Distribución normal

Ana Cecilia Plavan, Andrés Andruskevicius

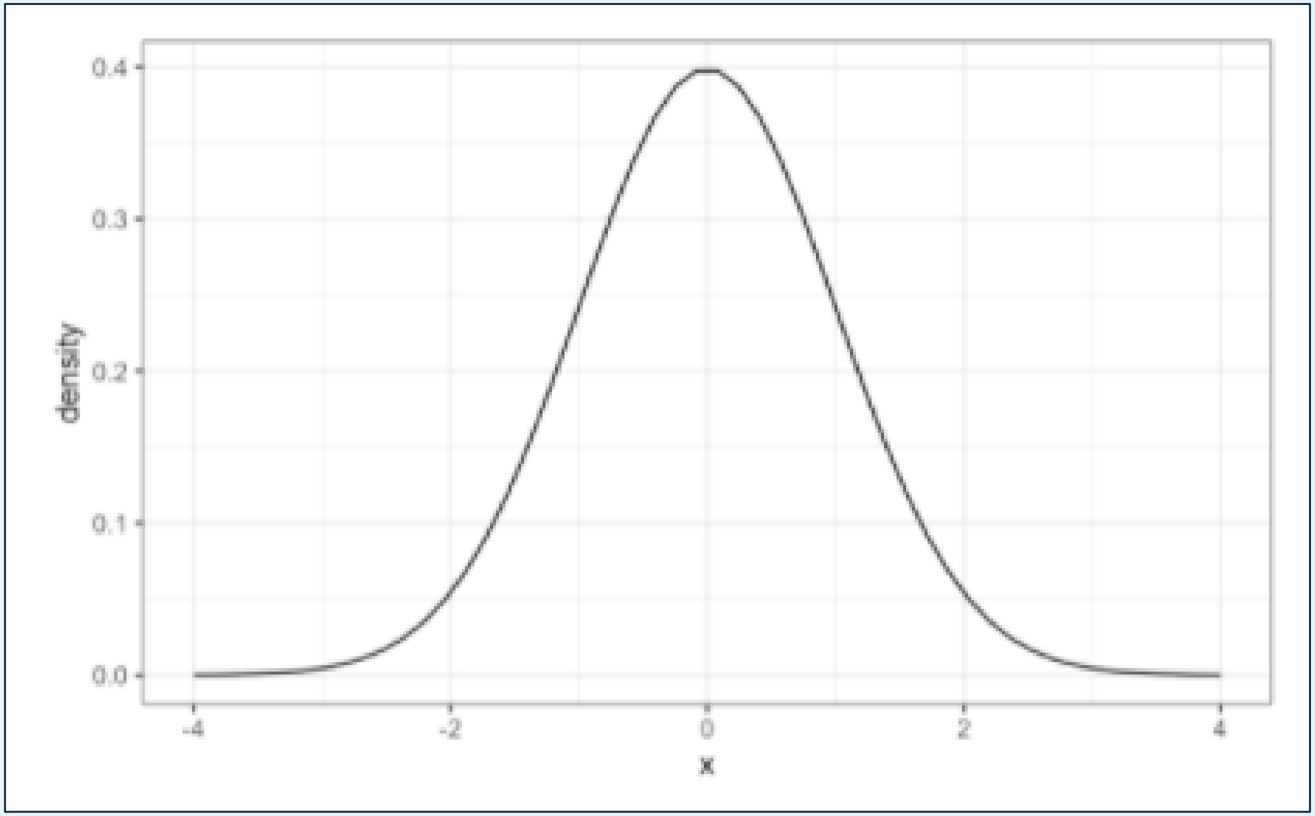
Fórmula de distribución normal

- Para un intervalo a - b la proporción de valores en ese intervalo se calcula:

$$\Pr(a < x < b) = \int_a^b \frac{1}{\sqrt{2\pi s}} e^{-\frac{1}{2}\left(\frac{x-m}{s}\right)^2} dx$$

- Siendo m es el promedio-mean, y s es la desviación estándar-standard deviation.

Resumen de tablas de datos en Distribución normal



Promedio ± desviación estándar

(m = 0 ± s = 1)

Valores importantes para una lista arbitraria de números

- Para una lista de números de un vector x
  - Promedio (mean) - m :

```
m <- sum(x) / length(x) → mean(x)
```

