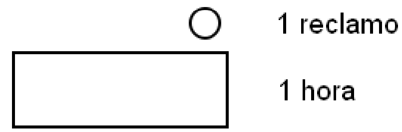
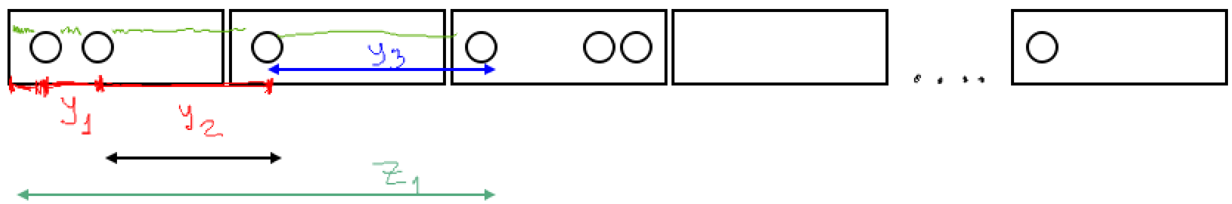
Distribución Poisson, Exponencial, Gamma





X = N° de reclamos por hora $\quad x_1 = 2,\ x_2 = 1,\ x_3 = 3,\ x_4 = 0, \ldots, x_n = 1 \qquad X \sim Pois(\lambda)$

EXP.: Y = Tiempo que transcurre entre 2 reclamos consecutivos, $Y \sim Exp(\beta)$

GAMMA: Z = Tiempo que transcurre entre 4 reclamos consecutivos, $Z \sim Gamma(\alpha = 4, \beta)$

Si $\alpha = 1 \Rightarrow Z \sim Gamma(\alpha = 1, \beta) \equiv Exp(\beta)$

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right), & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Si $\alpha = 1$, $f(x) = \dfrac{1}{\cancel{\Gamma(1)}^{\,1}\beta^1} \cancel{x^{1-1}}^{\,1} e^{-x/\beta} = \dfrac{1}{\beta} e^{-\frac{x}{\beta}}$

$\Gamma(\alpha) = (\alpha-1)!$

$\Gamma(1) = (1-1)! = 0! = 1$

Parametrización 1

$$f(x) = \frac{1}{\Gamma(\alpha)\,\beta^{\alpha}}\, x^{\alpha-1} e^{-x/\beta}$$

$$\mu_x = \alpha\beta$$

$$\sigma_x^2 = \alpha\beta^2$$

Parametrización 2

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\, x^{\alpha-1} e^{-x\beta}$$

$$\mu_x = \alpha/\beta$$

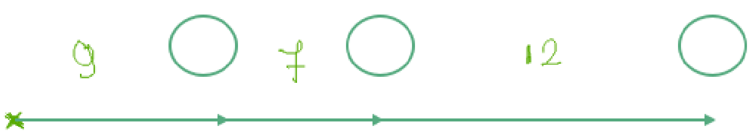
$$\sigma_x^2 = \alpha/\beta^2$$

Aplicación

En una central de monitoreo de emergencias, se sabe que, en promedio, ocurre un evento cada 10 minutos. Queremos modelar el tiempo hasta que ocurran 3 eventos consecutivos.

Variable aleatoria X =Tiempo que transcurre hasta 3 eventos

$$X \sim Gamma(\alpha = 3, \beta = 10)$$



dgamma $\longrightarrow$ Densidad (altura de f(x))

pgamma $\longrightarrow$ Probabilidad acumulada

qgamma $\longrightarrow$ Cuantil / Percentil

rgamma $\longrightarrow$ Muestra aleatoria

Probabilidad de que ocurran los 3 eventos antes de 25 minutos:

$$P(X < 25) = 0.4562$$

Probabilidad de que ocurran los 3 eventos hasta los 25 minutos:

