2.59 1.84 2.13 1.92 2.13 2.23 ...
 $ Avg.2.5.1  : num  2.05 2.58 1.77 1.6 1.49 1.98 1.72 1.89 1.72 1.66 ...
 $ AHh        : num  0.75 0.75 -0.75 0 -0.5 -1 -0.25 0.25 -0.75 -1 ...
 $ B365AHH    : num  1.99 2.04 1.91 2.05 2.08 2.05 1.95 1.88 1.86 1.88 ...
 $ B365AHA    : num  1.94 1.89 2.02 1.88 1.85 1.75 1.98 2.05 2.07 2.05 ...
 $ PAHH       : num  1.98 2.01 1.91 2.07 2.1 2.11 1.96 1.9 1.84 1.88 ...
 $ PAHA       : num  1.94 1.91 2.01 1.85 1.82 1.81 1.96 2.02 2.08 2.04 ...
 $ MaxAHH     : num  2 2.05 1.93 2.07 2.1 2.14 1.97 1.9 1.87 1.89 ...
 $ MaxAHA     : num  1.95 1.91 2.03 1.88 1.85 1.85 1.99 2.06 2.08 2.08 ...
 $ AvgAHH     : num  1.96 2 1.89 2.04 2.06 2.07 1.93 1.87 1.83 1.85 ...
 $ AvgAHA     : num  1.92 1.88 1.99 1.85 1.83 1.8 1.95 2.01 2.06 2.03 ...
 $ B365CH     : num  5.25 5.25 1.66 2.87 1.9 1.53 2.3 3 1.8 1.5 ...
 $ B365CD     : num  3.8 4.2 3.75 3.2 3.1 4 3.4 3.4 3.6 4 ...
 $ B365CA     : num  1.65 1.57 5.5 2.55 5 6.5 3.2 2.4 4.75 8 ...
 $ BWCH       : num  4.75 4.5 1.65 2.95 1.95 1.57 2.35 3 1.8 1.5 ...
 $ BWCD       : num  3.75 4.1 3.8 3.1 3.2 3.8 3.2 3.4 3.4 3.9 ...
 $ BWCA       : num  1.75 1.7 5.5 2.6 4.5 6.5 3.2 2.35 5 7.75 ...
 $ IWCH       : num  5 4.6 1.67 2.9 1.9 1.55 2.35 3 1.85 1.5 ...
 $ IWCD       : num  3.8 3.8 3.8 3.1 3.15 4.05 3.25 3.35 3.55 3.9 ...
 $ IWCA       : num  1.7 1.75 5.3 2.6 4.85 6.3 3.15 2.35 4.4 7.6 ...
 $ PSCH       : num  5.34 5.1 1.69 2.96 1.9 1.54 2.43 3.13 1.82 1.57 ...
 $ PSCD       : num  3.62 4.46 3.88 3.26 3.18 4.19 3.27 3.38 3.53 3.78 ...
 $ PSCA       : num  1.78 1.65 5.47 2.6 5.3 6.87 3.2 2.41 5.07 7.66 ...
 $ WHCH       : num  5 5 1.65 2.9 2.05 1.62 2.25 3 1.78 1.5 ...
 $ WHCD       : num  3.8 4.2 3.9 3.1 3.2 3.9 3.3 3.4 3.5 3.8 ...
 $ WHCA       : num  1.7 1.63 5.25 2.6 4 5.8 3.3 2.35 5 8 ...
 $ VCCH       : num  4.8 5.2 1.7 3 1.9 1.57 2.45 3.13 1.87 1.55 ...
 $ VCCD       : num  3.8 4.4 3.9 3.13 3.2 4 3.3 3.4 3.5 3.9 ...
 $ VCCA       : num  1.8 1.65 5.5 2.63 5.2 7 3.13 2.4 4.6 8 ...
 $ MaxCH      : num  5.8 6 1.72 3.05 1.95 1.58 2.46 3.38 1.87 1.58 ...
 $ MaxCD      : num  3.9 4.52 3.95 3.29 3.26 4.2 3.42 3.47 3.65 4.05 ...
 $ MaxCA      : num  1.81 1.75 6.2 2.72 5.3 7.3 3.58 2.48 5.35 8.9 ...
 $ AvgCH      : num  5.03 4.93 1.68 2.93 1.9 1.54 2.37 3.05 1.83 1.53 ...
 $ AvgCD      : num  3.66 4.26 3.82 3.14 3.16 4.05 3.25 3.34 3.5 3.84 ...
 $ AvgCA      : num  1.76 1.65 5.37 2.59 4.91 6.66 3.18 2.39 4.74 7.68 ...
 $ B365C.2.5  : num  1.9 1.44 2 2.2 2.75 1.9 2.1 2 2 2.37 ...
 $ B365C.2.5.1: num  1.9 2.75 1.8 1.66 1.44 1.9 1.72 1.8 1.8 1.57 ...
 $ PC.2.5     : num  1.98 1.49 2.06 2.2 2.84 1.95 2.18 2.04 2.03 2.43 ...
 $ PC.2.5.1   : num  1.93 2.76 1.85 1.74 1.47 1.95 1.75 1.85 1.87 1.61 ...
 $ MaxC.2.5   : num  1.99 1.51 2.08 2.38 2.85 1.98 2.18 2.09 2.07 2.46 ...
 $ MaxC.2.5.1 : num  2.11 2.88 1.98 1.74 1.5 2.1 1.83 2.05 1.92 1.65 ...
 $ AvgC.2.5   : num  1.86 1.47 2 2.24 2.69 1.9 2.1 1.97 1.99 2.36 ...
 $ AvgC.2.5.1 : num  1.97 2.63 1.82 1.66 1.46 1.92 1.74 1.85 1.83 1.59 ...
 $ AHCh       : num  0.75 1 -0.75 0 -0.5 -1 -0.25 0.25 -0.75 -1 ...
 $ B365CAHH   : num  1.93 1.82 1.94 2.11 1.89 1.96 2.08 1.86 2.02 2.06 ...
 $ B365CAHA   : num  2 1.97 1.99 1.82 2.04 1.97 1.85 2.07 1.77 1.87 ...
  [list output truncated]
```

```r
head(d1718); head(d1819); head(d1920)
```

```
  Div     Date  HomeTeam   AwayTeam FTHG FTAG FTR HTHG HTAG HTR HS AS HST AST
1 SP1 18/08/17   Leganes     Alaves    1    0   H    1    0   H 16  6   9   3
2 SP1 18/08/17  Valencia Las Palmas    1    0   H    1    0   H 22  5   6   4
3 SP1 19/08/17     Celta   Sociedad    2    3   A    1    1   D 16 13   5   6
```

```
4 SP1 19/08/17     Girona Ath Madrid    2    2   D    2    0   H 13  9   6   3
5 SP1 19/08/17    Sevilla    Espanol    1    1   D    1    1   D  9  9   4   6
6 SP1 20/08/17 Ath Bilbao     Getafe    0    0   D    0    0   D 12  8   2   2
  HF AF HC AC HY AY HR AR B365H B365D B365A  BWH  BWD  BWA  IWH IWD  IWA  LBH
1 14 18  4  2  0  1  0  0  2.05  3.20  4.10 2.05 3.10 4.10 2.10 3.4 3.50 2.05
2 25 13  5  2  3  3  0  1  1.75  3.80  4.50 1.75 3.90 4.60 1.75 3.6 4.80 1.75
3 12 11  5  4  3  1  0  0  2.38  3.25  3.20 2.40 3.30 3.00 2.50 3.3 2.85 2.35
4 15 15  6  0  2  4  0  1  8.00  4.33  1.45 7.50 4.33 1.45 7.20 4.4 1.45 7.50
5 14 12  7  3  2  4  1  0  1.62  4.00  5.50 1.62 3.90 5.75 1.55 4.0 6.20 1.60
6 16 15  7  6  1  3  0  1  1.50  4.00  7.50 1.48 4.25 7.00 1.50 4.2 6.50 1.50
   LBD  LBA  PSH  PSD  PSA  WHH  WHD  WHA  VCH VCD  VCA Bb1X2 BbMxH BbAvH BbMxD
1 3.00 4.20 2.03 3.25 4.52 2.05 3.10 4.00 2.05 3.2 4.40    35  2.12  2.03  3.40
2 3.80 4.33 1.78 4.01 4.83 1.80 3.75 4.20 1.80 4.0 4.60    35  1.83  1.77  4.04
3 3.25 3.00 2.44 3.40 3.16 2.40 3.40 2.90 2.40 3.4 3.13    35  2.50  2.39  3.50
4 4.00 1.50 8.36 4.38 1.49 8.00 4.20 1.44 7.50 4.3 1.50    35  8.36  7.53  4.40
5 3.90 5.50 1.62 4.17 6.18 1.67 3.60 5.50 1.65 4.0 5.75    35  1.69  1.63  4.17
6 4.00 7.00 1.53 4.37 7.31 1.50 4.00 7.00 1.50 4.2 7.00    34  1.53  1.50  4.40
  BbAvD BbMxA BbAvA BbOU BbMx.2.5 BbAv.2.5 BbMx.2.5.1 BbAv.2.5.1 BbAH BbAHh
1  3.15  4.52  4.17   31     2.84     2.68       1.53       1.46   18 -0.50
2  3.86  4.83  4.46   33     1.69     1.64       2.40       2.27   16 -0.75
3  3.32  3.20  3.01   34     2.03     1.98       1.90       1.84   18 -0.25
4  4.17  1.51  1.48   34     2.20     2.11       1.80       1.74   16  1.25
5  3.93  6.20  5.58   33     1.81     1.75       2.14       2.09   16 -1.00
6  4.17  7.50  6.94   32     2.01     1.94       1.96       1.87   17 -1.00
  BbMxAHH BbAvAHH BbMxAHA BbAvAHA PSCH PSCD PSCA
1    2.07    2.03    1.90    1.86 1.98 3.35 4.63
2    2.05    1.97    1.96    1.91 1.78 4.24 4.43
3    2.08    2.05    1.87    1.83 2.12 3.53 3.74
4    1.77    1.75    2.25    2.16 6.93 3.83 1.63
5    2.12    2.06    1.86    1.82 1.64 4.18 5.82
6    1.90    1.86    2.05    2.01 1.53 4.48 6.91
```