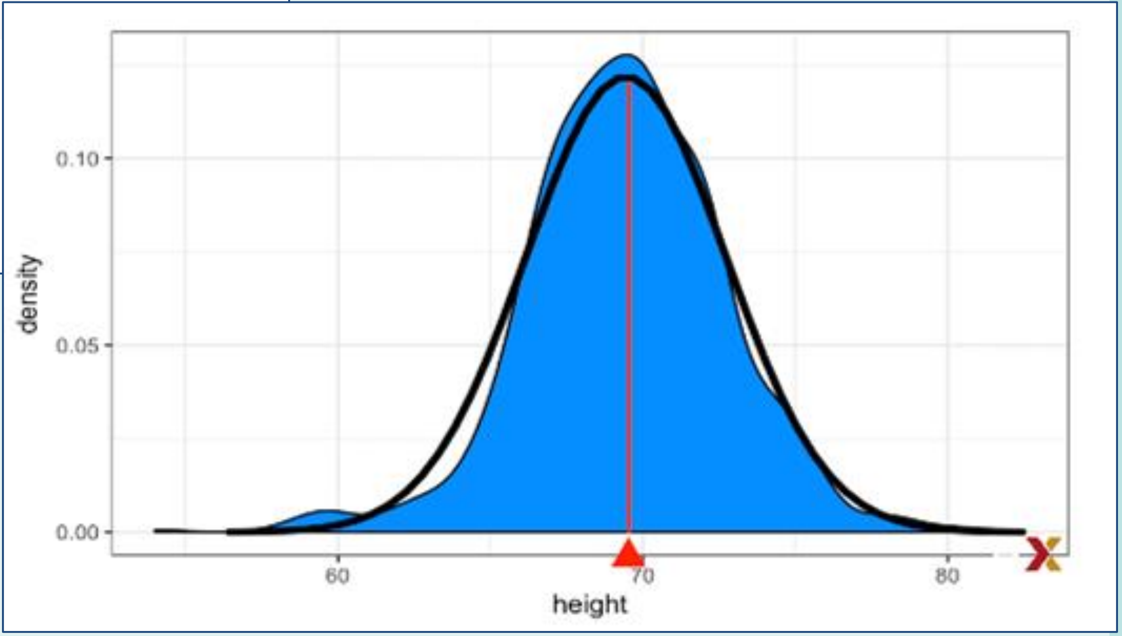
  - Desviación estándar (sd) - s :

```
s <- sqrt(sum((x-mu)^2) / length(x)) → sd(x)
```

Se interpreta como la distancia promedio entre los valores y su promedio

Para nuestros datos de alturas

```
index <- heights$sex == "Male"

x <- heights$height[index]

m <- mean(x)

s <- sd(x)

c(average = m, sd = s)

#> average       sd

#>   69.31     3.61
```



Unidades estándar

La unidad estándar de un valor nos dice cuán lejos, en desviaciones estándar está del promedio.

Para un valor x:

```
z = (x - m)/s → scale(x)
```

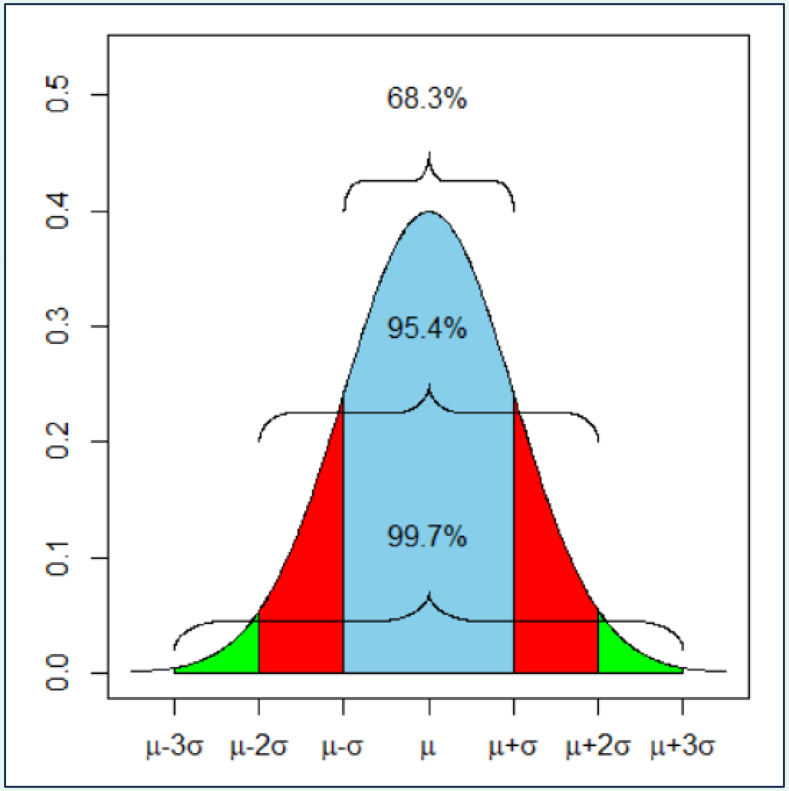
Siendo s y m la desviación estándar y el promedio respectivamente.