$$P(X \leqslant 25) = 0.4562$$

Parametrización 1

$$f(x) = \frac{1}{\Gamma(\alpha)\,\beta^{\alpha}}\, x^{\alpha-1} e^{-x/\beta}$$

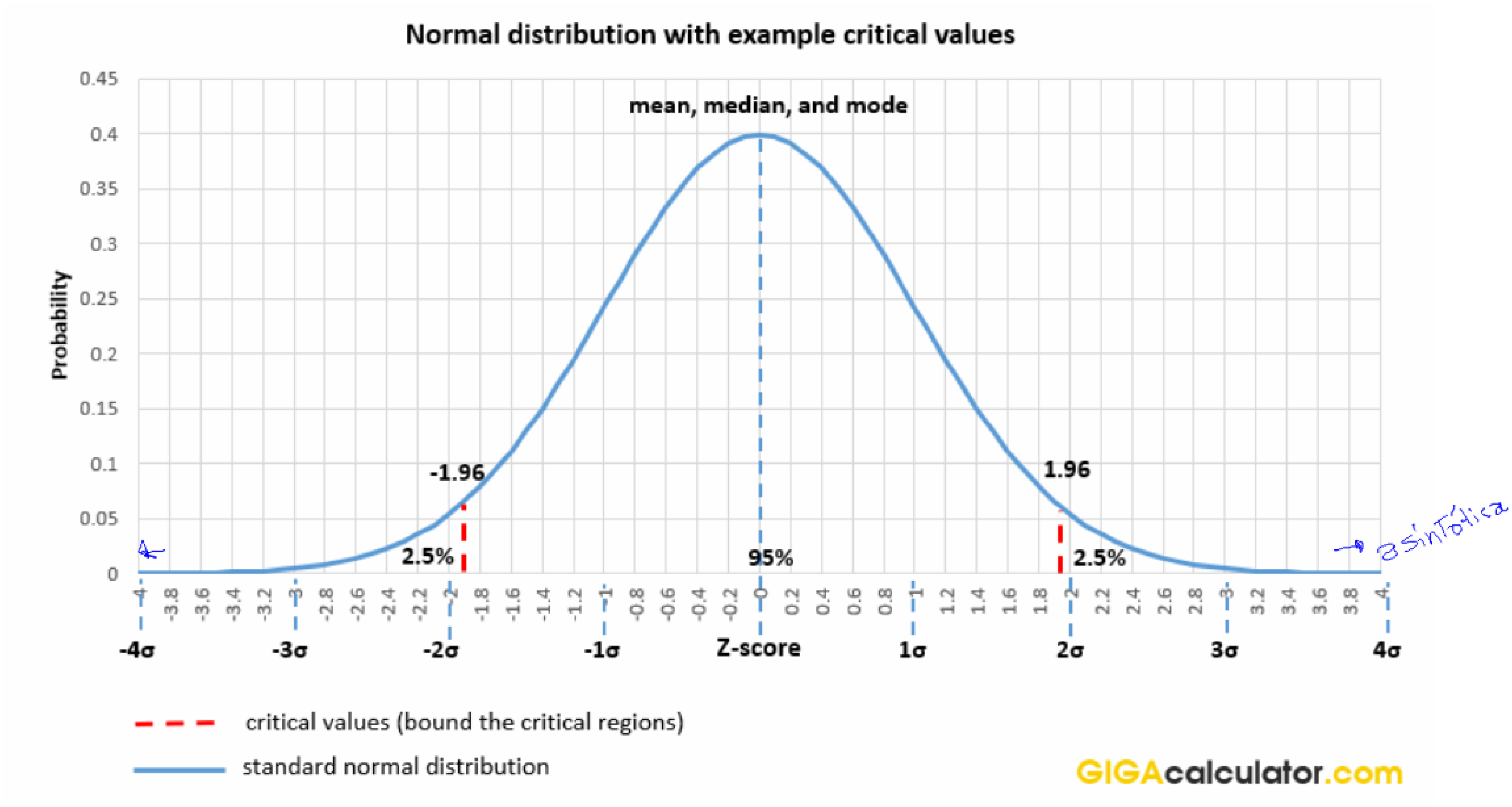
```
> pgamma(q = 25, shape = 3, scale = 10)
[1] 0.4561869
```

Parametrización 2

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\, x^{\alpha-1} e^{-x\beta}$$

```
> pgamma(q = 25, shape = 3, rate = 1/10)
[1] 0.4561869
```

Distribución Normal



$X \sim \text{Unif}(2,9)$, $P(4<X<6) = \int_4^6 \frac{1}{7}\,dx$

$X \sim \text{Exp}(\lambda=4)$, $P(X>10) = \int_{10}^{\infty} 4e^{-4x}\,dx$

$X \sim \text{Gamma}(\alpha=5, \beta=2)$, $P(X<8) = \int_0^8 \frac{1}{\Gamma(5)\,2^5} x^4 e^{-2x}\,dx$

