# Distribuciones y Funciones de Probabilidad

Febrero 26, 2024

Prof. Sergio Béjar

Departamento de Estudios Políticos, CIDE

# Objetivo de hoy

*Discutir distribuciones de probabilidad.*

# Introducción

La semana pasada repasamos probabilidad y conteo.

- Aunque son conceptos abstractos, constituyen una base important para lo que hacemos en estadística aplicada.

Hoy vamos a hablar de distribuciones de probabilidad.

- Nuestra herramienta más importante de inferencia estadística hace supuestos sobre los parámetros dado que conocemos su distribución (por ejemplo, la normal).

# Recordatorio

Acordémonos de la notación de las *combinaciones*:

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} \tag{1}$$

El signo de admiración indica factorial.

- e.g. $5! = 5 * 4 * 3 * 2 * 1$.

# Binomial Theorem

El uso más común de la notación de combinaciomes es el **teorema binomial**.

- Para los números reales $X$ and $Y$ y un entero no negativo $n$,

$$(X+Y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \tag{2}$$

Un caso especial ocurre cuando $X = 1$ y $Y = 1$.

$$2^n = \sum_{k=0}^{n} \binom{n}{k} \tag{3}$$

La forma general fue derivada por Pascal en 1654.

# Binomial Theorem

La expansión binomial se incrementa polinomialmente de la siguiente forma:

$$\begin{aligned}
(X+Y)^0 &= 1 \\
(X+Y)^1 &= X+Y \\
(X+Y)^2 &= X^2+2XY+Y^2 \\
(X+Y)^3 &= X^3+3X^2Y+3XY^2+Y^3 \\
(X+Y)^4 &= X^4+4X^3Y+6X^2Y^2+4XY^3+Y^4 \\
(X+Y)^5 &= X^5+5X^4Y+10X^3Y^2+10X^2Y^3+5XY^4+Y^5
\end{aligned} \tag{4}$$

Aquí hay que notar la simetría.

# Triángulo de Pascal

los coeficientes del **triángulo de Pascal** resumen los coeficientes en una expansión binomial.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $n = 0$: | | | | | | 1 | | | | | |
| $n = 1$: | | | | | 1 | | 1 | | | | |
| $n = 2$: | | | | 1 | | 2 | | 1 | | | |
| $n = 3$: | | | 1 | | 3 | | 3 | | 1 | | |
| $n = 4$: | | 1 | | 4 | | 6 | | 4 | | 1 | |
| $n = 5$: | 1 | | 5 | | 10 | | 10 | | 5 | | 1 |

# Triángulo de Pascal

Además de la simetría piramidal, el triángulo de Pascal tiene otras características interesantes.

- Cualquier valor es igual a la suma de los dos números encima de él.
- La suma de los elementos de cualquier fila es igual al doble de la suma de los elementos de la suma anterior (i.e. $\sum_{j=0}^{k} \binom{k}{j} = 2^k$).
- Todos los números que se encuentran en el borde del triángulo son iguales a 1.

El triángulo de Pascal tiene más propiedades matemáticas y estas son algunas de las más famosas.

# Estos tienen relación importante con la estadística

**Ejemplo**: ¿cuántas veces nos puede salir un águila si lanzamos una moneda 10 veces?

- Espacio muestra $S = \{\ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\ \}$
- Sacar 10 águilas (o no águilas) es altamente improbable.

# Función de Densidad de Probabilidad Binomial

*Este es un problema de combinaciones*.

- Para tener 0 águilas, *cada* lanzamiento debe ser un sol.
- Para sacar sólo un águila, hay más combinaciones posibles.

¿cuántas maneras hay de que una serie de lanzamientos de una moneda resulte en solo un águila?

- Para un número pequeño de lanzamientos podemos ver el triángulo de Pascal.
- Para 5 lanzamientos, hay 1 forma de obtener 0 águilas, 5 maneras de obtener 1 águila, 10 formas de obtener 2 y 3 águilas, 5 formas de obtener 4 águilas, y 1 forma de obtener 5 águilas.

