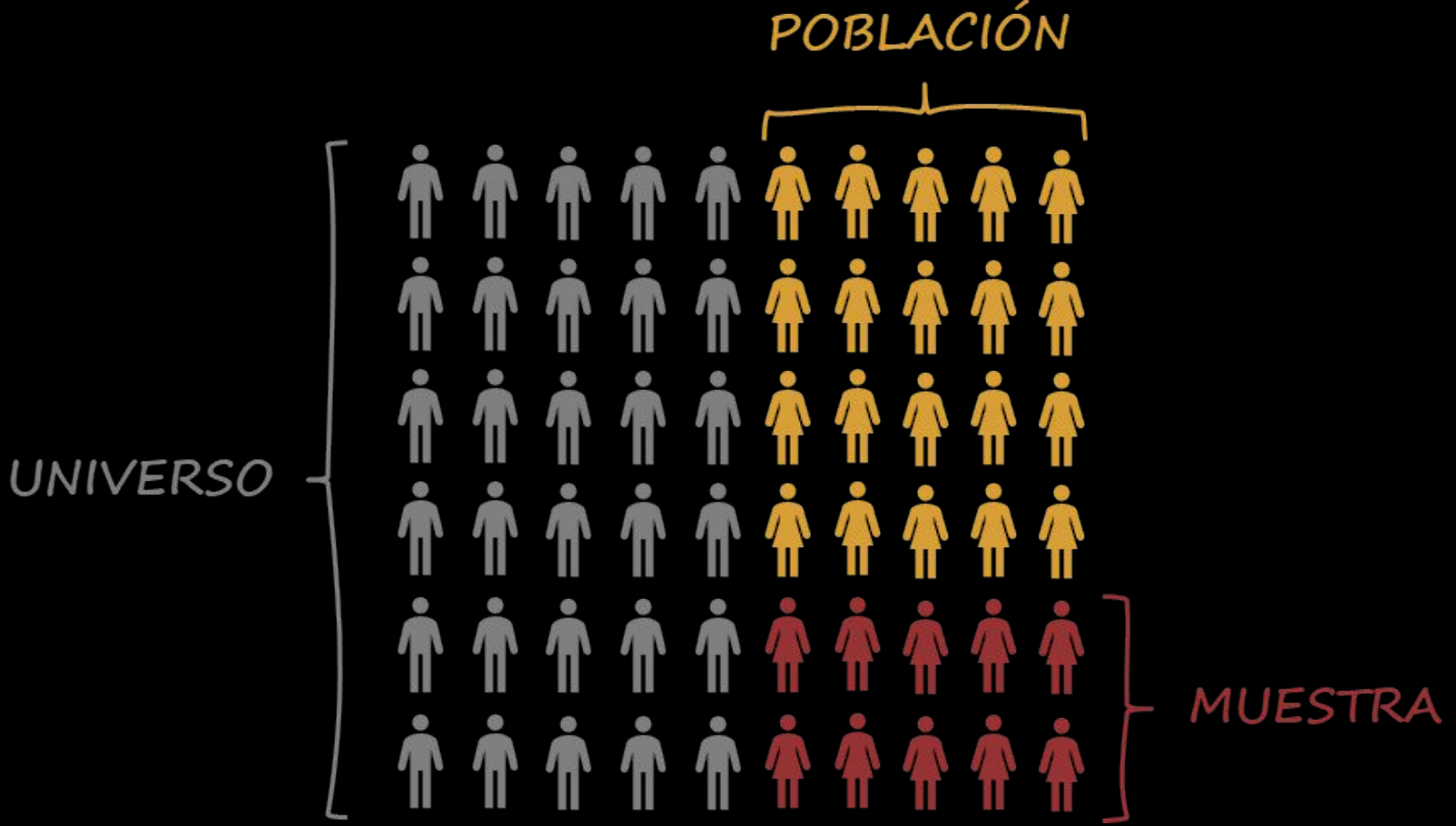
		Outcome variable					
		Nominal	Categorical (>2 Categories)	Ordinal	Quantitative Discrete	Quantitative Non-Normal	Quantitative Normal
Input Variable	Nominal	X^2 or Fisher's	X^2	X^2-trend or Mann - Whitney	Mann-Whitney	Mann-Whitney or log-rank[a]	Student's *t* test
	Categorical (2>categories)	X^2	X^2	Kruskal-Wallis[b]	Kruskal-Wallis[b]	Kruskal-Wallis[b]	Analysis of variance[c]
	Ordinal (Ordered categories)	X^2-trend or Mann - Whitney	e	Spearman rank	Spearman rank	Spearman rank	Spearman rank or linear regression[d]
	Quantitative Discrete	Logistic regression	e	e	Spearman rank	Spearman rank	Spearman rank or linear regression[d]
	Quantitative non-Normal	Logistic regression	e	e	e	Plot data and Pearson or Spearman rank	Plot data and Pearson or Spearman rank and linear regression
	Quantitative Normal	Logistic regression	e	e	e	Linear regression[d]	Pearson and linear regression