Se pueden resolver analíticamente

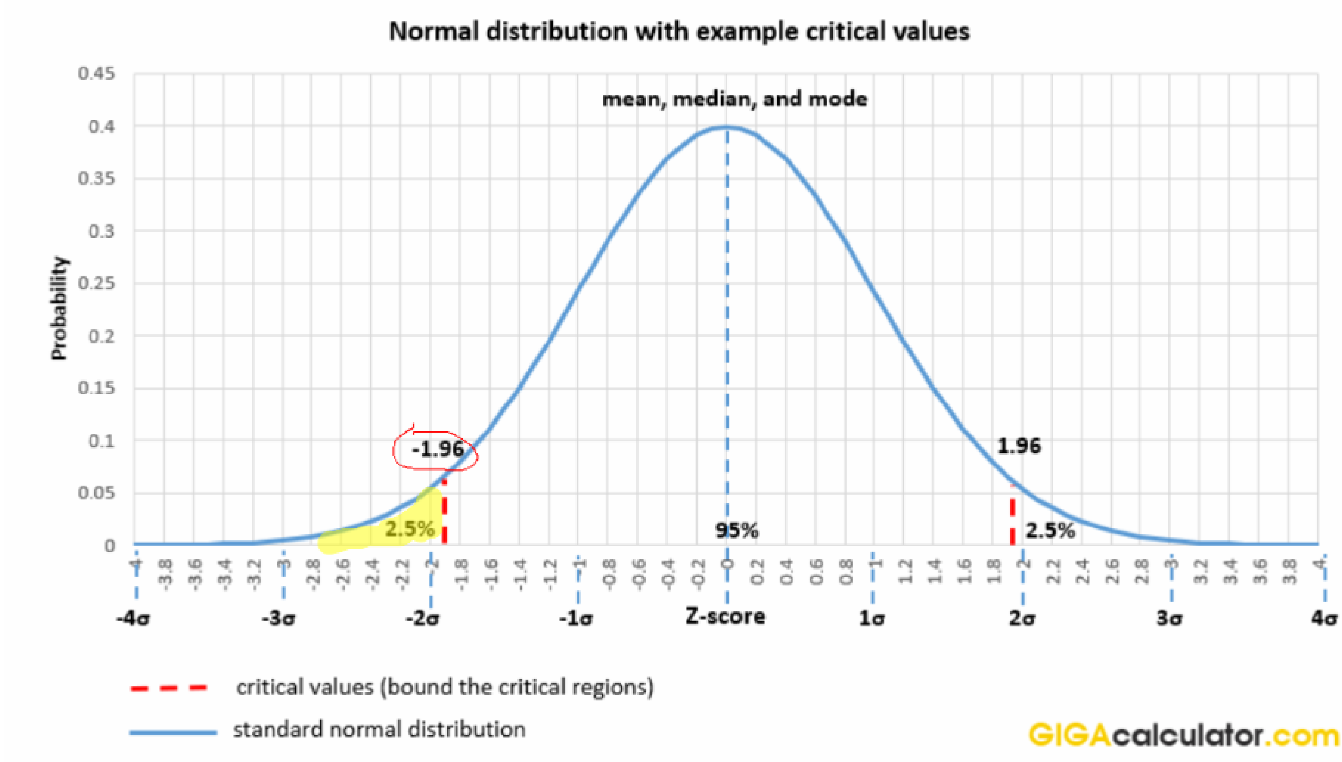
$X \sim N(\mu=20, \sigma^2=4)$, $P(X>21) = \int_{21}^{\infty} \frac{1}{\sqrt{2\pi} \times 2} e^{-\frac{1}{2}\left(\frac{x-20}{2}\right)^2}\,dx$

No tiene solución cerrada

R, Python, (Tablas)

$X \sim N(\mu=7, \sigma^2=2) \xrightarrow[\text{estandarizo}]{} Z \sim N(\mu=0, \sigma^2=1)$

$X \sim Ga(\alpha=6, \beta=12) \xrightarrow[\text{estandarizo}]{} D \sim ?$



$Z \sim N(\mu = 0, \sigma^2 = 1)$

$Z_{0.025}$ = Percentil 2.5

```
> qnorm(p = 0.025, mean = 0, sd = 1)
[1] -1.959964
> qnorm(p = 0.025)
[1] -1.959964
```

$Z_{0.05}$ = Percentil 5

```
> qnorm(p = 0.05, mean = 0, sd = 1)
[1] -1.644854
> qnorm(p = 0.05)
[1] -1.644854
```