```
  Div       Date   HomeTeam   AwayTeam FTHG FTAG FTR HTHG HTAG HTR HS AS HST
1 SP1 17/08/2018      Betis    Levante    0    3   A    0    1   A 22  6   8
2 SP1 17/08/2018     Girona Valladolid    0    0   D    0    0   D 13  2   1
3 SP1 18/08/2018  Barcelona     Alaves    3    0   H    0    0   D 25  3   9
4 SP1 18/08/2018      Celta    Espanol    1    1   D    0    1   A 12 14   2
5 SP1 18/08/2018 Villarreal   Sociedad    1    2   A    1    1   D 16  8   7
6 SP1 19/08/2018      Eibar     Huesca    1    2   A    0    2   A 18  8   6
  AST HF AF HC AC HY AY HR AR B365H B365D B365A  BWH  BWD   BWA  IWH  IWD  IWA
1   4 10 10  5  3  0  2  0  0  1.66  4.00   5.0 1.70  3.7  5.25 1.75 3.60  4.9
2   1 21 20  3  2  1  1  0  0  1.75  3.60   5.0 1.75  3.5  5.25 1.80 3.60  4.5
3   0  6 13  7  1  0  2  0  0  1.11 10.00  21.0 1.11 10.0 20.00 1.12 9.00 20.0
4   5 13 14  8  7  3  2  0  0  1.85  3.50   4.5 1.91  3.4  4.25 1.90 3.50  4.1
5   4 16 10  4  6  2  3  0  0  2.04  3.40   3.8 2.05  3.3  3.90 2.00 3.40  3.8
6   6 12 13  7  0  1  1  0  0  1.66  3.75   5.5 1.70  3.7  5.25 1.70 3.75  5.0
   PSH   PSD   PSA  WHH WHD   WHA  VCH  VCD  VCA Bb1X2 BbMxH BbAvH BbMxD BbAvD
1 1.69  4.19  5.11 1.67 3.9  4.75 1.67  4.2  5.2    40  1.75  1.68  4.25  4.00
2 1.80  3.70  4.99 1.75 3.6  4.60 1.80  3.7  4.8    40  1.85  1.78  3.83  3.60
3 1.11 11.27 25.40 1.08 9.0 29.00 1.10 10.5 34.0    40  1.13  1.10 11.50  9.82
4 1.93  3.64  4.27 1.91 3.5  4.00 1.93  3.5  4.4    38  1.97  1.90  3.73  3.53
5 2.06  3.51  3.91 2.05 3.3  3.60 2.05  3.5  3.9    40  2.11  2.03  3.62  3.43
6 1.72  3.90  5.26 1.73 3.6  4.75 1.70  3.8  5.0    40  1.76  1.70  3.93  3.77
  BbMxA BbAvA BbOU BbMx.2.5 BbAv.2.5 BbMx.2.5.1 BbAv.2.5.1 BbAH BbAHh BbMxAHH
```

```
1  5.25  4.95   38     1.82     1.76       2.15       2.06   20 -0.75    1.89
2  5.27  4.79   38     2.21     2.13       1.78       1.71   20 -0.75    2.06
3 41.00 25.67   32     1.39     1.34       3.40       3.18   19 -2.50    1.95
4  4.50  4.20   36     2.13     2.06       1.84       1.76   18 -0.75    2.26
5  3.93  3.76   37     2.05     1.99       1.88       1.81   18 -0.25    1.76
6  5.50  5.08   37     1.95     1.88       1.98       1.91   19 -0.75    1.96
  BbAvAHH BbMxAHA BbAvAHA PSCH  PSCD  PSCA
1    1.85    2.07    2.00 1.59  4.42  5.89
2    2.01    1.90    1.85 1.76  3.57  5.62
3    1.91    2.00    1.95 1.10 11.85 32.17
4    2.18    1.74    1.71 2.18  3.26  3.85
5    1.74    2.23    2.14 2.32  3.21  3.53
6    1.91    2.01    1.94 1.77  3.68  5.32

  Div       Date  Time   HomeTeam    AwayTeam FTHG FTAG FTR HTHG HTAG HTR HS AS
1 SP1 16/08/2019 20:00 Ath Bilbao   Barcelona    1    0   H    0    0   D 11 11
2 SP1 17/08/2019 16:00      Celta Real Madrid    1    3   A    0    1   A  7 17
3 SP1 17/08/2019 18:00   Valencia    Sociedad    1    1   D    0    0   D 14 12
4 SP1 17/08/2019 19:00   Mallorca       Eibar    2    1   H    1    0   H 16 11
5 SP1 17/08/2019 20:00    Leganes     Osasuna    0    1   A    0    0   D 13  4
6 SP1 17/08/2019 20:00 Villarreal     Granada    4    4   D    1    1   D 12 14
  HST AST HF AF HC AC HY AY HR AR B365H B365D B365A  BWH  BWD  BWA  IWH  IWD
1   5   2 14  9  3  8  1  1  0  0  5.25  3.80  1.65 5.50 3.80 1.65 5.00 3.80
2   4  11 17 12  6  4  5  2  0  1  4.75  4.20  1.65 4.40 4.20 1.72 5.30 4.20
3   6   3 13 14  3  3  4  4  1  0  1.66  3.75  5.50 1.67 3.75 5.50 1.67 3.75
4   4   5 13 14  9  3  2  3  0  0  2.80  3.20  2.60 2.95 3.10 2.60 2.90 3.10
5   2   2 17 11  8  0  1  4  1  0  2.00  3.20  4.20 2.05 3.25 3.90 2.05 3.10
6   7   7 10 16  2  7  3  1  0  0  1.60  3.80  6.50 1.60 3.80 6.25 1.63 4.00
   IWA  PSH  PSD  PSA  WHH WHD  WHA  VCH  VCD  VCA MaxH MaxD MaxA AvgH AvgD
1 1.70 5.15 3.84 1.74 5.00 3.8 1.70 5.00 3.80 1.75 5.50 3.95 1.76 5.07 3.81
2 1.60 4.73 4.18 1.72 5.25 4.2 1.60 4.75 4.20 1.73 5.30 4.40 1.73 4.67 4.12
3 5.30 1.68 3.94 5.47 1.67 3.8 5.25 1.67 3.90 5.75 1.72 3.98 5.75 1.68 3.80
4 2.60 2.98 3.14 2.66 2.90 3.1 2.62 2.90 3.13 2.70 3.05 3.20 2.70 2.91 3.09
5 4.05 2.10 3.21 4.13 2.05 3.2 4.00 2.10 3.20 4.10 2.10 3.30 4.25 2.06 3.18
6 5.50 1.62 3.99 6.13 1.60 3.9 5.80 1.65 4.00 5.75 1.65 4.15 6.50 1.61 3.95
  AvgA B365.2.5 B365.2.5.1 P.2.5 P.2.5.1 Max.2.5 Max.2.5.1 Avg.2.5 Avg.2.5.1
1 1.71     1.80       2.00  1.81    2.09    1.85      2.11    1.79      2.05
2 1.69     1.53       2.50  1.52    2.66    1.53      2.72    1.49      2.58
3 5.29     2.00       1.80  2.08    1.82    2.14      1.83    2.07      1.77
4 2.62     2.30       1.61  2.45    1.60    2.47      1.65    2.34      1.60
5 4.02     2.50       1.53  2.72    1.50    2.75      1.54    2.59      1.49
6 5.80     1.80       2.00  1.88    2.02    1.90      2.05    1.84      1.98
    AHh B365AHH B365AHA PAHH PAHA MaxAHH MaxAHA AvgAHH AvgAHA B365CH B365CD
1  0.75    1.99    1.94 1.98 1.94   2.00   1.95   1.96   1.92   5.25   3.80
2  0.75    2.04    1.89 2.01 1.91   2.05   1.91   2.00   1.88   5.25   4.20
3 -0.75    1.91    2.02 1.91 2.01   1.93   2.03   1.89   1.99   1.66   3.75
4  0.00    2.05    1.88 2.07 1.85   2.07   1.88   2.04   1.85   2.87   3.20
5 -0.50    2.08    1.85 2.10 1.82   2.10   1.85   2.06   1.83   1.90   3.10
6 -1.00    2.05    1.75 2.11 1.81   2.14   1.85   2.07   1.80   1.53   4.00
  B365CA BWCH BWCD BWCA IWCH IWCD IWCA PSCH PSCD PSCA WHCH WHCD WHCA VCCH VCCD
1   1.65 4.75 3.75 1.75 5.00 3.80 1.70 5.34 3.62 1.78 5.00  3.8 1.70 4.80 3.80
2   1.57 4.50 4.10 1.70 4.60 3.80 1.75 5.10 4.46 1.65 5.00  4.2 1.63 5.20 4.40
3   5.50 1.65 3.80 5.50 1.67 3.80 5.30 1.69 3.88 5.47 1.65  3.9 5.25 1.70 3.90
4   2.55 2.95 3.10 2.60 2.90 3.10 2.60 2.96 3.26 2.60 2.90  3.1 2.60 3.00 3.13
```

