# PRUEBA DE HIPÓTESIS

ANGÉLICA URQUIZO

# PRUEBA DE HIPOTESIS

## - Técnicas estadísticas


**Las hipótesis planteadas deben ser sometidas a prueba para ver si son apoyadas o refutadas de acuerdo a los resultados de los datos obtenidos. Mientras más investigaciones apoyen una hipótesis, más credibilidad y validez tendrá en el contexto en el que se la planteó.**

## 4.6.1 Errores de tipo I y de tipo II

**Cuando aceptamos o rechazamos una hipótesis podemos cometer errores, por ejemplo:**

**Se conoce como error de TIPO I ($\alpha$) a la probabilidad de rechazar la hipótesis nula cuando ésta es verdadera.**

**Se conoce como error de TIPO II ($\beta$) a la probabilidad de aceptar la hipótesis nula cuando ésta es falsa.**

| | $H_o$ es verdadera | $H_o$ es falsa |
|---|---|---|
| Se rechaza $H_o$ | Error de TIPO I $\alpha$ | Decisión correcta |
| Se acepta $H_o$ | Decisión correcta | Error de TIPO II $\beta$ |

## 4.6.2 Nivel de significación y nivel de confianza

El **error de tipo I** se denota por $\alpha$ y se lo llama **nivel de significación** y generalmente se toma (fijo) los valores entre 0.01 (1%) y 0.15 (15%) (los más usuales son: 0.05 y 0.01). Representa áreas de riesgo (de rechazo de la hipótesis nula ) o confianza (de aceptación de la hipótesis de investigación ), por ejemplo, en la distribución normal (muestral) cuya área total bajo la curva se considera 1.

Se llama **nivel de confianza** al valor $1 - \alpha$, es la zona de aceptación de la hipótesis nula.

A mayor conocimiento del tema menor será el valor de $\alpha$ y viceversa.

## 4.6.3 Pasos para la prueba de hipótesis

**A continuación se indican los pasos generales que se siguen para realizar la prueba de la(s) hipótesis:**

1) **Planteamiento de las hipótesis** (estadísticas)

2) **Nivel de significación** $\alpha$ con el que se pretende rechazar o aceptar la hipótesis nula

3) **Criterio con el que se rechaza la hipótesis nula**

4) **Cálculos**, esto es, la aplicación de la o las fórmulas para hallar los valores calculados y contrastarlos con los valores teóricos

5) **Decisión que se toma de acuerdo a los valores calculados y teóricos.**

## 4.6.4 Técnicas Estadísticas

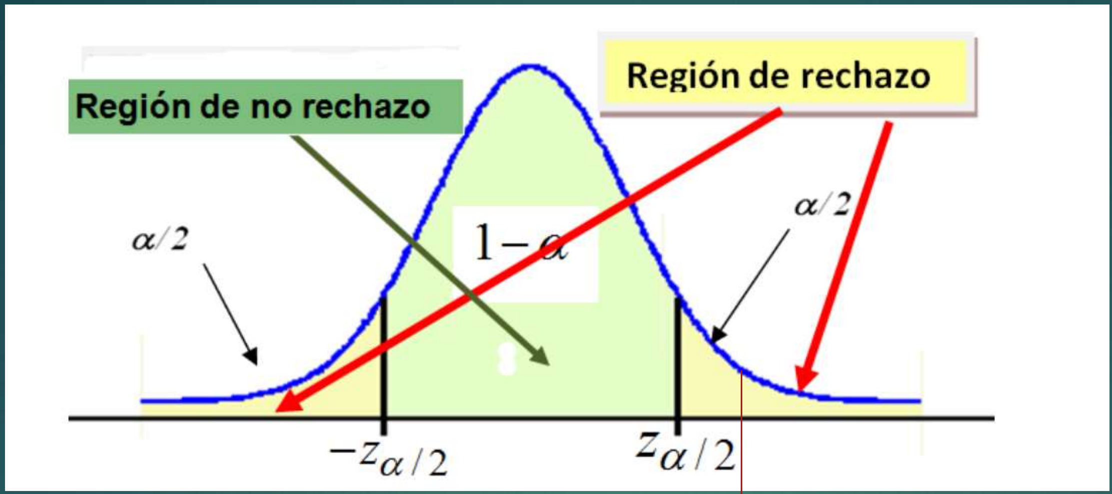
### 4.6.4.1 z Normalizado

**Si el valor de la varianza poblacional $\sigma^2$ es conocido, entonces la estadística de prueba es la media muestral. La distribución *muestral* es una distribución *normal* de puntuaciones z en unidades de desviación estándar.**

**DEF. Se llama *puntuación z* de la distribución (normal) al *valor crítico* que separa las áreas de rechazo y aceptación de la hipótesis nula.**