$Z_{0.50}$ = Mediana = 0

$X \sim N(\mu = 10, \sigma^2 = 1)$

$10 - 4 = 6$

$10 + 4 = 14$

$Z_{0.025} = -1.96$

$Z_{0.975} = 1.96$

$Z_{0.05} = -1.65$

$Z_{0.95} = 1.65$

$Z_{\alpha} = -Z_{1-\alpha}$

Aproximación de la Binomial a la Normal

$$X \sim Bin(n=80, \pi=0.3) \Rightarrow X \approx N(\mu = 24, \sigma^2 = 16.8)$$

$$\left.\begin{array}{l} n\pi = 80 \times 0.3 = 24 \\ n(1-\pi) = 80 \times 0.7 = 56 \end{array}\right\} \checkmark$$

$$n\pi(1-\pi) = 80 \times 0.3 \times 0.7 = 16.8$$

$$P(X=25) = P(24.5 \leq X \leq 25.5)$$

$\downarrow$ Binomial $\qquad\qquad$ $\downarrow$ Normal

```
> dbinom(x = 25, size = 80, prob = 0.3)
[1] 0.09307325
```

```
> pnorm(q = 25.5, mean = 24, sd = sqrt(16.8)) - pnorm(q = 24.5, mean = 24, sd = sqrt(16.8))
[1] 0.09425794
```

Aproximación de la Poisson a la Normal

¿Qué pasaría si aproximamos una Poisson con lambda pequeño a una Normal?

$X \sim Pois(\lambda = 5)$

$p(X=3) = 0.1404$

```
> dpois(x = 3, lambda = 5)
[1] 0.1403739
```

```
> rpois(n = 10, lambda = 5)
 [1] 3 2 4 7 7 6 6 8 2 6
```

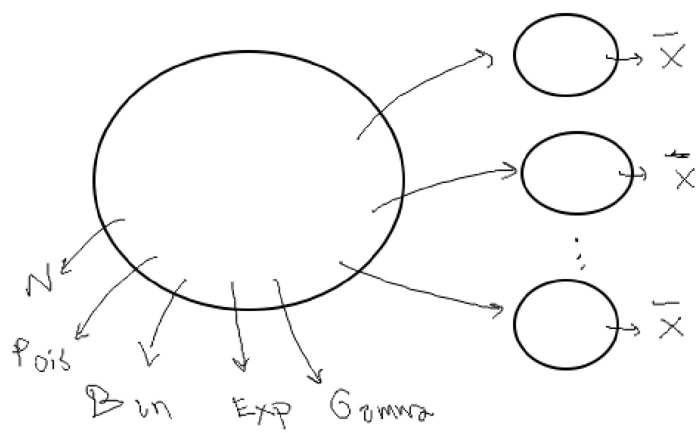
$X \approx N(\mu = 5, \sigma^2 = 5)$

$P(X=3) = P(2.5 < X < 3.5)$

```
> pnorm(q = 3.5, mean = 5, sd = sqrt(5)) - pnorm(q = 2.5, mean = 5, sd = sqrt(5))
[1] 0.1193912
```

```
> rnorm(n = 10, mean = 5, sd = sqrt(5)) |> round(0)
 [1]  1  4  3  2 -1  5  2  5  5  5
```

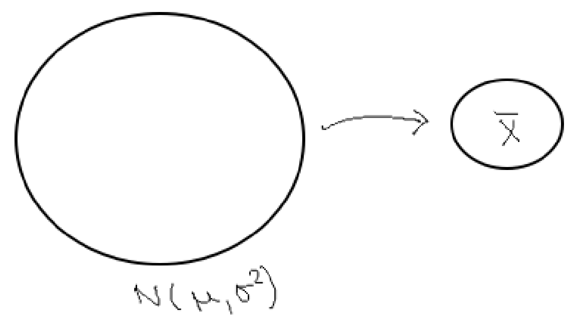
Distribución Muestral de la Media



$\bar{X}$ es una v.aleatoria

Caso 1:

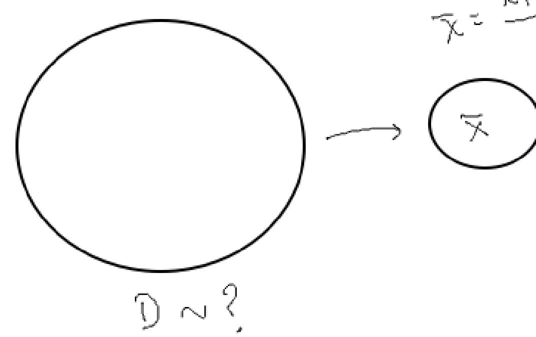
$$\bar{X} = \frac{x_1 + \ldots + x_n}{n}$$



$$\bar{X} \sim N(\mu, \sigma^2/n)$$

Caso 2:

$$\bar{x} = \frac{x_1 + \ldots + x_n}{n}, \quad \boxed{n \geq 30}$$



$$\bar{X} \sim N(\mu, \sigma^2/n)$$

Una clínica registra el tiempo de atención médica de sus pacientes como una variable que sigue una distribución normal con media 18 minutos y desviación estándar 6 minutos. Se toma una muestra aleatoria de n=9 pacientes. ¿Cuál es la probabilidad de que el tiempo promedio de atención sea mayor a 20 minutos?

X = Tiempo de atención

$X \sim N(\mu = 18, \sigma^2 = 6^2)$

$n = 9$

$P(\bar{X} > 20) = ?$

$P(X > 20)$: para un paciente

$P(\bar{X} > 20)$: para la muestra de pacientes

distribución muestral de la media

Por el T.L.C. : $\bar{X} \sim N(\mu = 18, \sigma^2 = 6^2/9) \sim N(\mu = 18, \sigma^2 = 4)$

$P(\bar{X} > 20) = 0.159$

```
> 1 - pnorm(q = 20, mean = 18, sd = 2)
[1] 0.1586553
> pnorm(q = 20, mean = 18, sd = 2, lower.tail = F)
[1] 0.1586553
```

En una ciudad, se estudian los ingresos diarios de trabajadores independientes. Se sabe que los ingresos son altamente asimétricos a la derecha, con algunos trabajadores que ganan mucho más que el promedio. Estudios previos han estimado que la media poblacional es de 120 soles y la desviación estándar es de 90 soles.Un investigador toma una muestra aleatoria de n = 40 trabajadores, ¿cuál es la probabilidad de que el ingreso promedio de la muestra esté entre 100 y 140 soles?

X = Ingreso (no es Normal)

$E(X) = \mu_X = 120$

$\sigma_X = 90$

$n = 40 \quad (n > 30)$

$$\bar{X} \sim N\left(\mu = 120,\ \sigma^2 = \underbrace{90^2/40}_{202.5}\right) \sim N(\mu = 120,\ \sigma^2 = 202.5)$$

dist. muestral de la media

$$P(100 < \bar{X} < 140) = P(\bar{X} < 140) - P(\bar{X} < 100)$$
$$= 0.92 - 0.08$$
$$= 0.84$$

```
> pnorm(q = 140, mean = 120, sd = sqrt(202.5))
[1] 0.9200572

> pnorm(q = 100, mean = 120, sd = sqrt(202.5))
[1] 0.07994275
```

Ejercicio Una empresa fabrica un tipo de artículo cuya producción está sujeta a pequeñas variaciones de peso. Según estudios de control de calidad, el peso (en kg) de los artículos sigue la siguiente distribución:

$$f(x) = \begin{cases} \frac{2}{3}x, & 1 \leq x \leq 2 \\ 0, & \text{para otros valores de } x \end{cases}$$

Se toma una muestra aleatoria de 75 artículos, y se quiere conocer la probabilidad de que el peso promedio esté por debajo de 1.6 kg

X = Peso (no es Normal)

$\mu_X = 1.556$

```
> Ex= function(x){
+   x*2/3*x
+ }
> integrate(Ex,1,2)$value
[1] 1.555556
```

$\sigma_X^2 = 2.5 - 1.556^2 = 0.079$

```
> Ex2= function(x){
+   x^2*2/3*x
+ }
> integrate(Ex2,1,2)$value
[1] 2.5
```

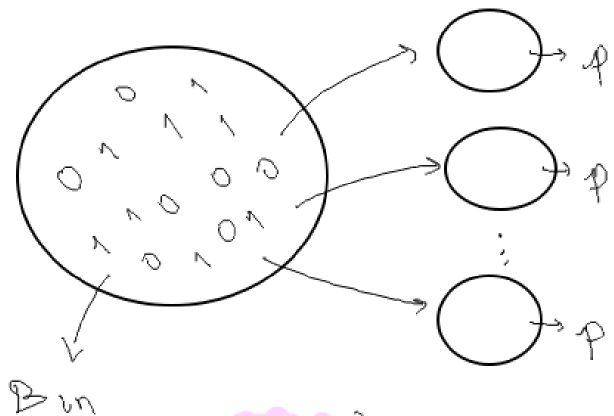
$n = 75$ ($n > 30$)