```
5   5.00 1.95 3.20 4.50 1.90 3.15 4.85 1.90 3.18 5.30 2.05  3.2 4.00 1.90 3.20
6   6.50 1.57 3.80 6.50 1.55 4.05 6.30 1.54 4.19 6.87 1.62  3.9 5.80 1.57 4.00
  VCCA MaxCH MaxCD MaxCA AvgCH AvgCD AvgCA B365C.2.5 B365C.2.5.1 PC.2.5
1 1.80  5.80  3.90  1.81  5.03  3.66  1.76      1.90        1.90   1.98
2 1.65  6.00  4.52  1.75  4.93  4.26  1.65      1.44        2.75   1.49
3 5.50  1.72  3.95  6.20  1.68  3.82  5.37      2.00        1.80   2.06
4 2.63  3.05  3.29  2.72  2.93  3.14  2.59      2.20        1.66   2.20
5 5.20  1.95  3.26  5.30  1.90  3.16  4.91      2.75        1.44   2.84
6 7.00  1.58  4.20  7.30  1.54  4.05  6.66      1.90        1.90   1.95
  PC.2.5.1 MaxC.2.5 MaxC.2.5.1 AvgC.2.5 AvgC.2.5.1  AHCh B365CAHH B365CAHA
1     1.93     1.99       2.11     1.86       1.97  0.75     1.93     2.00
2     2.76     1.51       2.88     1.47       2.63  1.00     1.82     1.97
3     1.85     2.08       1.98     2.00       1.82 -0.75     1.94     1.99
4     1.74     2.38       1.74     2.24       1.66  0.00     2.11     1.82
5     1.47     2.85       1.50     2.69       1.46 -0.50     1.89     2.04
6     1.95     1.98       2.10     1.90       1.92 -1.00     1.96     1.97
  PCAHH PCAHA MaxCAHH MaxCAHA AvgCAHH AvgCAHA
1  1.91  2.01    2.02    2.03    1.91    1.98
2  1.85  2.07    2.00    2.20    1.82    2.06
3  1.92  2.00    1.96    2.12    1.89    2.00
4  2.09  1.83    2.12    1.88    2.07    1.83
5  1.90  2.01    1.95    2.06    1.90    1.99
6  1.96  1.96    1.98    2.12    1.93    1.95
```

```r
View(d1718); View(d1819); View(d1920)
summary(d1718); summary(d1819); summary(d1920)
```

```
     Div                Date             HomeTeam           AwayTeam        
 Length:380         Length:380         Length:380         Length:380        
 Class :character   Class :character   Class :character   Class :character  
 Mode  :character   Mode  :character   Mode  :character   Mode  :character  




      FTHG            FTAG            FTR                 HTHG       
 Min.   :0.000   Min.   :0.000   Length:380         Min.   :0.0000  
 1st Qu.:0.750   1st Qu.:0.000   Class :character   1st Qu.:0.0000  
 Median :1.000   Median :1.000   Mode  :character   Median :0.0000  
 Mean   :1.547   Mean   :1.147                      Mean   :0.6605  
 3rd Qu.:2.000   3rd Qu.:2.000                      3rd Qu.:1.0000  
 Max.   :7.000   Max.   :6.000                      Max.   :5.0000  

      HTAG            HTR                  HS              AS       
 Min.   :0.0000   Length:380         Min.   : 2.00   Min.   : 1.00  
 1st Qu.:0.0000   Class :character   1st Qu.:10.00   1st Qu.: 8.00  
 Median :0.0000   Mode  :character   Median :13.00   Median :10.00  
 Mean   :0.4868                      Mean   :13.53   Mean   :10.47  
 3rd Qu.:1.0000                      3rd Qu.:16.00   3rd Qu.:13.00  
 Max.   :3.0000                      Max.   :30.00   Max.   :24.00  

      HST              AST               HF              AF       
 Min.   : 0.000   Min.   : 0.000   Min.   : 4.00   Min.   : 0.00  
 1st Qu.: 3.000   1st Qu.: 2.000   1st Qu.:11.00   1st Qu.:11.00  
```

```
 Median : 4.500   Median : 3.000   Median :13.00   Median :14.00  
 Mean   : 4.758   Mean   : 3.805   Mean   :13.73   Mean   :13.95  
 3rd Qu.: 6.000   3rd Qu.: 5.000   3rd Qu.:17.00   3rd Qu.:17.00  
 Max.   :14.000   Max.   :13.000   Max.   :29.00   Max.   :29.00  
                                                                  
       HC               AC               HY              AY       
 Min.   : 0.000   Min.   : 0.000   Min.   :0.000   Min.   :0.000  
 1st Qu.: 4.000   1st Qu.: 2.000   1st Qu.:1.000   1st Qu.:2.000  
 Median : 5.000   Median : 4.000   Median :2.000   Median :3.000  
 Mean   : 5.613   Mean   : 4.192   Mean   :2.339   Mean   :2.676  
 3rd Qu.: 7.000   3rd Qu.: 6.000   3rd Qu.:3.000   3rd Qu.:4.000  
 Max.   :16.000   Max.   :14.000   Max.   :8.000   Max.   :9.000  
                                                                  
       HR               AR               B365H            B365D       
 Min.   :0.0000   Min.   :0.00000   Min.   : 1.050   Min.   : 2.790  
 1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.: 1.617   1st Qu.: 3.290  
 Median :0.0000   Median :0.00000   Median : 2.075   Median : 3.500  
 Mean   :0.1105   Mean   :0.07895   Mean   : 2.777   Mean   : 4.259  
 3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.: 2.790   3rd Qu.: 4.330  
 Max.   :2.0000   Max.   :2.00000   Max.   :17.000   Max.   :15.000  
                                                                     
     B365A             BWH              BWD              BWA        
 Min.   : 1.170   Min.   : 1.050   Min.   : 2.950   Min.   : 1.180  
 1st Qu.: 2.600   1st Qu.: 1.650   1st Qu.: 3.300   1st Qu.: 2.600  
 Median : 3.700   Median : 2.100   Median : 3.600   Median : 3.700  
 Mean   : 5.192   Mean   : 2.744   Mean   : 4.278   Mean   : 5.204  
 3rd Qu.: 5.500   3rd Qu.: 2.750   3rd Qu.: 4.330   3rd Qu.: 5.500  
 Max.   :34.000   Max.   :14.500   Max.   :15.500   Max.   :34.000  
                                                                    
      IWH              IWD              IWA              LBH        
 Min.   : 1.070   Min.   : 3.050   Min.   : 1.200   Min.   : 1.050  
 1st Qu.: 1.650   1st Qu.: 3.300   1st Qu.: 2.600   1st Qu.: 1.610  
 Median : 2.100   Median : 3.500   Median : 3.500   Median : 2.050  
 Mean   : 2.721   Mean   : 4.161   Mean   : 5.041   Mean   : 2.742  
 3rd Qu.: 2.700   3rd Qu.: 4.200   3rd Qu.: 5.300   3rd Qu.: 2.750  
 Max.   :15.000   Max.   :12.000   Max.   :27.000   Max.   :19.000  
                                                    NA's   :1       
      LBD              LBA              PSH              PSD        
 Min.   : 2.900   Min.   : 1.170   Min.   : 1.050   Min.   : 3.020  
 1st Qu.: 3.250   1st Qu.: 2.575   1st Qu.: 1.660   1st Qu.: 3.410  
 Median : 3.500   Median : 3.600   Median : 2.120   Median : 3.705  
 Mean   : 4.152   Mean   : 5.375   Mean   : 2.857   Mean   : 4.539  
 3rd Qu.: 4.200   3rd Qu.: 5.500   3rd Qu.: 2.850   3rd Qu.: 4.455  
 Max.   :17.000   Max.   :41.000   Max.   :19.650   Max.   :20.380  
 NA's   :1        NA's   :1                                         
      PSA              WHH              WHD              WHA        
 Min.   : 1.180   Min.   : 1.060   Min.   : 2.900   Min.   : 1.170  
 1st Qu.: 2.670   1st Qu.: 1.665   1st Qu.: 3.250   1st Qu.: 2.600  
 Median : 3.845   Median : 2.100   Median : 3.500   Median : 3.550  
 Mean   : 5.522   Mean   : 2.738   Mean   : 4.092   Mean   : 5.041  
 3rd Qu.: 5.942   3rd Qu.: 2.750   3rd Qu.: 4.200   3rd Qu.: 5.500  
 Max.   :36.500   Max.   :17.000   Max.   :15.000   Max.   :26.000  
                                                                    
      VCH              VCD              VCA              Bb1X2      
```

```
 Min.   : 1.040   Min.   : 3.000   Min.   : 1.180   Min.   : 3.00
 1st Qu.: 1.650   1st Qu.: 3.400   1st Qu.: 2.630   1st Qu.:35.00
 Median : 2.100   Median : 3.700   Median : 3.700   Median :37.00
 Mean   : 2.762   Mean   : 4.416   Mean   : 5.472   Mean   :37.71
 3rd Qu.: 2.800   3rd Qu.: 4.400   3rd Qu.: 5.750   3rd Qu.:40.00
 Max.   :15.000   Max.   :17.000   Max.   :36.000   Max.   :43.00