# Función de Densidad de Probabilidad Binomial

Con más generalidad podemos usar la **función de densidad de probabilidad binomial** para resolver el ejemplo anterior.

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \tag{5}$$

donde:

- $x =$ la cuenta de "éxitos" o resultados favorables (e.g. número de águilas en una secuencia de lanzamientos)
- $n =$ el número de intentos.
- $p =$ probabilidad de éxito en cualquier intento.

# Función de Densidad de Probabilidad Binomial

**Ejemplo:** ¿Cuál es la probabilidad de obtener cinco soles en 10 lanzamientos de una moneda?

$$\begin{aligned} p(x = 5 \mid n = 10, p = .5) &= \binom{10}{5}(.5)^5(1-.5)^{10-5} \\ &= (252) * (.03125) * (.03125) \\ &= 0.2460938 \end{aligned} \tag{6}$$

En R:

```
dbinom (5,10,.5)
```

```
## [1] 0.2460938
```

# Aplicación: La Decadencia de la Guerra?

Pinker (2011) argumenta que la ausencia de guerras mundiales desde la Segunda Guerra Mundial demuestra que hay una disminución de la violencia. Pero:

- Este tipo de guerras son extremadamente raras.
- Gibler and Miller (2022) codifican 1,958 confrontaciones de 1816 a 2014.
- De esas: 84 son guerras ($p = .042$)
- De las guerras, solo 24 son guerras que podríamos decir que son guerras "realmente grandes" ($p = .012$)

# La Decadencia de la Guerra?

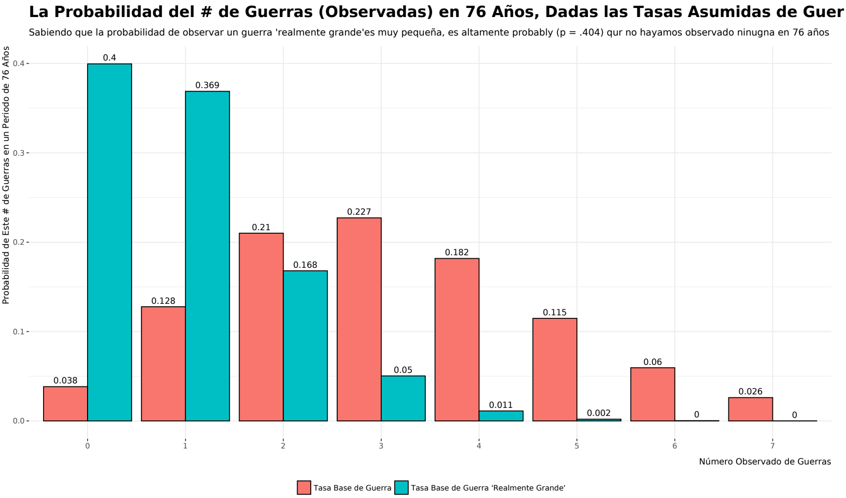
Estamos en 2024. No hemos observado una guerra de magnitud similar a la Segunda Guerra Mundial en 76 años in.

Cuál es la probabilidad de que *no* observemos esto donde:

- $p = .042$, la probabilidad base de guerra vs. no guerra?
- $p = .012$, la probabilidad base de una guerra "realmente grande"?