		Outcome variable					
		Nominal	Categorical (>2 Categories)	Ordinal	Quantitative Discrete	Quantitative Non-Normal	Quantitative Normal
Input Variable	Nominal	X^2 or Fisher's	X^2	X^2-trend or Mann - Whitney	Mann-Whitney	Mann-Whitney or log-rank[a]	Student's *t* test
	Categorical (2>categories)	X^2	X^2	Kruskal-Wallis[b]	Kruskal-Wallis[b]	Kruskal-Wallis[b]	Analysis of variance[c]
	Ordinal (Ordered categories)	X^2-trend or Mann - Whitney	e	Spearman rank	Spearman rank	Spearman rank	Spearman rank or linear regression[d]
	Quantitative Discrete	Logistic regression	e	e	Spearman rank	Spearman rank	Spearman rank or linear regression[d]
	Quantitative non-Normal	Logistic regression	e	e	e	Plot data and Pearson or Spearman rank	Plot data and Pearson or Spearman rank and linear regression
	Quantitative Normal	Logistic regression	e	e	e	Linear regression[d]	Pearson and linear regression

PRUEBAS PARAMÉTRICAS Y NO PARAMÉTRICAS

Ciencia de Datos

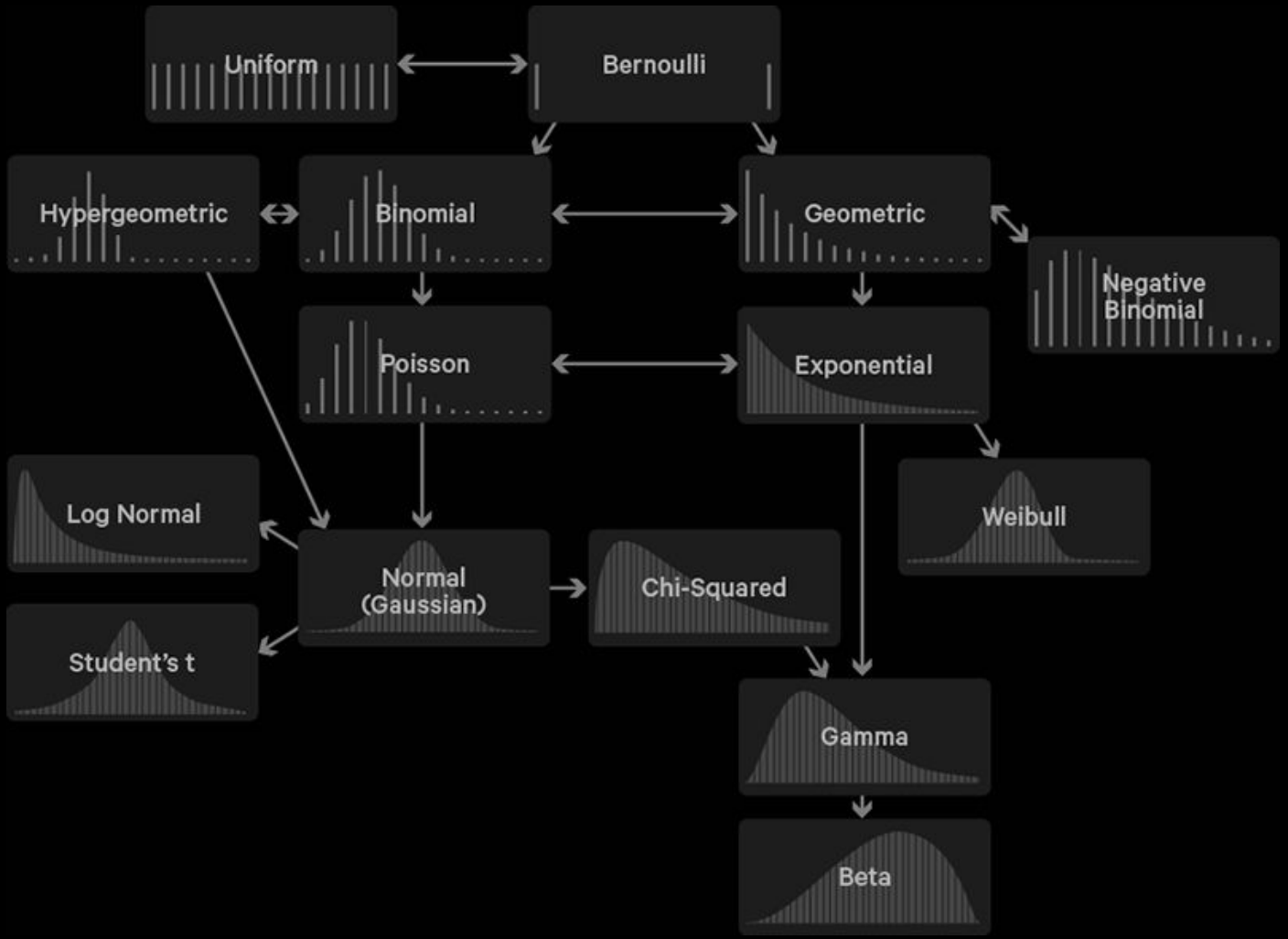
INFERENCIA ESTADÍSTICA



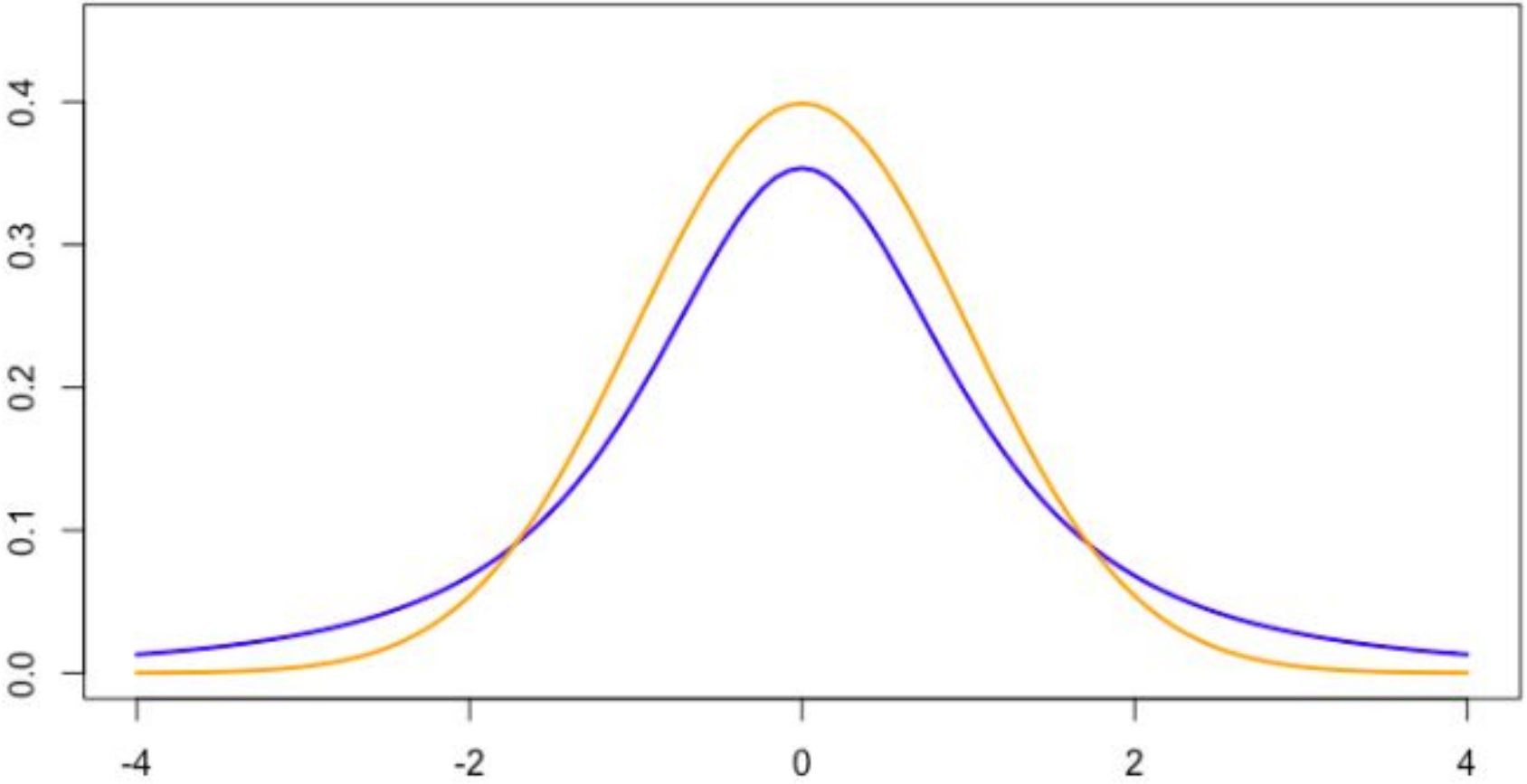
MUESTRA

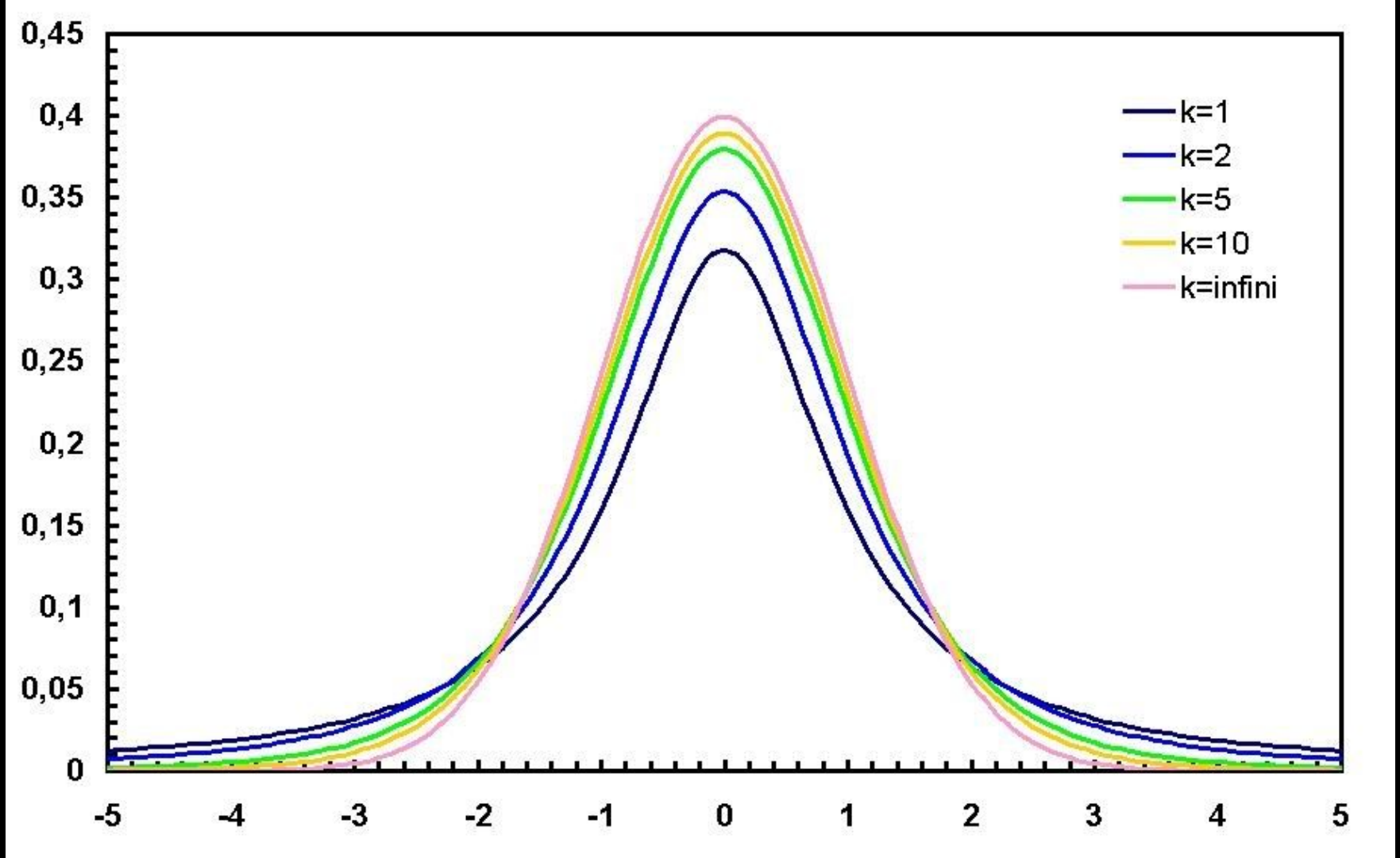
POBLACIÓN

DISTRIBUCIONES DE FRECUENCIA



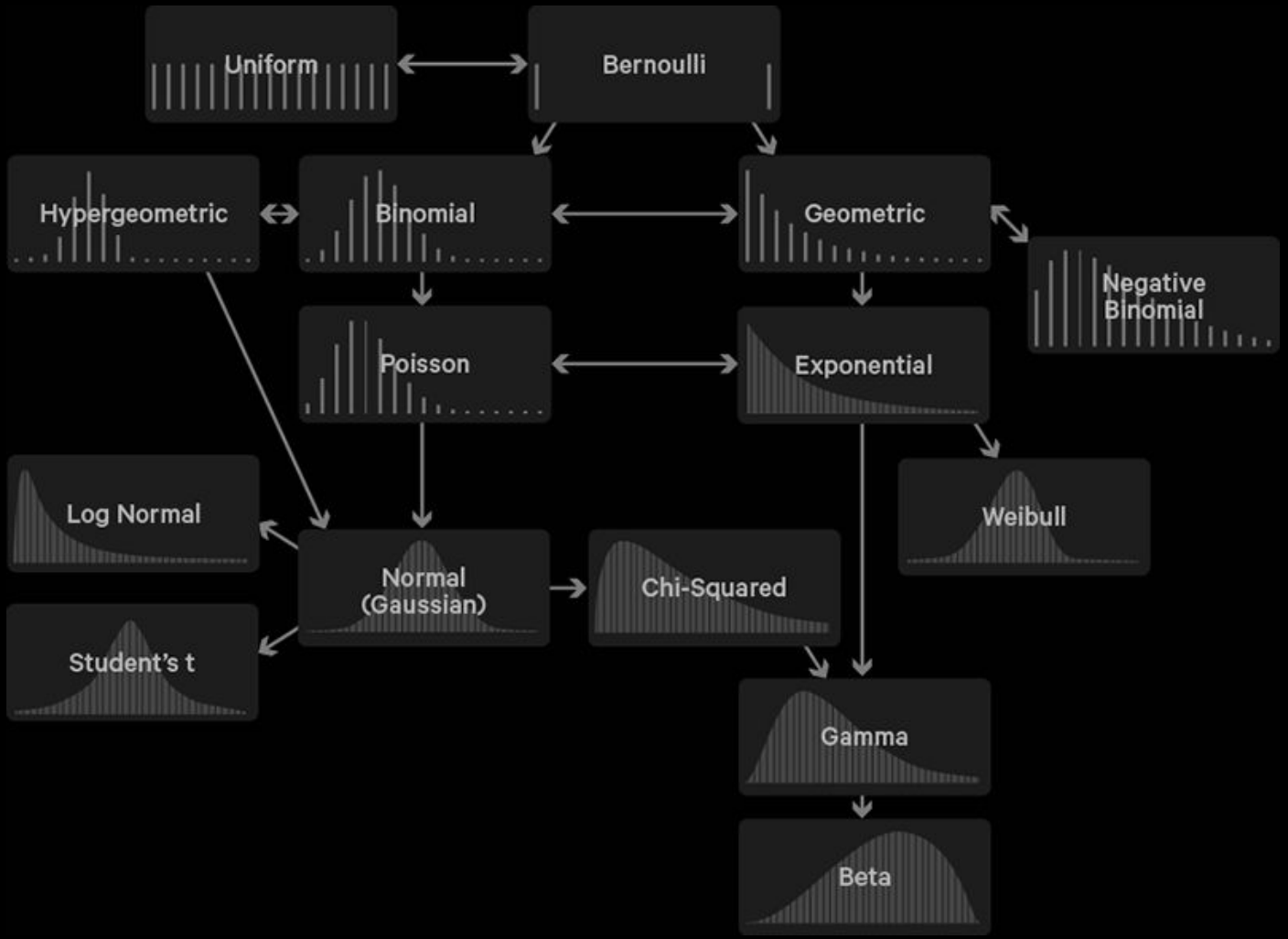
Distribución t de Student (azul) y distribución Normal estándar N(0,1) (naranja)