     BbMxH            BbAvH            BbMxD            BbAvD
 Min.   : 1.080   Min.   : 1.050   Min.   : 3.110   Min.   : 2.940
 1st Qu.: 1.700   1st Qu.: 1.640   1st Qu.: 3.478   1st Qu.: 3.328
 Median : 2.200   Median : 2.090   Median : 3.750   Median : 3.570
 Mean   : 2.966   Mean   : 2.743   Mean   : 4.636   Mean   : 4.261
 3rd Qu.: 2.882   3rd Qu.: 2.765   3rd Qu.: 4.553   3rd Qu.: 4.272
 Max.   :19.650   Max.   :16.300   Max.   :20.380   Max.   :15.320

     BbMxA            BbAvA             BbOU         BbMx.2.5
 Min.   : 1.210   Min.   : 1.170   Min.   : 3.00   Min.   :1.130
 1st Qu.: 2.728   1st Qu.: 2.607   1st Qu.:31.75   1st Qu.:1.667
 Median : 3.920   Median : 3.665   Median :34.00   Median :1.960
 Mean   : 6.107   Mean   : 5.190   Mean   :34.06   Mean   :1.950
 3rd Qu.: 6.105   3rd Qu.: 5.543   3rd Qu.:37.00   3rd Qu.:2.203
 Max.   :67.000   Max.   :33.420   Max.   :42.00   Max.   :3.080

    BbAv.2.5       BbMx.2.5.1      BbAv.2.5.1         BbAH
 Min.   :1.120   Min.   :1.470   Min.   :1.410   Min.   : 1.00
 1st Qu.:1.617   1st Qu.:1.780   1st Qu.:1.718   1st Qu.:17.00
 Median :1.880   Median :2.000   Median :1.920   Median :18.00
 Mean   :1.872   Mean   :2.284   Mean   :2.162   Mean   :18.16
 3rd Qu.:2.120   3rd Qu.:2.402   3rd Qu.:2.283   3rd Qu.:19.00
 Max.   :2.850   Max.   :7.000   Max.   :5.970   Max.   :24.00

     BbAHh             BbMxAHH         BbAvAHH         BbMxAHA
 Min.   :-3.2500   Min.   :1.610   Min.   :1.580   Min.   :1.680
 1st Qu.:-0.7500   1st Qu.:1.890   1st Qu.:1.840   1st Qu.:1.897
 Median :-0.2500   Median :1.985   Median :1.930   Median :1.970
 Mean   :-0.4059   Mean   :1.988   Mean   :1.938   Mean   :1.988
 3rd Qu.: 0.0625   3rd Qu.:2.070   3rd Qu.:2.020   3rd Qu.:2.080
 Max.   : 2.0000   Max.   :2.420   Max.   :2.340   Max.   :2.520

    BbAvAHA           PSCH             PSCD             PSCA
 Min.   :1.630   Min.   : 1.060   Min.   : 2.930   Min.   : 1.160
 1st Qu.:1.850   1st Qu.: 1.640   1st Qu.: 3.410   1st Qu.: 2.590
 Median :1.930   Median : 2.120   Median : 3.700   Median : 3.850
 Mean   :1.937   Mean   : 2.839   Mean   : 4.508   Mean   : 5.695
 3rd Qu.:2.030   3rd Qu.: 2.980   3rd Qu.: 4.560   3rd Qu.: 6.095
 Max.   :2.440   Max.   :18.700   Max.   :18.500   Max.   :46.000
                 NA's   :1        NA's   :1        NA's   :1

     Div                Date             HomeTeam           AwayTeam
 Length:380         Length:380         Length:380         Length:380
 Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character
```

```
     FTHG            FTAG           FTR                  HTHG       
 Min.   :0.000   Min.   :0.000   Length:380         Min.   :0.0000  
 1st Qu.:1.000   1st Qu.:0.000   Class :character   1st Qu.:0.0000  
 Median :1.000   Median :1.000   Mode  :character   Median :0.0000  
 Mean   :1.453   Mean   :1.134                      Mean   :0.5447  
 3rd Qu.:2.000   3rd Qu.:2.000                      3rd Qu.:1.0000  
 Max.   :8.000   Max.   :6.000                      Max.   :3.0000  
     HTAG             HTR                  HS              AS       
 Min.   :0.0000   Length:380         Min.   : 3.00   Min.   : 2.00  
 1st Qu.:0.0000   Class :character   1st Qu.:10.00   1st Qu.: 8.00  
 Median :0.0000   Mode  :character   Median :13.00   Median :10.00  
 Mean   :0.5132                      Mean   :13.87   Mean   :10.43  
 3rd Qu.:1.0000                      3rd Qu.:17.00   3rd Qu.:13.00  
 Max.   :5.0000                      Max.   :34.00   Max.   :21.00  
     HST              AST               HF              AF       
 Min.   : 0.000   Min.   : 0.000   Min.   : 1.00   Min.   : 3.00  
 1st Qu.: 3.000   1st Qu.: 2.000   1st Qu.:11.00   1st Qu.:11.00  
 Median : 5.000   Median : 3.000   Median :13.00   Median :13.00  
 Mean   : 4.834   Mean   : 3.589   Mean   :13.63   Mean   :13.45  
 3rd Qu.: 6.000   3rd Qu.: 5.000   3rd Qu.:16.00   3rd Qu.:16.00  
 Max.   :15.000   Max.   :11.000   Max.   :26.00   Max.   :27.00  
       HC               AC               HY              AY       
 Min.   : 0.000   Min.   : 0.000   Min.   :0.000   Min.   :0.000  
 1st Qu.: 4.000   1st Qu.: 2.000   1st Qu.:1.000   1st Qu.:2.000  
 Median : 5.000   Median : 4.000   Median :2.000   Median :3.000  
 Mean   : 5.574   Mean   : 4.021   Mean   :2.529   Mean   :2.642  
 3rd Qu.: 7.000   3rd Qu.: 6.000   3rd Qu.:4.000   3rd Qu.:3.000  
 Max.   :15.000   Max.   :12.000   Max.   :8.000   Max.   :7.000  
       HR                AR             B365H            B365D       
 Min.   :0.00000   Min.   :0.0000   Min.   : 1.080   Min.   : 2.870  
 1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.: 1.660   1st Qu.: 3.300  
 Median :0.00000   Median :0.0000   Median : 2.120   Median : 3.500  
 Mean   :0.08684   Mean   :0.1211   Mean   : 2.596   Mean   : 3.996  
 3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.: 2.800   3rd Qu.: 4.000  
 Max.   :1.00000   Max.   :2.0000   Max.   :17.000   Max.   :11.000  
     B365A             BWH              BWD              BWA        
 Min.   : 1.160   Min.   : 1.060   Min.   : 2.900   Min.   : 1.190  
 1st Qu.: 2.547   1st Qu.: 1.670   1st Qu.: 3.300   1st Qu.: 2.600  
 Median : 3.500   Median : 2.150   Median : 3.500   Median : 3.500  
 Mean   : 4.790   Mean   : 2.579   Mean   : 3.991   Mean   : 4.745  
 3rd Qu.: 5.062   3rd Qu.: 2.800   3rd Qu.: 4.000   3rd Qu.: 5.250  
 Max.   :29.000   Max.   :15.000   Max.   :12.000   Max.   :36.000  
      IWH              IWD              IWA              PSH        
 Min.   : 1.070   Min.   : 2.850   Min.   : 1.200   Min.   : 1.080  
 1st Qu.: 1.700   1st Qu.: 3.300   1st Qu.: 2.600   1st Qu.: 1.700  
 Median : 2.150   Median : 3.500   Median : 3.450   Median : 2.180  
 Mean   : 2.553   Mean   : 3.943   Mean   : 4.587   Mean   : 2.639  
 3rd Qu.: 2.763   3rd Qu.: 4.000   3rd Qu.: 4.950   3rd Qu.: 2.840  
 Max.   :13.000   Max.   :13.000   Max.   :28.000   Max.   :19.070  
      PSD              PSA              WHH              WHD        
 Min.   : 2.990   Min.   : 1.180   Min.   : 1.050   Min.   : 2.500  
 1st Qu.: 3.357   1st Qu.: 2.652   1st Qu.: 1.700   1st Qu.: 3.300  
 Median : 3.640   Median : 3.575   Median : 2.150   Median : 3.500  
```