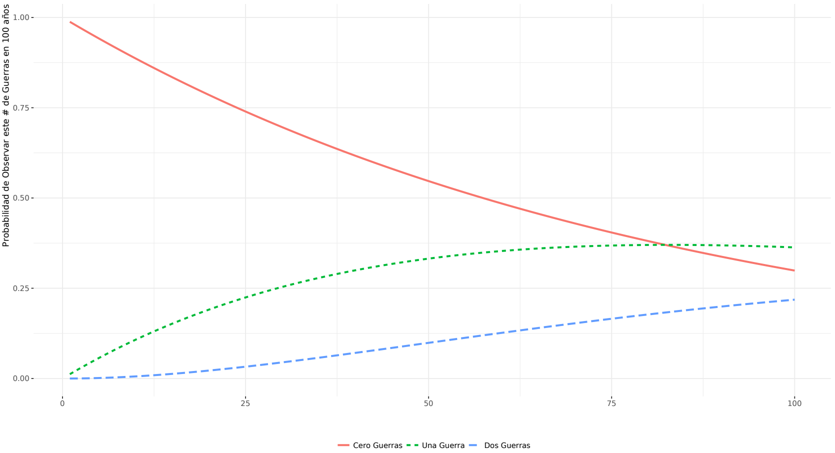
```r
tibble(num_wars = seq(0:7)-1,
       base = dbinom(num_wars, 76, .042),
       rbw = dbinom(num_wars, 76, .012))

tibble(num_wars = rep(c(0, 1, 2), 100)) %>%
  arrange(num_wars) %>%
  mutate(period = rep(seq(1:100), 3),
         p = dbinom(num_wars, period, 0.012))
```


La Probabilidad del # de Guerras (Observadas) en 76 Años, Dadas las Tasas Asumidas de Guer
Sabiendo que la probabilidad de observar un guerra 'realmente grande'es muy pequeña, es altamente probably (p = .404) qur no hayamos observado ninugna en 76 años
Probabilidad de Este # de Guerras en un Periodo de 76 Años
0.0
0.1
0.2
0.3
0.4
0.038
0.4
0.128
0.369
0.21
0.168
0.227
0.05
0.182
0.011
0.115
0.002
0.06
0
0.026
0
0
1
2
3
4
5
6
7
Número Observado de Guerras
Tasa Base de Guerra
Tasa Base de Guerra 'Realmente Grande'

## La Probabilidad de Observar un Número Fijo de 'Grandes Guerras' en un Periodo de 100 años

Después de 76 años, todavía es más probable que todavía no hayamos observado una 'Gran Guerra' que haber observado solo una.



# Funciones Normales

Son otro tipo de funciones muy comúnes.

- Los datos se distribuyen de tal forma que la mayoría se concentran alrededor de alguna tendencia central.
- Los casos en los extremos ocurren con menos frecuencia.

# Función de Densidad Normal

Lo vamos a modelar con la **función de densidad normal**.

- Algunas veces llamada distribución Gaussiana, llamada así Carl Friedrich Gauss, su descubridor.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e\{-\frac{(x-\mu)^2}{2\sigma^2}\}, \tag{7}$$

donde: $\mu =$ la media, $\sigma^2 =$ la varianza.

# Función de Densidad Normal

Propiedades:

- Las colas tienden asintóticamente hacia 0.
- El kernel (o núcleo) es una parábola básica.
- Denotado como una función en lugar de una probabilidad porque la distribución es continua.
- La distribución es perfectamente simétrica.
  - La moda/media/mediana son los mismos valores.
  - $-x$ es tan lejos de $\mu$ como $x$.

# Función de Densidad Normal

$x$ sin restricciones. Puede ser cualquier valor en la distribución.

- $\mu$ y $\sigma^2$ son parámetros que definen la forma de la distribución.
  - $\mu$ define la tendencia central.
  - $\sigma^2$ define que tan corta/larga es la distribución.

# Desmenuzando la Función de Densidad Normal

Desempaquemos la función de densidad normal (y usemos un poco de `R`).

# Desmenuzando la Función de Densidad Normal

Esta es la función de densidad normal.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e\{-\frac{(x-\mu)^2}{2\sigma^2}\} \tag{8}$$

Asumamos, para simplificar, $\mu = 0$ y $\sigma^2 = 1$.