0,45
0,4
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
-5
-4
-3
-2
-1
0
1
2
3
4
5
k=1
k=2
k=5
k=10
k=infini

TENEMOS UNA DIFICULTAD

DISTRIBUCIONES DE FRECUENCIA



OPCIÓN 1: HACEMOS SUPUESTOS

DISTRIBUCIÓN NORMAL

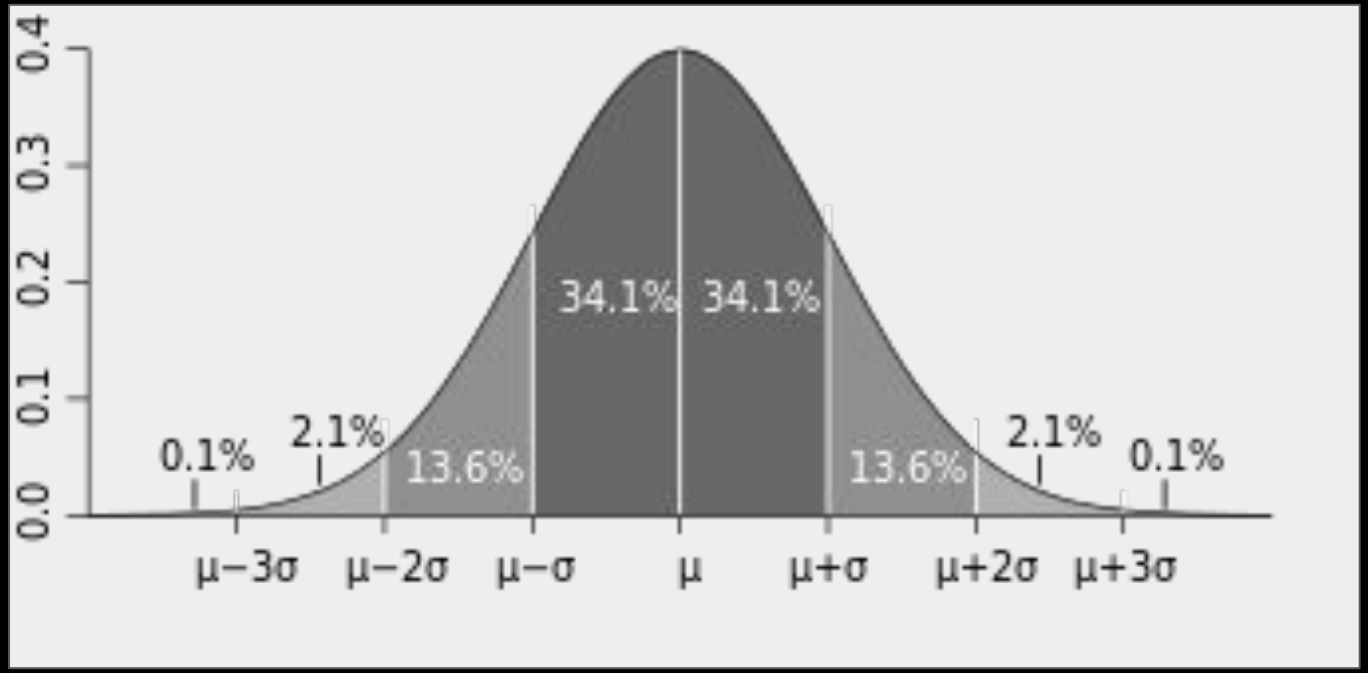
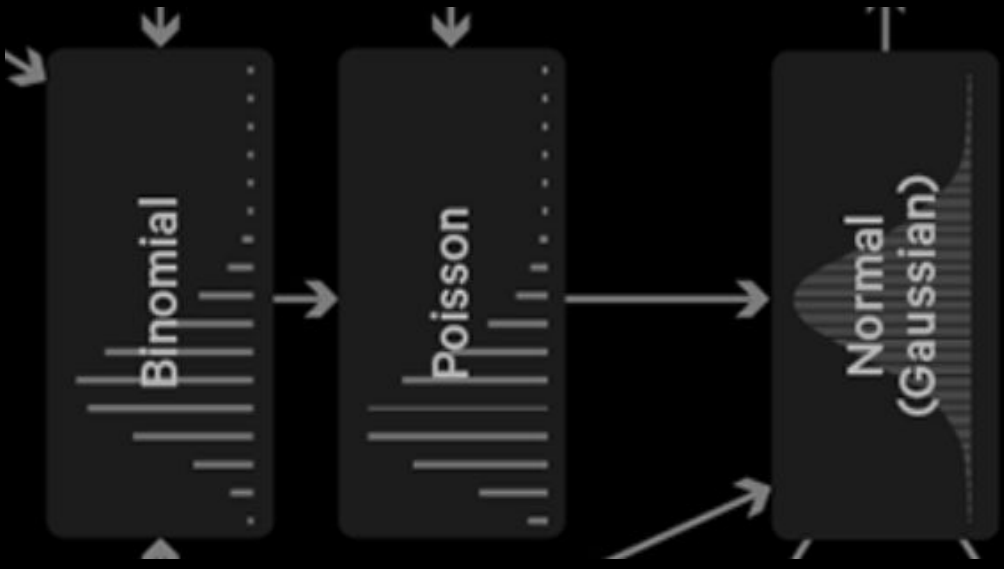


