



## Variables Aleatorias y Gráficos\_

Sesión Presencial 2

## ¿Qué es una Hipótesis?

- Idea fundacional de la inferencia  $\rightsquigarrow$  **tomar decisiones o esclarecer juicios en base a información limitada.**

### **Definición:**

*Juicio empíricamente comprobable sobre la relación entre dos o más variables. El juicio desarrollado se asume como incierto frente a la relación, dado que no tenemos certeza sobre su existencia bajo condiciones específicas.*

## Componentes necesarios

- **Hipótesis Nula:** Ausencia de Efecto.
- **Hipótesis Alternativa:** Juicio a contrastar. En su caso más general se asume que la  $H_a$  es distinta a cero.
- **Estadístico de Prueba:** Nuestro estimado.
- **Distribución Nula:** En su versión más general, se asume que  $H_0 \sim \mathcal{N}(0, 1)$ .
- **Puntaje de corte:** Criterio arbitrario para evaluar. Generalmente se manifiestan en *P-Values*.

## Forma canónica del estadístico de prueba

$$Z = \frac{\hat{\theta} - \theta}{\sigma/\sqrt{n}} \Rightarrow \frac{\underbrace{\hat{\theta}}_{\text{Estimador Muestral}} - \underbrace{\theta}_{\text{Valor Poblacional}}}{\underbrace{\sigma/\sqrt{n}}_{\text{Error estándar}}}$$

- $\hat{\theta}$  punto estimado a prueba.
- $\theta$  valor poblacional (o verdadero) a contrastar.
- $\sigma/\sqrt{n}$  permite ajustar la diferencia entre el punto estimado y el valor poblacional, siguiendo una distribución ~~asintóticamente~~ **asintóticamente** normal.

## Ejemplo: Prueba frente a una constante

- Utilizando `scipy.stats`:

```
In [18]: stats.ttest_1samp(wvs_subset['wvs_trust'], .7)
```

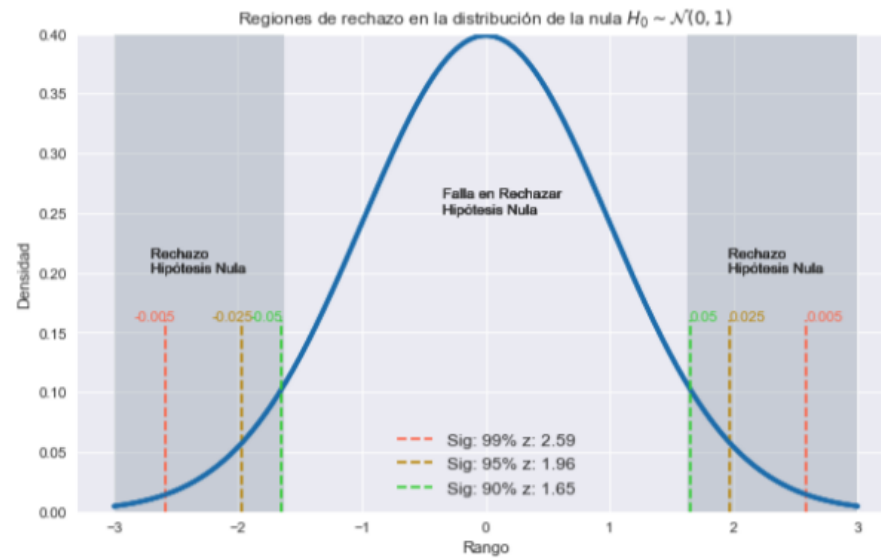
```
Out[18]: Ttest_1sampResult(statistic=-18.770596381293288, pvalue=1.2609756204715342e-22)
```

- Buscamos encontrar valores **absolutos** de `statistic` mayores a 1.96 para tener un 95% de confianza.
- Para este caso, observamos que la media mundial de la confianza interpersonal es distinta al valor a contrastar de .7.

## P-Values

- Cuando evaluamos hipótesis, contrastamos valores frente a una distribución nula. Ésta se puede dividir en dos áreas.

```
In [23]: plt.figure(figsize=(10, 6));gfx.graph_significance()
```



## ¿Qué significa que sea significativa?

**Mantra:** evaluamos si cada experimento (líneas verticales) contienen el parámetro verdadero ( $\theta$ ) o no. Cuando sí lo contienen las líneas verticales se visualizan como azules, de lo contrario se visualizan como rojas. Esta es la visualización más clara del 95% de confianza: **Esperamos que si repetimos este experimento 100 veces, obtendremos un resultado similar en 95 ocasiones.**

```
In [25]: plt.figure(figsize=(10, 6));gfc.confidence_intervals();plt.title(r'Iteración del experimento donde  $H_a = \hat{\theta} \sim \mathcal{N}(\theta, 1)$ ');
```



## Ejemplo: Prueba de hipótesis para muestras independientes

- Comparemos niveles de confianza entre sociedades occidentales y no.
- Las hipótesis correspondientes serían:
- **Hipótesis Nula:** *Las tasas de confianza interpersonal son similares entre los países de europa occidental y el resto del mundo.*
- **Hipótesis Alternativa:** *Las tasas de confianza interpersonal son distintas entre los países de europa occidental y el resto del mundo.*

```
In [21]: wvs_subset['western_dm'] = np.where(wvs_subset['region'] == 'WesternDem', 1, 0)
stats.ttest_ind(wvs_subset.query('western_dm ==1')['wvs_trust'],
                wvs_subset.query('western_dm == 0')['wvs_trust'])
```

```
Out[21]: Ttest_indResult(statistic=5.108950860705963, pvalue=7.079387187592779e-06)
```

**{desafío}**  
latam\_