# Desmenuzando la Función de Densidad Normal

Cuando $\mu = 0$ y $\sigma^2 = 1$, la función de densidad normal es más simple.

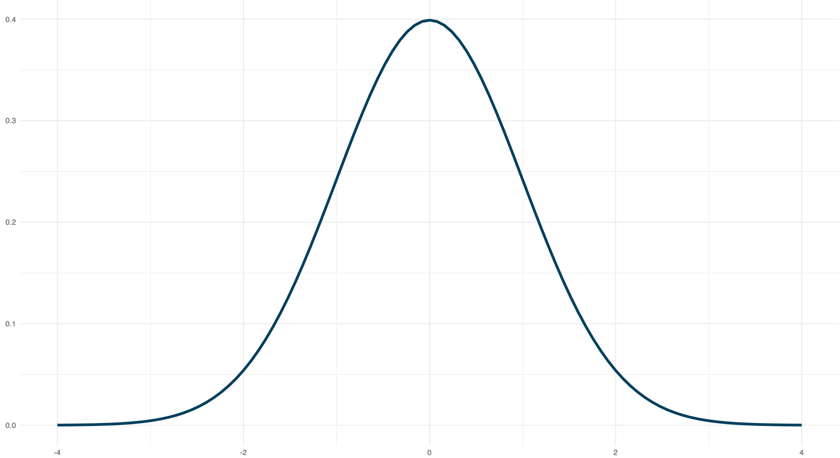
$$f(x) = \frac{1}{\sqrt{2\pi}} e\{-\frac{x^2}{2}\} \tag{9}$$

Ahora, grafiquemos en R.

```r
ggplot(data.frame(x = c(-4, 4)), aes(x)) +
  theme_minimal() +
  stat_function(fun = dnorm, color="#003f5c", size=1.5)
```

# Una Distribución Normal Simple

El parámetro mu determina la tendencia central y sigma^2 lo ancho.



# Desmenuzando la Función de Densidad Normal

Veámos dentro del exponencial.

- El término dentro de los paréntesis (-$x^2/2$) es una parábola.
- Al exponenciarlo, lo hacemos asintótico a 0.

# Código de R

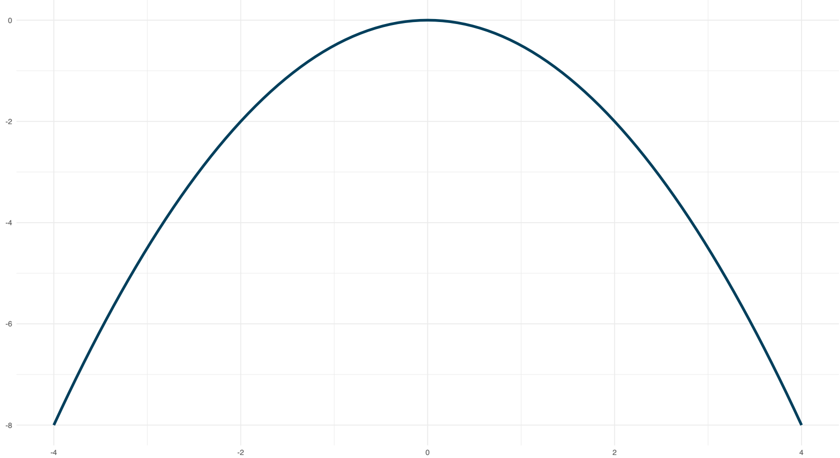
```r
library(ggplot2)
parab <- function(x) {-x^2/2}
expparab <- function(x) {exp(-x^2/2)}

ggplot(data.frame(x = c(-4, 4)), aes(x)) +
  stat_function(fun = parab, color="#003f5c", size=1.5) +
  theme_minimal()

ggplot(data.frame(x = c(-4, 4)), aes(x)) +
  stat_function(fun = expparab, color="#003f5c", size=1.5) +
  theme_minimal()
```