$P(\bar{X} < 1.6)$

$\bar{X} \sim N(\mu = 1.556, \sigma^2 = \underbrace{0.079/75}_{0.00105})$

```
> pnorm(q = 1.6, mean = 1.556, sd = sqrt(0.00105))
[1] 0.9127474
```

Distribución Muestral de la Proporción



$(n \geq 50)$

$$p \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

π : proporción conocida, previa, histórica

Una encuesta nacional indica que aproximadamente el 62% de los adultos leen al menos un libro al mes. Un investigador realiza una encuesta a $n = 150$ adultos seleccionados al azar en una región del país. ¿Cuál es la probabilidad de que menos del 58% de los encuestados en esta muestra lean al menos un libro al mes?

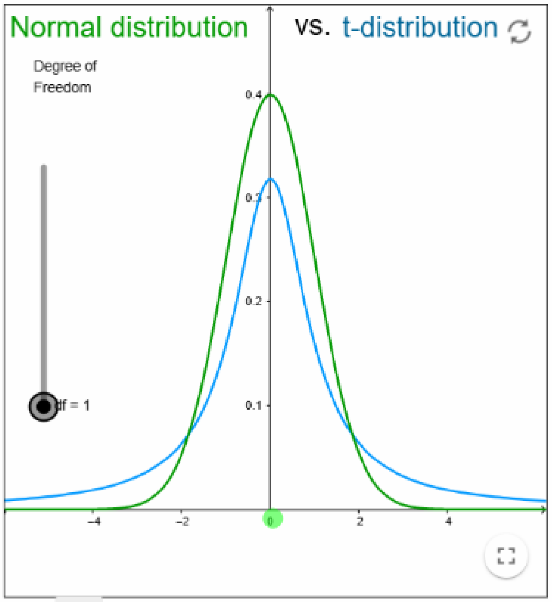
π

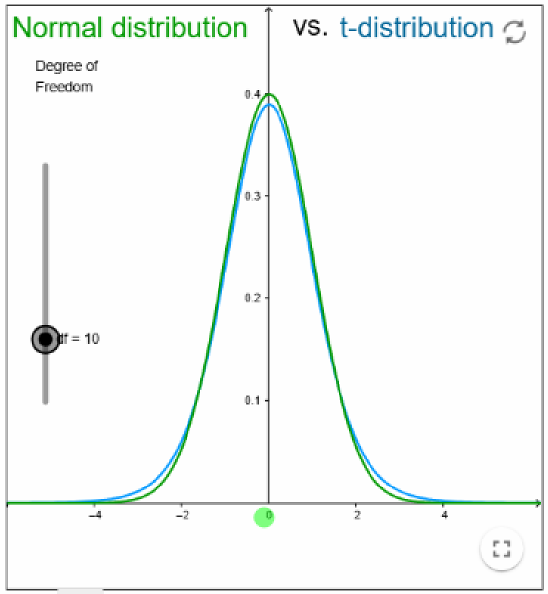
$$P(p < 0.58) = 0.156$$

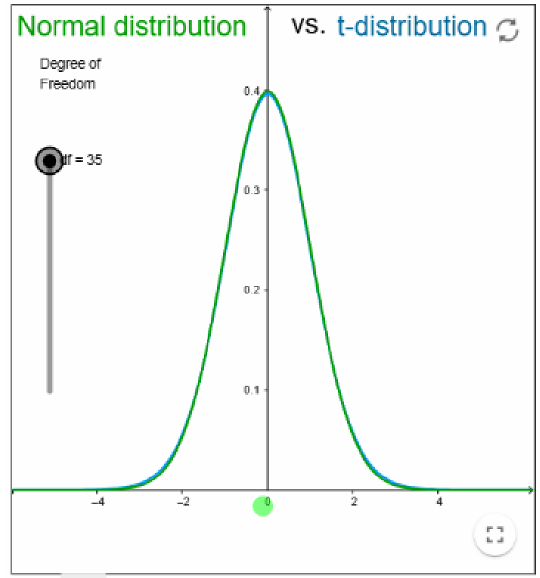
$$p \sim N\left(\mu = 0.62, \ \sigma^2 = \underbrace{\frac{0.62 \times 0.38}{150}}_{0.00157}\right)$$

```
> (sigma2 <- 0.62*0.38/150)
[1] 0.001570667
> pnorm(q = 0.58, mean = 0.62, sd = sqrt(sigma2))
[1] 0.1564167
```