**En un ensayo a dos colas, se tiene:**

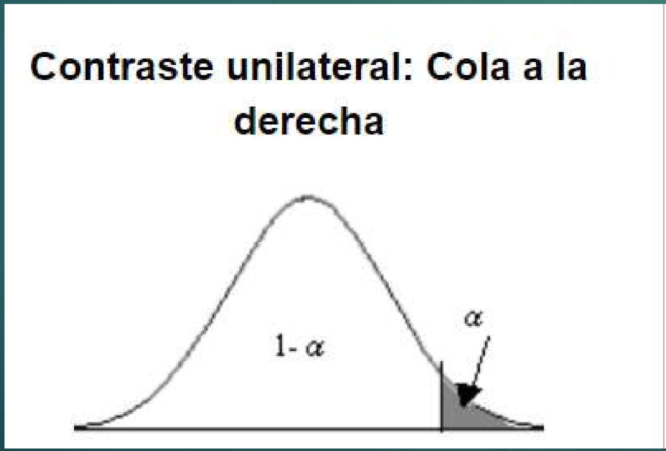
**Para un nivel de significación del 1%** $Z_t = 2.57$

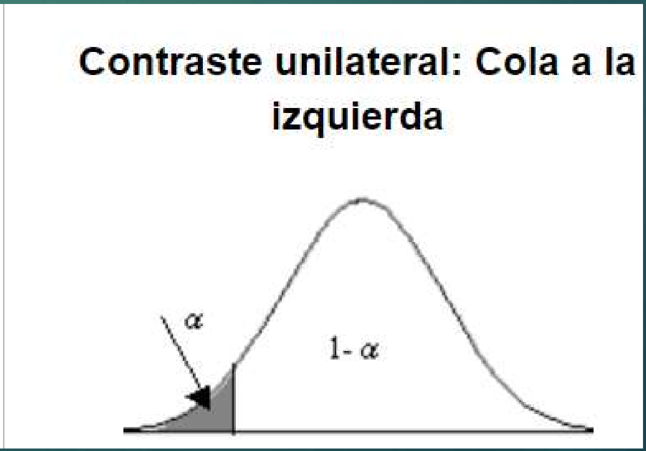



Prueba a dos colas
Hi: $\mu \neq k$

Prueba a una cola
Hi: $\mu > k$

Hi: $\mu < k$



El área entre el centro y el valor teórico se obtiene así: 0.5 – (0.01)/2 = 0.495. Viendo 0.495 en el interior de la tabla del APÉNDICE A, encontramos para 0.4949 que es el más próximo a 0.495, a la izquierda 2.5 y arriba 7; luego, el valor teórico es 2.57

**Para un nivel de significación del 5%** $Z_t$ = **1.96**

El área entre el centro y el valor teórico se obtiene así: 0.5 – (0.05)/2 = 0.475. Viendo 0.475 en el interior de la tabla del APÉNDICE A, encontramos a la izquierda 1.9 y arriba 6; luego, el valor teórico es 1.96

**Para un nivel de significación del 10%** $Z_t$ = **1.64**

El área entre el centro y el valor teórico se obtiene así: 0.5 – (0.10)/2 = 0.45. Viendo 0.45 en el interior de la tabla del APÉNDICE A, encontramos para 0.4495 que es el más próximo a 0.45, a la izquierda 1.6 y arriba 4; luego, el valor teórico es 1.64

**En un ensayo a una cola, se tiene:**

**Para un nivel de significación del 1%** $Z_t$ **= 2.33**

**El área entre el centro y el valor teórico se obtiene así: 0.5 – (0.01) = 0.49. Viendo 0.49 en el interior de la tabla del APÉNDICE A, encontramos para 0.4901 que es el más próximo a 0.49, a la izquierda 2.3 y arriba 3; luego, el valor teórico es 2.33**

**Para un nivel de significación del 5%** $Z_t$ **= 1.64**

**El área entre el centro y el valor teórico se obtiene así: 0.5 – (0.05) = 0.45. Viendo 0.45 en el interior de la tabla del APÉNDICE A, encontramos para 0.4495 que es el más próximo a 0.45, a la izquierda 1.6 y arriba 4; luego, el valor teórico es 1.64**

**Llamaremos valor calculado de la puntuación z al valor que se obtiene utilizando una de las fórmulas para z; así:**

**En el caso de una muestra A (con una distribución muestral normal de la población) :**

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

- **donde:** $\bar{x}$ **es la media aritmética muestral,** $\sigma$ **es la desviación típica poblacional,** $\mu$ **es la media poblacional o hipotetizada.**

**EJEMPLO.- En una prueba de rendimiento a 22 estudiantes se obtiene como promedio = 7.6, la desviación típica poblacional es** $\sigma$ **= 1.1. Pruebe que este promedio 7.6 difiere significativamente del promedio poblacional** $\mu$ **= 7 con un nivel de significación del 5%.**