```
 Mean   : 4.133   Mean   : 4.994   Mean   : 2.564   Mean   : 3.997  
 3rd Qu.: 4.170   3rd Qu.: 5.230   3rd Qu.: 2.800   3rd Qu.: 4.000  
 Max.   :13.220   Max.   :36.830   Max.   :17.000   Max.   :13.000  
      WHA              VCH              VCD              VCA       
 Min.   : 1.150   Min.   : 1.060   Min.   : 3.000   Min.   : 1.18  
 1st Qu.: 2.587   1st Qu.: 1.700   1st Qu.: 3.300   1st Qu.: 2.60  
 Median : 3.500   Median : 2.150   Median : 3.600   Median : 3.60  
 Mean   : 4.779   Mean   : 2.627   Mean   : 4.097   Mean   : 5.00  
 3rd Qu.: 5.000   3rd Qu.: 2.800   3rd Qu.: 4.200   3rd Qu.: 5.20  
 Max.   :34.000   Max.   :21.000   Max.   :13.000   Max.   :41.00  
     Bb1X2           BbMxH            BbAvH            BbMxD       
 Min.   :31.00   Min.   : 1.100   Min.   : 1.070   Min.   : 3.040  
 1st Qu.:34.00   1st Qu.: 1.750   1st Qu.: 1.690   1st Qu.: 3.420  
 Median :36.00   Median : 2.250   Median : 2.160   Median : 3.735  
 Mean   :36.13   Mean   : 2.739   Mean   : 2.582   Mean   : 4.251  
 3rd Qu.:38.00   3rd Qu.: 2.913   3rd Qu.: 2.792   3rd Qu.: 4.250  
 Max.   :41.00   Max.   :21.000   Max.   :16.550   Max.   :15.000  
     BbAvD            BbMxA            BbAvA             BbOU      
 Min.   : 2.920   Min.   : 1.200   Min.   : 1.170   Min.   :28.00  
 1st Qu.: 3.288   1st Qu.: 2.750   1st Qu.: 2.620   1st Qu.:32.00  
 Median : 3.550   Median : 3.695   Median : 3.495   Median :34.00  
 Mean   : 4.005   Mean   : 5.361   Mean   : 4.763   Mean   :33.84  
 3rd Qu.: 4.032   3rd Qu.: 5.370   3rd Qu.: 5.043   3rd Qu.:35.00  
 Max.   :12.430   Max.   :52.000   Max.   :33.380   Max.   :39.00  
    BbMx.2.5        BbAv.2.5       BbMx.2.5.1      BbAv.2.5.1   
 Min.   :1.200   Min.   :1.170   Min.   :1.420   Min.   :1.380  
 1st Qu.:1.710   1st Qu.:1.650   1st Qu.:1.710   1st Qu.:1.650  
 Median :2.000   Median :1.940   Median :1.950   Median :1.870  
 Mean   :2.029   Mean   :1.947   Mean   :2.116   Mean   :2.018  
 3rd Qu.:2.315   3rd Qu.:2.230   3rd Qu.:2.303   3rd Qu.:2.210  
 Max.   :3.200   Max.   :2.890   Max.   :5.250   Max.   :4.740  
      BbAH           BbAHh            BbMxAHH         BbAvAHH     
 Min.   :15.00   Min.   :-3.0000   Min.   :1.560   Min.   :1.520  
 1st Qu.:19.00   1st Qu.:-1.0000   1st Qu.:1.850   1st Qu.:1.800  
 Median :20.00   Median :-0.2500   Median :2.030   Median :1.970  
 Mean   :19.88   Mean   :-0.4033   Mean   :2.055   Mean   :1.992  
 3rd Qu.:21.00   3rd Qu.: 0.2500   3rd Qu.:2.200   3rd Qu.:2.130  
 Max.   :24.00   Max.   : 2.0000   Max.   :3.270   Max.   :3.020  
    BbMxAHA         BbAvAHA           PSCH             PSCD       
 Min.   :1.450   Min.   :1.410   Min.   : 1.070   Min.   : 2.860  
 1st Qu.:1.800   1st Qu.:1.750   1st Qu.: 1.698   1st Qu.: 3.310  
 Median :1.950   Median :1.890   Median : 2.190   Median : 3.610  
 Mean   :1.972   Mean   :1.915   Mean   : 2.726   Mean   : 4.102  
 3rd Qu.:2.130   3rd Qu.:2.070   3rd Qu.: 2.970   3rd Qu.: 4.202  
 Max.   :2.850   Max.   :2.670   Max.   :18.040   Max.   :14.910  
      PSCA       
 Min.   : 1.190  
 1st Qu.: 2.612  
 Median : 3.645  
 Mean   : 5.100  
 3rd Qu.: 5.575  
 Max.   :36.030  

     Div                   Date                   Time                HomeTeam
```

```
 Length:380          Length:380          Length:380          Length:380
 Class :character    Class :character    Class :character    Class :character
 Mode  :character    Mode  :character    Mode  :character    Mode  :character




   AwayTeam               FTHG            FTAG            FTR
 Length:380          Min.   :0.000   Min.   :0.000   Length:380
 Class :character    1st Qu.:1.000   1st Qu.:0.000   Class :character
 Mode  :character    Median :1.000   Median :1.000   Mode  :character
                     Mean   :1.437   Mean   :1.042
                     3rd Qu.:2.000   3rd Qu.:2.000
                     Max.   :6.000   Max.   :5.000

      HTHG              HTAG           HTR                  HS
 Min.   :0.0000   Min.   :0.00   Length:380          Min.   : 3.00
 1st Qu.:0.0000   1st Qu.:0.00   Class :character    1st Qu.: 9.00
 Median :0.0000   Median :0.00   Mode  :character    Median :12.00
 Mean   :0.6026   Mean   :0.45                       Mean   :12.46
 3rd Qu.:1.0000   3rd Qu.:1.00                       3rd Qu.:15.00
 Max.   :4.0000   Max.   :3.00                       Max.   :25.00

       AS             HST              AST               HF
 Min.   : 1.00   Min.   : 0.000   Min.   : 0.000   Min.   : 4.00
 1st Qu.: 7.00   1st Qu.: 3.000   1st Qu.: 2.000   1st Qu.:11.00
 Median :10.00   Median : 4.000   Median : 3.000   Median :13.00
 Mean   :10.14   Mean   : 4.337   Mean   : 3.511   Mean   :13.66
 3rd Qu.:12.25   3rd Qu.: 6.000   3rd Qu.: 5.000   3rd Qu.:16.00
 Max.   :24.00   Max.   :17.000   Max.   :12.000   Max.   :28.00

       AF              HC               AC                HY
 Min.   : 5.00   Min.   : 0.000   Min.   : 0.000   Min.   :0.000
 1st Qu.:11.00   1st Qu.: 3.000   1st Qu.: 2.750   1st Qu.:1.000
 Median :13.00   Median : 5.000   Median : 4.000   Median :2.000
 Mean   :13.79   Mean   : 5.042   Mean   : 4.195   Mean   :2.547
 3rd Qu.:16.00   3rd Qu.: 7.000   3rd Qu.: 6.000   3rd Qu.:4.000
 Max.   :30.00   Max.   :14.000   Max.   :12.000   Max.   :7.000

       AY              HR            AR               B365H
 Min.   :0.000   Min.   :0.0   Min.   :0.0000   Min.   : 1.120
 1st Qu.:2.000   1st Qu.:0.0   1st Qu.:0.0000   1st Qu.: 1.700
 Median :2.000   Median :0.0   Median :0.0000   Median : 2.200
 Mean   :2.584   Mean   :0.1   Mean   :0.1263   Mean   : 2.595
 3rd Qu.:4.000   3rd Qu.:0.0   3rd Qu.:0.0000   3rd Qu.: 2.900
 Max.   :8.000   Max.   :2.0   Max.   :2.0000   Max.   :10.000

     B365D           B365A             BWH               BWD
 Min.   :2.800   Min.   : 1.280   Min.   : 1.120   Min.   :2.700
 1st Qu.:3.200   1st Qu.: 2.500   1st Qu.: 1.700   1st Qu.:3.200
 Median :3.400   Median : 3.500   Median : 2.200   Median :3.400
 Mean   :3.805   Mean   : 4.465   Mean   : 2.599   Mean   :3.813
 3rd Qu.:4.000   3rd Qu.: 5.062   3rd Qu.: 2.900   3rd Qu.:4.000
 Max.   :9.500   Max.   :21.000   Max.   :10.000   Max.   :9.500
```

```
      BWA              IWH              IWD              IWA        
 Min.   : 1.280   Min.   :1.130   Min.   :2.750   Min.   : 1.280  
 1st Qu.: 2.550   1st Qu.:1.730   1st Qu.:3.188   1st Qu.: 2.550  
 Median : 3.500   Median :2.200   Median :3.400   Median : 3.450  
 Mean   : 4.438   Mean   :2.603   Mean   :3.770   Mean   : 4.378  
 3rd Qu.: 5.250   3rd Qu.:2.913   3rd Qu.:4.000   3rd Qu.: 5.000  
 Max.   :20.000   Max.   :9.900   Max.   :8.800   Max.   :19.500  
                                                                  
      PSH              PSD              PSA              WHH        
 Min.   : 1.120   Min.   :2.780   Min.   : 1.290   Min.   : 1.110  
 1st Qu.: 1.720   1st Qu.:3.230   1st Qu.: 2.600   1st Qu.: 1.700  
 Median : 2.260   Median :3.515   Median : 3.520   Median : 2.225  
 Mean   : 2.650   Mean   :3.894   Mean   : 4.673   Mean   : 2.616  
 3rd Qu.: 2.958   3rd Qu.:4.088   3rd Qu.: 5.230   3rd Qu.: 2.900  
 Max.   :10.020   Max.   :9.950   Max.   :25.500   Max.   :10.000  
 NA's   :2        NA's   :2       NA's   :2                        
      WHD             WHA              VCH              VCD       
 Min.   :2.800   Min.   : 1.270   Min.   : 1.100   Min.   :2.800  
 1st Qu.:3.200   1st Qu.: 2.550   1st Qu.: 1.700   1st Qu.:3.200  
 Median :3.400   Median : 3.450   Median : 2.200   Median :3.500  
 Mean   :3.802   Mean   : 4.623   Mean   : 2.582   Mean   :3.831  
 3rd Qu.:4.000   3rd Qu.: 5.250   3rd Qu.: 2.885   3rd Qu.:4.025  
 Max.   :9.000   Max.   :26.000   Max.   :10.500   Max.   :9.500  
                                                                  