Tabla 1. Pruebas paramétricas

Prueba t de 1 muestra
Prueba t de 2 muestras
ANOVA de un solo factor
Correlación Pearson
Regresión lineal

DISTRIBUCIÓN MÁS FRECUENTES



Recomendables si

1. La muestra es grande
2. La distribución poblacional es conocida
3. Funcionan más rapidamente
4. Se usan variables de intervalo o razón

Ventajas:

Más potencia estadística

OPCIÓN 2: LIBRES DE DISTRIBUCIÓN

Tabla 2. Pruebas no paramétricas

Wilcoxon para 1 muestra
Prueba de Mann-Whitney
Kruskal-Wallis
Mediana de Wood
Prueba de Friedman

Recomendables si

1. La muestra es pequeña
2. La distribución poblacional es desconocida
3. Se usan variables categoriales u ordinales, aunque también hay para intervalo o razón

Análisis con variables numéricas:

Análisis	Paramétrico	No paramétrico
Describir un grupo	μ, σ^2	Mediana, rango intercuartil
Comparar un grupo a un valor	T Student de una muestra	Prueba Wilcoxon
Comparar medias en 2 grupos	T Student de dos muestras	Mann-Whitney
Comparar medias en 2 grupos apareados	T Student apareada	Prueba Wilcoxon
Comparar medias en 3 o mas grupos	ANOVA	Kruskal-Wallis
Correlación entre dos variables	Pearson (lineal)	Spearman (monotónica)

¿CÓMO DETERMINAR?

Criterios

1. Fuente de los datos
2. Información en literatura
3. Prueba de normalidad

```python
from scipy import stats
rng = np.random.default_rng()
pts = 1000
a = rng.normal(0, 1, size=pts)
b = rng.normal(2, 1, size=pts)
x = np.concatenate((a, b))
k2, p = stats.normaltest(x)
alpha = 1e-3
print("p = {:g}".format(p))
p = 8.4713e-19
if p < alpha:  # null hypothesis: x comes from a normal distribution
    print("The null hypothesis can be rejected")
else:
    print("The null hypothesis cannot be rejected")
```

```
p = 1.21879e-12
The null hypothesis can be rejected
```

PRUEBA CON CARS

```
df = pd.read_csv("cars.csv")
odo = df[["odometer_value"]]
plt.hist(df[["odometer_value"]])
k2, p = stats.normaltest(odo)
alpha = 1e-3
print("p = ", p)
p = 8.4713e-19
if p < alpha:  # null hypothesis: x comes from a normal distribution
    print("The null hypothesis can be rejected")
else:
    print("The null hypothesis cannot be rejected")
```

```
p =  [0.]
The null hypothesis can be rejected
```