En alturas

- Persona de altura promedio z=0
- Uno de los más altos z≈2
- Uno de los más pequeños z≈−2
- Una altura muy rara z>3 o z<-3

Distribución normal estándar

- Distribución con m = μ y s = σ



- Se describe con la regla de 68-95-99.7

Distribución normal

Decimos que los datos son normalmente distribuidos con promedio m y desviación estándar s si la distribución de probabilidades de define por:

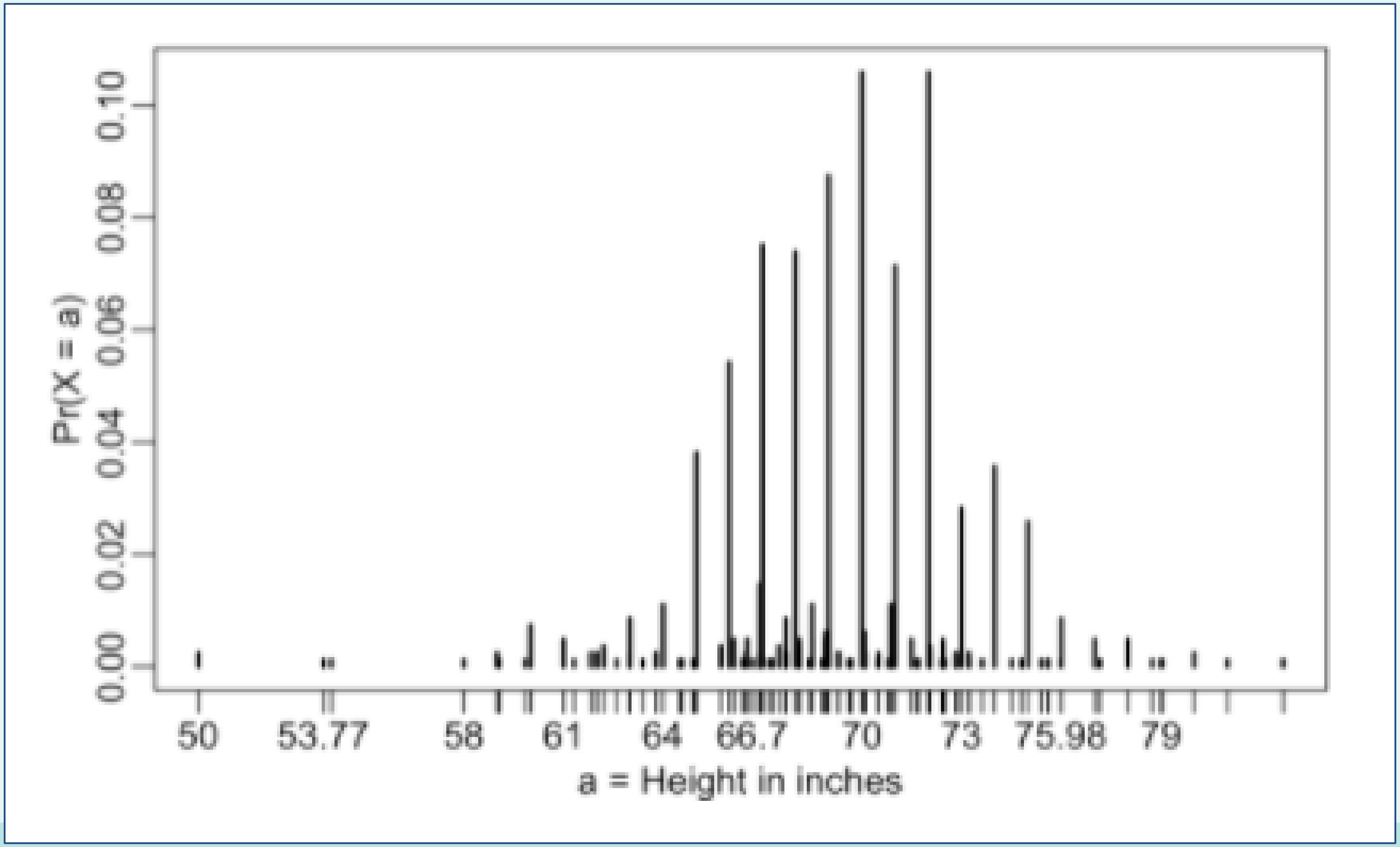
```
F(a) = pnorm(a, m, s)
```

Estimamos probabilidad de que un hombre sea más alto que 70.5inches

```
library(tidyverse)

library(dslabs)

data(heights)

x <- heights %>% filter(sex=="Male") %>% pull(height)

1 - pnorm(70.5, mean(x), sd(x))
```

Discretización y aproximación normal

Si para nuestros datos de altura consideramos cada altura como una categoría única y graficamos la proporción de estudiantes que tiene cada altura.



Discretización de datos

Rango alturas x1<= X <=x2	Proporción de estudiantes mean(x<=x2)-mean(x<=x1)	Aprox. a distribución normal pnorm(x2,m,s)-pnorm(x1,m,s)
67.5≦ X ≦68.5	*0.115*	*0.103*
68.5≦ X ≦69.5	*0.119*	*0.11*
69.5≦ X ≦70.5	*0.122*	*0.108*
70.1≦ X ≦70.9	*0.0222*	*0.0836*