      VCA              MaxH              MaxD              MaxA       
 Min.   : 1.250   Min.   : 1.160   Min.   : 2.900   Min.   : 1.320  
 1st Qu.: 2.500   1st Qu.: 1.778   1st Qu.: 3.340   1st Qu.: 2.678  
 Median : 3.400   Median : 2.310   Median : 3.600   Median : 3.625  
 Mean   : 4.453   Mean   : 2.757   Mean   : 4.019   Mean   : 4.951  
 3rd Qu.: 5.000   3rd Qu.: 3.055   3rd Qu.: 4.202   3rd Qu.: 5.500  
 Max.   :26.000   Max.   :11.000   Max.   :10.500   Max.   :31.000  
                                                                  
      AvgH            AvgD            AvgA            B365.2.5    
 Min.   :1.120   Min.   :2.780   Min.   : 1.280   Min.   :1.280  
 1st Qu.:1.718   1st Qu.:3.210   1st Qu.: 2.550   1st Qu.:1.800  
 Median :2.220   Median :3.435   Median : 3.455   Median :2.100  
 Mean   :2.608   Mean   :3.813   Mean   : 4.483   Mean   :2.108  
 3rd Qu.:2.917   3rd Qu.:4.022   3rd Qu.: 5.093   3rd Qu.:2.500  
 Max.   :9.910   Max.   :9.160   Max.   :21.610   Max.   :3.400  
                                                                  
   B365.2.5.1        P.2.5          P.2.5.1         Max.2.5     
 Min.   :1.330   Min.   :1.310   Min.   :1.370   Min.   :1.310  
 1st Qu.:1.530   1st Qu.:1.810   1st Qu.:1.570   1st Qu.:1.850  
 Median :1.720   Median :2.145   Median :1.770   Median :2.175  
 Mean   :1.867   Mean   :2.162   Mean   :1.912   Mean   :2.196  
 3rd Qu.:2.000   3rd Qu.:2.520   3rd Qu.:2.098   3rd Qu.:2.553  
 Max.   :3.750   Max.   :3.340   Max.   :3.640   Max.   :3.400  
                 NA's   :2       NA's   :2                       
   Max.2.5.1        Avg.2.5         Avg.2.5.1          AHh         
 Min.   :1.380   Min.   :1.260   Min.   :1.350   Min.   :-2.5000  
 1st Qu.:1.610   1st Qu.:1.780   1st Qu.:1.560   1st Qu.:-0.7500  
 Median :1.820   Median :2.080   Median :1.750   Median :-0.2500  
 Mean   :1.950   Mean   :2.099   Mean   :1.869   Mean   :-0.3289  
```

```
3rd Qu.:2.132   3rd Qu.:2.440   3rd Qu.:2.040   3rd Qu.: 0.0000
Max.   :3.950   Max.   :3.160   Max.   :3.680   Max.   : 1.7500

   B365AHH         B365AHA          PAHH            PAHA
Min.   :1.670   Min.   :1.650   Min.   :1.750   Min.   :1.680
1st Qu.:1.900   1st Qu.:1.890   1st Qu.:1.890   1st Qu.:1.880
Median :1.970   Median :1.960   Median :1.970   Median :1.950
Mean   :1.963   Mean   :1.954   Mean   :1.966   Mean   :1.954
3rd Qu.:2.040   3rd Qu.:2.030   3rd Qu.:2.040   3rd Qu.:2.030
Max.   :2.200   Max.   :2.160   Max.   :2.310   Max.   :2.200
NA's   :10      NA's   :10      NA's   :2       NA's   :2
    MaxAHH          MaxAHA          AvgAHH          AvgAHA
Min.   :1.800   Min.   :1.710   Min.   :1.740   Min.   :1.670
1st Qu.:1.920   1st Qu.:1.910   1st Qu.:1.870   1st Qu.:1.860
Median :1.990   Median :1.980   Median :1.940   Median :1.930
Mean   :1.995   Mean   :1.988   Mean   :1.939   Mean   :1.932
3rd Qu.:2.070   3rd Qu.:2.060   3rd Qu.:2.010   3rd Qu.:2.000
Max.   :2.330   Max.   :2.250   Max.   :2.270   Max.   :2.170

    B365CH           B365CD           B365CA            BWCH
Min.   : 1.100   Min.   : 2.700   Min.   : 1.250   Min.   :1.100
1st Qu.: 1.700   1st Qu.: 3.200   1st Qu.: 2.600   1st Qu.:1.715
Median : 2.150   Median : 3.400   Median : 3.550   Median :2.200
Mean   : 2.598   Mean   : 3.857   Mean   : 4.628   Mean   :2.603
3rd Qu.: 2.870   3rd Qu.: 4.000   3rd Qu.: 5.250   3rd Qu.:2.900
Max.   :11.000   Max.   :10.000   Max.   :26.000   Max.   :9.750

     BWCD             BWCA            IWCH             IWCD
Min.   : 2.750   Min.   : 1.28   Min.   : 1.120   Min.   :2.700
1st Qu.: 3.200   1st Qu.: 2.60   1st Qu.: 1.730   1st Qu.:3.150
Median : 3.400   Median : 3.50   Median : 2.225   Median :3.400
Mean   : 3.817   Mean   : 4.55   Mean   : 2.609   Mean   :3.747
3rd Qu.: 4.000   3rd Qu.: 5.25   3rd Qu.: 2.900   3rd Qu.:4.000
Max.   :10.000   Max.   :23.00   Max.   :11.000   Max.   :9.000

     IWCA             PSCH             PSCD             PSCA
Min.   : 1.250   Min.   : 1.100   Min.   : 2.710   Min.   : 1.270
1st Qu.: 2.600   1st Qu.: 1.720   1st Qu.: 3.188   1st Qu.: 2.688
Median : 3.500   Median : 2.275   Median : 3.480   Median : 3.640
Mean   : 4.368   Mean   : 2.672   Mean   : 3.900   Mean   : 4.873
3rd Qu.: 5.100   3rd Qu.: 2.975   3rd Qu.: 4.062   3rd Qu.: 5.440
Max.   :20.000   Max.   :10.930   Max.   :11.520   Max.   :28.530

     WHCH             WHCD             WHCA             VCCH
Min.   : 1.080   Min.   : 2.620   Min.   : 1.250   Min.   : 1.080
1st Qu.: 1.700   1st Qu.: 3.200   1st Qu.: 2.600   1st Qu.: 1.722
Median : 2.225   Median : 3.400   Median : 3.600   Median : 2.225
Mean   : 2.647   Mean   : 3.827   Mean   : 4.793   Mean   : 2.600
3rd Qu.: 2.900   3rd Qu.: 4.000   3rd Qu.: 5.250   3rd Qu.: 2.880
Max.   :11.000   Max.   :11.000   Max.   :26.000   Max.   :10.500

     VCCD             VCCA            MaxCH            MaxCD
Min.   : 2.750   Min.   : 1.250   Min.   : 1.130   Min.   : 2.860
1st Qu.: 3.200   1st Qu.: 2.600   1st Qu.: 1.788   1st Qu.: 3.320
```

```
 Median : 3.450   Median : 3.400   Median : 2.355   Median : 3.610  
 Mean   : 3.845   Mean   : 4.549   Mean   : 2.836   Mean   : 4.062  
 3rd Qu.: 4.000   3rd Qu.: 5.050   3rd Qu.: 3.105   3rd Qu.: 4.242  
 Max.   :10.500   Max.   :26.000   Max.   :13.000   Max.   :12.400  
                                                                    
     MaxCA            AvgCH            AvgCD            AvgCA       
 Min.   : 1.320   Min.   : 1.100   Min.   : 2.710   Min.   : 1.280  
 1st Qu.: 2.743   1st Qu.: 1.710   1st Qu.: 3.167   1st Qu.: 2.618  
 Median : 3.870   Median : 2.235   Median : 3.430   Median : 3.545  
 Mean   : 5.215   Mean   : 2.625   Mean   : 3.824   Mean   : 4.630  
 3rd Qu.: 5.750   3rd Qu.: 2.908   3rd Qu.: 4.032   3rd Qu.: 5.282  
 Max.   :31.370   Max.   :10.390   Max.   :10.410   Max.   :24.600  
                                                                    
   B365C.2.5      B365C.2.5.1       PC.2.5          PC.2.5.1          MaxC.2.5    
 Min.   :1.220   Min.   :1.30   Min.   :1.220   Min.   :1.320   Min.   :1.260  
 1st Qu.:1.720   1st Qu.:1.53   1st Qu.:1.795   1st Qu.:1.570   1st Qu.:1.857  
 Median :2.100   Median :1.72   Median :2.125   Median :1.790   Median :2.190  
 Mean   :2.142   Mean   :1.87   Mean   :2.188   Mean   :1.922   Mean   :2.241  
 3rd Qu.:2.500   3rd Qu.:2.10   3rd Qu.:2.530   3rd Qu.:2.110   3rd Qu.:2.560  
 Max.   :3.500   Max.   :4.33   Max.   :3.720   Max.   :4.520   Max.   :3.720  
                                                                              