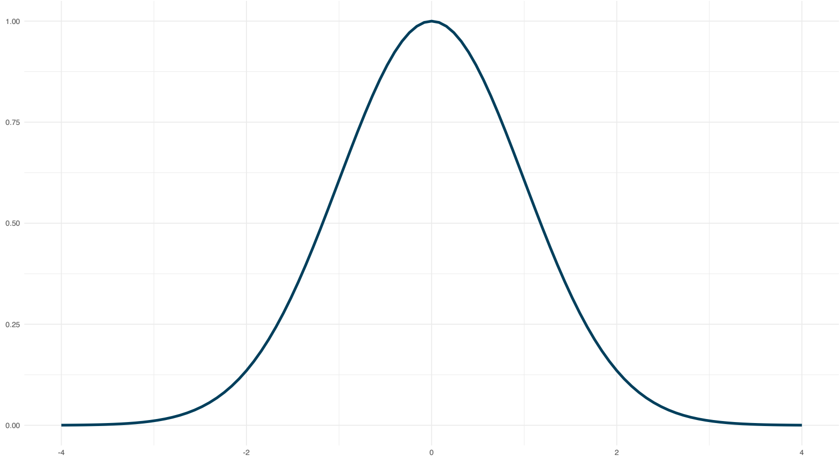
# Una Parábola Básica

Hay que notar que la altura está en el 0 porque la parte negativa voltea la parábola hacia abajo.



# Una Parábola Exponencial Negativa

Al exponenciar, comprimímos la parábola, ajustamos la altura , y hacemos las colas asintóticas a 0.



# Desmenuzando la Distribución Normal

Cuando el número entre paréntesis es cero (i.e. $x = \mu$, aquí: 0), esto se convierte en un exponente de 0.

- $exp(0) = 1$ (inversamente, $log(1) = 0$).
- El logaritmo de $x$ para una base $b$ es el valor del exponente que convierte $b$ a $x$.
  - $log_b(x) = a \quad \Longrightarrow \quad b^a = x$
- Noten como el tope de la curva estaba en 1 en la parábola exponencial.

# Desmenuzando la Distribución Normal

Teniendo en mente lo anterior, debe ser claro que $\frac{1}{\sqrt{2\pi\sigma^2}}$ (recordar: $\sigma^2 = 1$ en nuestro caso simple) determina la altura de la distribución.

# Desmenuzando la Distribución Normal

Observe:

```
1/sqrt(2*pi)
```

```
## [1] 0.3989423
```

```
dnorm(0,mean=0,sd=1)
```

```
## [1] 0.3989423
```

La altura de la distribución para $x = 0$ cuando $\mu = 0$ y $\sigma^2 = 1$ es .3989423.

# Desmenuzando la Distribución Normal

Hemos hablado de la altura y la forma de la distribución como una *función*. Pero no hemos dicho nada de las probabilidades.

- La distribución normal es continua. Por tanto, la probabilidad para cada uno de los valores posibles es practicamente 0.

Dicho esto, el área *debajo* de la curva es el dominio completo y es igual a 1.

- La probabilidad de seleccionar un número entre dos puntos del eje de las Xs es igual al área bajo la curva que está *entre* esos dos puntos.