Distribución t de Student







media poblacional

$$\mu_X = E(X) = 0 \qquad \text{para } \nu > 1$$ ✓

```
> rt(n = 600, df = 1) |> mean()
[1] 0.6251954
> rt(n = 600, df = 1) |> mean()
[1] 0.6248533
> rt(n = 600, df = 1) |> mean()
[1] -241.2673
> rt(n = 600, df = 1) |> mean()
[1] -0.4439622
> rt(n = 600, df = 1) |> mean()
[1] 0.4811869
> rt(n = 600, df = 1) |> mean()
[1] -1.614717
```

```
> rt(n = 600, df = 2) |> mean()
[1] -0.03575703 ≈ 0
> rt(n = 600, df = 2) |> mean()
[1] 0.1333855 ≈ 0
> rt(n = 600, df = 2) |> mean()
[1] 0.1025686 ≈ 0
> rt(n = 600, df = 2) |> mean()
[1] 0.04014817 ≈ 0
> rt(n = 600, df = 2) |> mean()
[1] 0.03692146 ≈ 0
> rt(n = 600, df = 2) |> mean()
[1] 0.006413797 ≈ 0
```

✗

```
> rt(n = 600, df = 1) |> median()
[1] -0.02297411
> rt(n = 600, df = 1) |> median()
[1] 0.02515488
> rt(n = 600, df = 1) |> median()
[1] 0.03453413
> rt(n = 600, df = 1) |> median()
[1] 0.008825645
> rt(n = 600, df = 1) |> median()
[1] -0.01210182
> rt(n = 600, df = 1) |> median()
[1] 0.07435611
```

✓

$$\sigma_X^2 = V(X) = \frac{\nu}{\nu - 2} \qquad \text{para } \nu > 2$$

$\nu = 3$

$\Rightarrow \sigma^2 = \frac{3}{3-2} = 3$

```
> rt(n = 600, df = 1) |> var()
[1] 10140.67
> rt(n = 600, df = 1) |> var()
[1] 36200.02
> rt(n = 600, df = 1) |> var()
[1] 1200.476
> rt(n = 600, df = 1) |> var()
[1] 13178.74
> rt(n = 600, df = 1) |> var()
[1] 84.30659
> rt(n = 600, df = 1) |> var()
[1] 375.9205
```

```
> rt(n = 600, df = 2) |> var()
[1] 7.271294
> rt(n = 600, df = 2) |> var()
[1] 5.810418
> rt(n = 600, df = 2) |> var()
[1] 7.642794
> rt(n = 600, df = 2) |> var()
[1] 10.91618
> rt(n = 600, df = 2) |> var()
[1] 8.976006
> rt(n = 600, df = 2) |> var()
[1] 4.2733
```

```
> rt(n = 600, df = 3) |> var()
[1] 2.560216
> rt(n = 600, df = 3) |> var()
[1] 2.467452
> rt(n = 600, df = 3) |> var()
[1] 4.138305
> rt(n = 600, df = 3) |> var()
[1] 2.393392
> rt(n = 600, df = 3) |> var()
[1] 2.167202
> rt(n = 600, df = 3) |> var()
[1] 2.488638
```

Distribución Chi cuadrado

$$\left.\begin{array}{l} Z_1 \sim N(0,1) \\ Z_2 \sim N(0,1) \\ \vdots \\ Z_n \sim N(0,1) \end{array}\right\} \left(Z_1^2 + Z_2^2 + \cdots + Z_n^2\right) \sim \chi^2_{(n)}$$

$$Z^2 \sim \chi^2_{(1)}$$

Función de densidad

Se define la V.A. X con distribución Chi cuadrado con grado de libertad k, cuya función de densidad viene dada por:

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2-1)} e^{-x/2}, \quad x \geq 0$$

Notación: $X \sim \chi^2_{(k)}$

$$X \sim Gamma\left(\alpha = \frac{k}{2}, \beta = \frac{1}{2}\right)$$

$$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

$$\frac{\left(\frac{1}{2}\right)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$$

Distribución F (de Fisher)

$$X_1 \sim \chi^2_{(m)} \qquad X_2 \sim \chi^2_{(n)} \Rightarrow \frac{X_1/m}{X_2/n} \sim F(m,n)$$