   MaxC.2.5.1       AvgC.2.5       AvgC.2.5.1         AHCh        
 Min.   :1.330   Min.   :1.220   Min.   :1.290   Min.   :-2.7500  
 1st Qu.:1.627   1st Qu.:1.760   1st Qu.:1.550   1st Qu.:-0.7500  
 Median :1.840   Median :2.080   Median :1.750   Median :-0.2500  
 Mean   :1.997   Mean   :2.118   Mean   :1.878   Mean   :-0.3329  
 3rd Qu.:2.223   3rd Qu.:2.440   3rd Qu.:2.062   3rd Qu.: 0.0000  
 Max.   :4.610   Max.   :3.520   Max.   :4.130   Max.   : 1.7500  
                                                                  
    B365CAHH        B365CAHA         PCAHH           PCAHA      
 Min.   :1.700   Min.   :1.670   Min.   :1.750   Min.   :1.750  
 1st Qu.:1.880   1st Qu.:1.880   1st Qu.:1.880   1st Qu.:1.880  
 Median :1.960   Median :1.960   Median :1.960   Median :1.960  
 Mean   :1.959   Mean   :1.956   Mean   :1.962   Mean   :1.959  
 3rd Qu.:2.040   3rd Qu.:2.040   3rd Qu.:2.040   3rd Qu.:2.040  
 Max.   :2.160   Max.   :2.160   Max.   :2.200   Max.   :2.210  
                                                                
    MaxCAHH         MaxCAHA          AvgCAHH         AvgCAHA     
 Min.   :1.780   Min.   :1.800   Min.   :1.72   Min.   :1.750  
 1st Qu.:1.940   1st Qu.:1.930   1st Qu.:1.86   1st Qu.:1.860  
 Median :2.020   Median :2.020   Median :1.93   Median :1.940  
 Mean   :2.021   Mean   :2.016   Mean   :1.94   Mean   :1.938  
 3rd Qu.:2.100   3rd Qu.:2.100   3rd Qu.:2.02   3rd Qu.:2.010  
 Max.   :2.260   Max.   :2.270   Max.   :2.15   Max.   :2.180  
```

```r
#Ahora seleccionaremos únicamente las columnas Date, HomeTeam, AwayTeam,
# FTHG, FTAG y FTR en cada uno de los data frames. Primero guardaremos los
#data frames en una lista con nombre lista y después con ayuda de las funciones
#lapply y select (del paquete dplyr), seleccionaremos las columnas deseadas.
#Los nuevos data frames quedarán guardados en nlista.
lista <- list(d1718, d1819, d1920)
nlista <- lapply(lista, select, Date, HomeTeam, AwayTeam, FTHG, FTAG, FTR)
```

```r
#Con las funciones lapply y str observaremos la estrucura de nuestros nuevos data frames
lapply(nlista, str)
```

```
'data.frame':   380 obs. of  6 variables:
 $ Date    : chr  "18/08/17" "18/08/17" "19/08/17" "19/08/17" ...
 $ HomeTeam: chr  "Leganes" "Valencia" "Celta" "Girona" ...
 $ AwayTeam: chr  "Alaves" "Las Palmas" "Sociedad" "Ath Madrid" ...
 $ FTHG    : int  1 1 2 2 1 0 2 0 1 0 ...
 $ FTAG    : int  0 0 3 2 1 0 0 3 0 1 ...
 $ FTR     : chr  "H" "H" "A" "D" ...
'data.frame':   380 obs. of  6 variables:
 $ Date    : chr  "17/08/2018" "17/08/2018" "18/08/2018" "18/08/2018" ...
 $ HomeTeam: chr  "Betis" "Girona" "Barcelona" "Celta" ...
 $ AwayTeam: chr  "Levante" "Valladolid" "Alaves" "Espanol" ...
 $ FTHG    : int  0 0 3 1 1 1 2 1 2 1 ...
 $ FTAG    : int  3 0 0 1 2 2 0 4 1 1 ...
 $ FTR     : chr  "A" "D" "H" "D" ...
'data.frame':   380 obs. of  6 variables:
 $ Date    : chr  "16/08/2019" "17/08/2019" "17/08/2019" "17/08/2019" ...
 $ HomeTeam: chr  "Ath Bilbao" "Celta" "Valencia" "Mallorca" ...
 $ AwayTeam: chr  "Barcelona" "Real Madrid" "Sociedad" "Eibar" ...
 $ FTHG    : int  1 1 1 2 0 4 1 0 1 1 ...
 $ FTAG    : int  0 3 1 1 1 4 0 2 2 0 ...
 $ FTR     : chr  "H" "A" "D" "H" ...

[[1]]
NULL

[[2]]
NULL

[[3]]
NULL
```

```r
#Arreglamos las columnas Date para que R reconozca los elementos como fechas,
# esto lo hacemos con las funciones mutate (paquete dplyr) y as.Date.
nlista[[1]] <- mutate(nlista[[1]], Date = as.Date(Date, "%d/%m/%y"))
nlista[[2]] <- mutate(nlista[[2]], Date = as.Date(Date, "%d/%m/%Y"))
nlista[[3]] <- mutate(nlista[[3]], Date = as.Date(Date, "%d/%m/%Y"))

#Verificamos que nuestros cambios se hayan llevado a cabo
lapply(nlista, str)
```

```
'data.frame':   380 obs. of  6 variables:
 $ Date    : Date, format: "2017-08-18" "2017-08-18" ...
 $ HomeTeam: chr  "Leganes" "Valencia" "Celta" "Girona" ...
 $ AwayTeam: chr  "Alaves" "Las Palmas" "Sociedad" "Ath Madrid" ...
 $ FTHG    : int  1 1 2 2 1 0 2 0 1 0 ...
 $ FTAG    : int  0 0 3 2 1 0 0 3 0 1 ...
 $ FTR     : chr  "H" "H" "A" "D" ...
'data.frame':   380 obs. of  6 variables:
 $ Date    : Date, format: "2018-08-17" "2018-08-17" ...
```

```
 $ HomeTeam: chr  "Betis" "Girona" "Barcelona" "Celta" ...
 $ AwayTeam: chr  "Levante" "Valladolid" "Alaves" "Espanol" ...
 $ FTHG    : int  0 0 3 1 1 1 2 1 2 1 ...
 $ FTAG    : int  3 0 0 1 2 2 0 4 1 1 ...
 $ FTR     : chr  "A" "D" "H" "D" ...
'data.frame':	380 obs. of  6 variables:
 $ Date    : Date, format: "2019-08-16" "2019-08-17" ...
 $ HomeTeam: chr  "Ath Bilbao" "Celta" "Valencia" "Mallorca" ...
 $ AwayTeam: chr  "Barcelona" "Real Madrid" "Sociedad" "Eibar" ...
 $ FTHG    : int  1 1 1 2 0 4 1 0 1 1 ...
 $ FTAG    : int  0 3 1 1 1 4 0 2 2 0 ...
 $ FTR     : chr  "H" "A" "D" "H" ...

[[1]]
NULL

[[2]]
NULL

[[3]]
NULL
```

```r
#Finalmente, con ayuda de las funciones rbind y do.call
#combinamos los data frames contenidos en nlista como un único data frame
data <- do.call(rbind, nlista)
dim(data)
```

```
[1] 1140    6
```

```r
str(data)
```

```
'data.frame':	1140 obs. of  6 variables:
 $ Date    : Date, format: "2017-08-18" "2017-08-18" ...
 $ HomeTeam: chr  "Leganes" "Valencia" "Celta" "Girona" ...
 $ AwayTeam: chr  "Alaves" "Las Palmas" "Sociedad" "Ath Madrid" ...
 $ FTHG    : int  1 1 2 2 1 0 2 0 1 0 ...
 $ FTAG    : int  0 0 3 2 1 0 0 3 0 1 ...
 $ FTR     : chr  "H" "H" "A" "D" ...
```

```r
tail(data)
```

```
           Date HomeTeam    AwayTeam FTHG FTAG FTR
1135 2020-07-19  Espanol       Celta    0    0   D
1136 2020-07-19  Granada  Ath Bilbao    4    0   H
1137 2020-07-19  Leganes Real Madrid    2    2   D
1138 2020-07-19  Levante      Getafe    1    0   H
1139 2020-07-19  Osasuna    Mallorca    2    2   D
1140 2020-07-19  Sevilla    Valencia    1    0   H
```

```r
View(data)
summary(data)
```

```
      Date                HomeTeam           AwayTeam              FTHG      
 Min.   :2017-08-18   Length:1140        Length:1140        Min.   :0.000  
 1st Qu.:2018-03-17   Class :character   Class :character   1st Qu.:1.000  
 Median :2019-01-16   Mode  :character   Mode  :character   Median :1.000  
 Mean   :2019-01-15                                         Mean   :1.479  
 3rd Qu.:2019-10-27                                         3rd Qu.:2.000  
 Max.   :2020-07-19                                         Max.   :8.000  
      FTAG           FTR           
 Min.   :0.000   Length:1140       
 1st Qu.:0.000   Class :character  
 Median :1.000   Mode  :character  
 Mean   :1.108                     
 3rd Qu.:2.000                     
 Max.   :6.000                     
```

```r
#Con ayuda de la función table obtenemos las estimaciones
# de probabilidades solicitadas
(pcasa <- round(table(data$FTHG)/dim(data)[1], 3)) # Probabilidades marginales
```