# Desmenuzando la Función de Densidad Normal

Observemos que:

```r
pnorm(0, mean=0, sd=1)
```

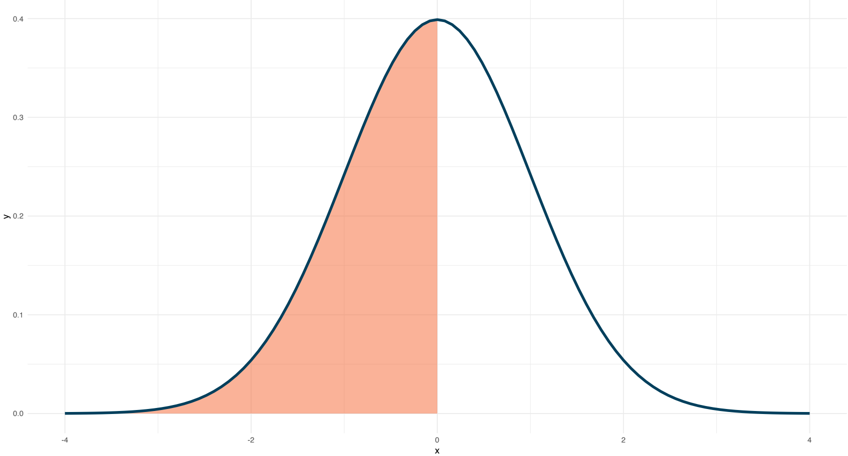
```
## [1] 0.5
```

# Desmenuzando la Función de Densidad Normal

```
ggplot(data.frame(x = c(-4, 4)), aes(x)) +
  theme_minimal() +
  theme(plot.title = element_text(size=30)) +
  stat_function(fun = dnorm,
                xlim = c(-4,0),
                size=0,
                geom = "area", fill="#F66733", alpha=.5) +
    stat_function(fun = dnorm, color="#003f5c", size=1.5)
```

# Una Distribución Normal Stándard

Hay que notar que la mitad de la distribución está entre -Infinito y 0.




-Infinito a 0 tiene el 50% del área debajo de la curva

# 68-90-95-99

```r
pnorm(1,mean=0,sd=1)-pnorm(-1,mean=0,sd=1)
```

```
## [1] 0.6826895
```

```r
pnorm(1.645,mean=0,sd=1)-pnorm(-1.645,mean=0,sd=1)
```

```
## [1] 0.9000302
```

```r
pnorm(1.96,mean=0,sd=1)-pnorm(-1.96,mean=0,sd=1)
```

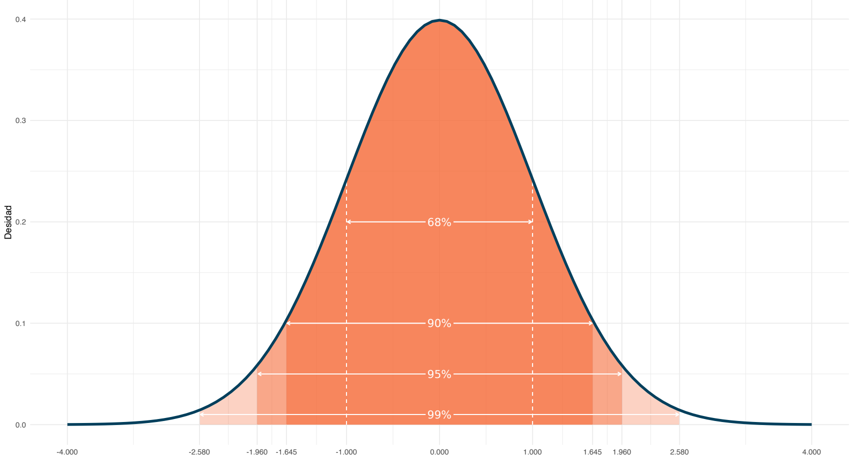
```
## [1] 0.9500042
```

```r
pnorm(2.58,mean=0,sd=1)-pnorm(-2.58,mean=0,sd=1)
```

```
## [1] 0.99012
```

# El Área por Debajo de la Distribución Normal

Las colas se extienden hacia el infinito y son asintóticas a 0, pero el área completa es igual a 1. 95% de todos los posibles valores están aproximadamente 1.96 unidades standard de la media.



# Conclusión

- Hay mucha información que tenemos que digerir en esta presentación y la anterior.
- La probabilidad y las funciones de probabilidad son básicas para la estadística inferencial que empezaremos a hacer la próxima clase.

# Table of Contents