1)Planteamiento de las hipótesis

Ho: El promedio de rendimiento del grupo es 7

Hi: El promedio de rendimiento del grupo es diferente a 7

$H_o$ : $\mu = 7$

$H_i$ : $\mu \neq 7$

2) Nivel de significación $\alpha = 0.05$

3) Criterio

Rechace la hipótesis nula si $z_c < -1.96$, o $z_c > 1.96$

(Donde 1.96 es el valor teórico de z en un ensayo a dos colas con un nivel de significación de 0.05, y es el valor calculado de z que se obtiene aplicando la fórmula: $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ )

## 4) Cálculos

Reemplazando los datos $\bar{x} = 7.6$, $\sigma = 1.1$, $\mu = 7$ y $n = 22$, en la fórmula, se obtiene:

$$z_c = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{7.6 - 7}{1.1 / \sqrt{22}} = \frac{0.6}{0.2345} = 2.56$$

## 5) Decisión

**z calculado = 2.56≥1.96 = z teórico**

**2.56 está en la zona de rechazo de la hipótesis nula, luego queda aceptada la hipótesis de investigación, esto es : "El promedio de rendimiento del grupo es diferente a 7".**

- **En el caso de dos muestras A y B de medias $\bar{x}_A$ y $\bar{x}_B$ respectivamente, y con la hipótesis nula $H_o$ : μA= μB**

**Si se conocen los valores de las varianzas poblacionales, $\sigma_A^2$, $\sigma_B^2$ se utiliza la *puntuación z*, cuyo valor se calcula con la siguiente fórmula:**

$$z_c = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \qquad \textbf{(2)}$$

**EJEMPLO**

**En una prueba de rendimiento a los grupos A con $n_A$ = 20, y B con $n_B$ = 18 se han obtenido los siguientes resultados:**

$\bar{x}_A$ **= 7.9,** $\bar{x}_B$ **= 7.1,** $\sigma_A = 0.9$, $\sigma_B = 1.3$

**Pruebe que el rendimiento de los dos grupos es significativamente diferente con un nivel de significación α = 0.05**

**1) Planteamiento de las hipótesis**

$H_o$ **: $\mu_A = \mu_B$ (El promedio de rendimiento del grupo A es igual al promedio de rendimiento del grupo B )**

$H_i$ **: $\mu_A \neq \mu_B$ (El promedio de rendimiento del grupo A es diferente al promedio de rendimiento del grupo B)**

**2) Nivel de significación**

$\alpha$ **= 0.05**

**3) Criterio**

**Rechace la $H_o$ si $z_c <$ -1.96 o $z_c >$1.96**

**Donde 1.96 es el valor teórico de z en un ensayo a dos colas con un nivel de significación de 0.05, y $z_c$ es el valor calculado de z que se obtiene aplicando la fórmula:**

$$z_c = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}}$$

**4) Cálculos**

**Reemplazando los datos**

$\bar{x}_A = 7.9 \quad \bar{x}_B = 7.1$

$\sigma_A^2 = (0.9)^2 = 0.81 \quad \sigma_B^2 = (1.3)^2 = 1.69$

$n_A = 20 \quad n_B = 18$

**en la fórmula correspondiente, se obtiene:**

$$z_c = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} = \frac{7.9 - 7.1}{\sqrt{\frac{0.81}{20} + \frac{1.69}{18}}} = \frac{0.8}{0.3666} = 2.18$$

**5) Decisión**

**Como el valor de z calculado es mayor al valor de z teórico; esto es:**

$$z_c = 2.18 \geq 1.96 = z_t$$

**2.18 está en la zona de rechazo de la hipótesis nula, luego queda aceptada la hipótesis de investigación, esto es: "El promedio de rendimiento del grupo A es diferente al promedio de rendimiento del grupo B".**

# Prueba z en software.

En términos simples, el valor *p* o p-valor ayuda a diferenciar resultados que son producto del azar del muestreo, de resultados que son estadísticamente significativos.
Por ejemplo un p-valor de 5% da una confianza del 95% de que el resultado no es por azar.

# Prueba de normalidad

**PRUEBA DE SHAPIRO-WILK**
Cuando la muestra es como máximo de tamaño 50 se puede contrastar la normalidad con la prueba de shapiro Shapiro-Wilk

**PRUEBA DE KOLMOGOROV-SMIRNOV**

https://www.socscistatistics.com/tests/kolmogorov/default.aspx

# Prueba z una muestra

https://mathcracker.com/es/prueba-z-para-una-media#results

http://www.learningaboutelectronics.com/Articulos/Calculadora-de-prueba-de-hipotesis-estadistica.php#respuesta