```
    0     1     2     3     4     5     6     7     8 
0.232 0.327 0.267 0.112 0.035 0.019 0.005 0.001 0.001 
```

```r
#estimadas para los equipos que juegan en casa

(pvisita <- round(table(data$FTAG)/dim(data)[1], 3)) # Probabilidades marginales
```

```
    0     1     2     3     4     5     6 
0.352 0.340 0.212 0.054 0.029 0.010 0.003 
```

```r
#estimadas para los equipos que juegan como visitante

(pcta <- round(table(data$FTHG, data$FTAG)/dim(data)[1], 3)) # Probabilidades
```

```
   
        0     1     2     3     4     5     6
  0 0.078 0.081 0.046 0.018 0.005 0.004 0.000
  1 0.116 0.115 0.068 0.018 0.009 0.002 0.000
  2 0.088 0.094 0.061 0.011 0.009 0.002 0.002
  3 0.045 0.032 0.025 0.006 0.002 0.002 0.001
  4 0.014 0.011 0.007 0.000 0.004 0.000 0.000
  5 0.009 0.005 0.004 0.000 0.001 0.000 0.000
  6 0.003 0.002 0.000 0.001 0.000 0.000 0.000
  7 0.000 0.001 0.000 0.000 0.000 0.000 0.000
  8 0.000 0.000 0.001 0.000 0.000 0.000 0.000
```

```r
# conjuntas estimadas para los partidos

#Con la función apply primero dividimos cada elemento de las columnas de la
#matriz de probabilidades conjuntas, por las probabilidades marginales asociadas
#y que corresponden al equipo de casa. Note como hemos definido una función anómima
```

```r
# dentro de apply. Luego dividimos cada elemento de las filas de la matriz que resulta,
# por las probabilidades marginales asociadas y que corresponden al equivo visitante.
#Finalmente hacemos obtenemos la transpuesta de la matriz que resulta. Esta última matriz,
# es la matriz de cocientes buscada, es decir, hemos dividio cada probabilidad conjunta,
# por el producto de probabilidades marginales correspondientes.
(cocientes <- apply(pcta, 2, function(col) col/pcasa))
```

```
          0         1         2          3          4           5           6
  0 0.3362069 0.3491379 0.1982759 0.07758621 0.02155172 0.017241379 0.000000000
  1 0.3547401 0.3516820 0.2079511 0.05504587 0.02752294 0.006116208 0.000000000
  2 0.3295880 0.3520599 0.2284644 0.04119850 0.03370787 0.007490637 0.007490637
  3 0.4017857 0.2857143 0.2232143 0.05357143 0.01785714 0.017857143 0.008928571
  4 0.4000000 0.3142857 0.2000000 0.00000000 0.11428571 0.000000000 0.000000000
  5 0.4736842 0.2631579 0.2105263 0.00000000 0.05263158 0.000000000 0.000000000
  6 0.6000000 0.4000000 0.0000000 0.20000000 0.00000000 0.000000000 0.000000000
  7 0.0000000 1.0000000 0.0000000 0.00000000 0.00000000 0.000000000 0.000000000
  8 0.0000000 0.0000000 1.0000000 0.00000000 0.00000000 0.000000000 0.000000000
```

```r
(cocientes <- apply(cocientes, 1, function(fila) fila/pvisita))
```

```
          0         1         2         3         4         5        6
  0 0.9551332 1.0077843 0.9363296 1.1414367 1.1363636 1.3456938 1.704545
  1 1.0268763 1.0343587 1.0354704 0.8403361 0.9243697 0.7739938 1.176471
  2 0.9352635 0.9809013 1.0776624 1.0528976 0.9433962 0.9930487 0.000000
  3 1.4367816 1.0193680 0.7629352 0.9920635 0.0000000 0.0000000 3.703704
  4 0.7431629 0.9490668 1.1623402 0.6157635 3.9408867 1.8148820 0.000000
  5 1.7241379 0.6116208 0.7490637 1.7857143 0.0000000 0.0000000 0.000000
  6 0.0000000 0.0000000 2.4968789 2.9761905 0.0000000 0.0000000 0.000000

         7        8
  0 0.000000 0.000000
  1 2.941176 0.000000
  2 0.000000 4.716981
  3 0.000000 0.000000
  4 0.000000 0.000000
  5 0.000000 0.000000
  6 0.000000 0.000000
```

```r
(cocientes <- t(cocientes))
```

```
          0         1         2         3         4         5        6
  0 0.9551332 1.0268763 0.9352635 1.4367816 0.7431629 1.7241379 0.000000
  1 1.0077843 1.0343587 0.9809013 1.0193680 0.9490668 0.6116208 0.000000
  2 0.9363296 1.0354704 1.0776624 0.7629352 1.1623402 0.7490637 2.496879
  3 1.1414367 0.8403361 1.0528976 0.9920635 0.6157635 1.7857143 2.976190
  4 1.1363636 0.9243697 0.9433962 0.0000000 3.9408867 0.0000000 0.000000
  5 1.3456938 0.7739938 0.9930487 0.0000000 1.8148820 0.0000000 0.000000
  6 1.7045455 1.1764706 0.0000000 3.7037037 0.0000000 0.0000000 0.000000
  7 0.0000000 2.9411765 0.0000000 0.0000000 0.0000000 0.0000000 0.000000
  8 0.0000000 0.0000000 4.7169811 0.0000000 0.0000000 0.0000000 0.000000
```

```r
#Lo anterior igual lo pudimos lograr de la siguiente manera:

pcta/outer(pcasa, pvisita, "*")
```

```
            0         1         2         3         4         5         6
  0 0.9551332 1.0268763 0.9352635 1.4367816 0.7431629 1.7241379 0.0000000
  1 1.0077843 1.0343587 0.9809013 1.0193680 0.9490668 0.6116208 0.0000000
  2 0.9363296 1.0354704 1.0776624 0.7629352 1.1623402 0.7490637 2.4968789
  3 1.1414367 0.8403361 1.0528976 0.9920635 0.6157635 1.7857143 2.9761905
  4 1.1363636 0.9243697 0.9433962 0.0000000 3.9408867 0.0000000 0.0000000
  5 1.3456938 0.7739938 0.9930487 0.0000000 1.8148820 0.0000000 0.0000000
  6 1.7045455 1.1764706 0.0000000 3.7037037 0.0000000 0.0000000 0.0000000
  7 0.0000000 2.9411765 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
  8 0.0000000 0.0000000 4.7169811 0.0000000 0.0000000 0.0000000 0.0000000
```

```r
#Primero extraemos de manera aleatoria algunas filas de nuestro data
# frame data, esto lo hacemos con ayuda de la función sample.
set.seed(2)
indices <- sample(dim(data)[1], size = 380, replace = TRUE)
newdata <- data[indices, ]

#Con ayuda de la función table obtenemos las estimaciones de probabilidades

(pcasa <- round(table(newdata$FTHG)/dim(newdata)[1], 3)) # Probabilidades marginales
```

```
    0     1     2     3     4     5     6
0.261 0.311 0.274 0.084 0.045 0.018 0.008
```

```r
# estimadas para los equipos que juegan en casa

(pvisita <- round(table(newdata$FTAG)/dim(newdata)[1], 3)) # Probabilidades
```

```
    0     1     2     3     4     5     6
0.361 0.400 0.187 0.034 0.008 0.008 0.003
```

```r
#marginales estimadas para los equipos que juegan como visitante

(pcta <- round(table(newdata$FTHG, newdata$FTAG)/dim(newdata)[1], 3)) # Probabilidades
```

```
        0     1     2     3     4     5     6
  0 0.103 0.100 0.039 0.011 0.000 0.008 0.000
  1 0.113 0.139 0.045 0.008 0.005 0.000 0.000
  2 0.082 0.103 0.074 0.011 0.003 0.000 0.003
  3 0.029 0.034 0.016 0.005 0.000 0.000 0.000
  4 0.018 0.018 0.008 0.000 0.000 0.000 0.000
  5 0.008 0.005 0.005 0.000 0.000 0.000 0.000
  6 0.008 0.000 0.000 0.000 0.000 0.000 0.000
```

```r
#conjuntas estimadas para los partidos

#Obtenemos nuevamente los cocientes de probabilidades conjuntas entre probabilidades marginales

(cocientes <- pcta/outer(pcasa, pvisita, "*"))
```

```
          0         1         2         3         4         5         6
  0 1.0931746 0.9578544 0.7990657 1.2395763 0.0000000 3.8314176 0.0000000
  1 1.0064932 1.1173633 0.7737676 0.7565727 2.0096463 0.0000000 0.0000000
  2 0.8290030 0.9397810 1.4442406 1.1807643 1.3686131 0.0000000 3.6496350
  3 0.9563382 1.0119048 1.0185893 1.7507003 0.0000000 0.0000000 0.0000000
  4 1.1080332 1.0000000 0.9506833 0.0000000 0.0000000 0.0000000 0.0000000
  5 1.2311480 0.6944444 1.4854427 0.0000000 0.0000000 0.0000000 0.0000000
  6 2.7700831 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
```

```r
#Repita el remuestreo anterior varias veces (unas 1000 veces)
# y obtenga una idea de las distribuciones de los cocientes.
#Finalmente mencione en cuales casos le parece razonable la suposición de que
#el cociente es igual a 1.
```