	Chi Cuadrado	F
parámetros	gl	gl1, gl2
Rango	> 0	> 0
Simetría (N)	gl ↑	gl1 ↑, gl2 ↑

$$\mu_X = E(X) = \frac{d_2}{d_2 - 2}, \quad \text{si } d_2 > 2$$

```
> rf(n = 600, df1 = 72, df2 = 1) |> mean()
[1] 16769.14
> rf(n = 600, df1 = 72, df2 = 1) |> mean()
[1] 913.4748
> rf(n = 600, df1 = 72, df2 = 1) |> mean()
[1] 295.162
> rf(n = 600, df1 = 72, df2 = 1) |> mean()
[1] 1252.862
> rf(n = 600, df1 = 72, df2 = 2) |> mean()
[1] 9.986078
> rf(n = 600, df1 = 72, df2 = 2) |> mean()
[1] 425.1741
> rf(n = 600, df1 = 72, df2 = 2) |> mean()
[1] 9.296252
> rf(n = 600, df1 = 72, df2 = 2) |> mean()
[1] 12.23178
> rf(n = 600, df1 = 72, df2 = 2) |> mean()
[1] 6.738412
> rf(n = 600, df1 = 72, df2 = 2) |> mean()
[1] 7.207723
```

> 2

```
> rf(n = 600, df1 = 72, df2 = 12) |> mean()
[1] 1.213212
> rf(n = 600, df1 = 72, df2 = 12) |> mean()
[1] 1.205092
> rf(n = 600, df1 = 72, df2 = 12) |> mean()
[1] 1.20865
> rf(n = 600, df1 = 72, df2 = 12) |> mean()
[1] 1.254181
> rf(n = 600, df1 = 72, df2 = 12) |> mean()
[1] 1.187558
> rf(n = 600, df1 = 72, df2 = 12) |> mean()
[1] 1.192083
```

$d_2 = 12$

$\mu = \frac{12}{12-2} = 1.2$

$$\sigma_X^2 = V(X) = \frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)}, \quad \text{si } d_2 > 4$$

```
> rf(n = 600, df1 = 72, df2 = 3) |> var()
[1] 44.86641
> rf(n = 600, df1 = 72, df2 = 3) |> var()
[1] 79.4811
> rf(n = 600, df1 = 72, df2 = 3) |> var()
[1] 24.36929
> rf(n = 600, df1 = 72, df2 = 3) |> var()
[1] 123.1916
> rf(n = 600, df1 = 72, df2 = 3) |> var()
[1] 69.18664
> rf(n = 600, df1 = 72, df2 = 3) |> var()
[1] 42.00017
```

```
> d1 = 72
> d2 = 30
> v  = 2*d2**2*(d1+d2-2)/(d1*(d2-2)^2*(d2-4))
> v
[1] 0.1226452
```

→ Teórico = poblacional

```
> rf(n = 600, df1 = 72, df2 = 30) |> var()
[1] 0.1303321
> rf(n = 600, df1 = 72, df2 = 30) |> var()
[1] 0.1366566
> rf(n = 600, df1 = 72, df2 = 30) |> var()
[1] 0.1294137
> rf(n = 600, df1 = 72, df2 = 30) |> var()
[1] 0.1343175
```

muestrales = estimadores

Otras distribuciones continuas

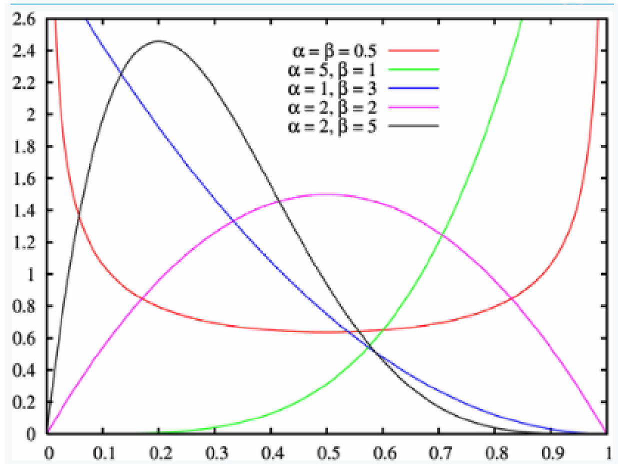
- Distribución Beta ✓
- Distribución Log-Normal
- Distribución Weibull
- Distribución Cauchy
- Distribución Pareto
- Distribución Rayleigh

$X \sim Bin(n = 1, \pi = 0.7)$

$R_X = \{0, 1\}$

$X \sim Beta(\alpha = 5, \beta = 3)$

$R_X \in [0, 1]$



Función de densidad de probabilidad

$Beta(1, 1) \equiv U(0, 1)$