

~~HENRY~~



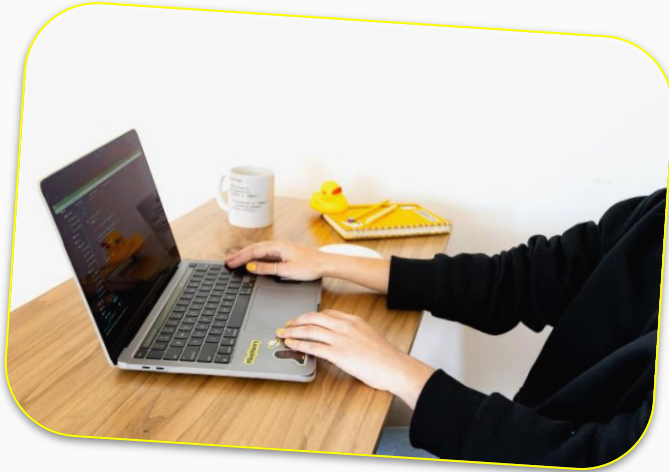
Test Estadísticos **Clase 9**

Data Science





Agenda



- **Tests** Estadísticos
- **Prueba** de una hipótesis de investigación
- **Prueba** de la validez de una afirmación
- **Prueba** en situaciones de toma de decisión
- **Formas** para las hipótesis nula y alternativa
- **Errores** tipo I y II
- **Pruebas** de hipótesis para la media poblacional



OBJETIVOS DE LA CLASE

Al finalizar esta lecture estarás en la capacidad de...

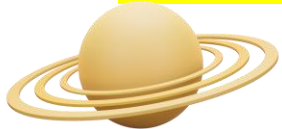
→ **Usar** los conceptos de Tests Estadísticos y Pruebas de Hipótesis.

Test



Por lo general, en la práctica, se tienen que tomar decisiones sobre poblaciones, partiendo de la información muestral de las mismas.

Tales decisiones se llaman, **decisiones estadísticas.**





Hipótesis Estadísticas

Para la toma de decisiones se hacen supuestos que son formulados respecto del valor de algún parámetro, que pueden ser o no ciertos. En general, lo son sobre las **distribuciones de probabilidad** de las poblaciones.



Pruebas de hipótesis



¿Qué son?

En el procedimiento de test estadísticos, se utilizan las denominadas pruebas de hipótesis y en ellas se usan datos de una muestra para probar dos afirmaciones contrarias indicadas por:

HA: Hipótesis alternativa

H0: Hipótesis nula



Hipótesis en Investigaciones



La **puntuación promedio** de un Henry Challenge es de 78, Henry determina que a través de métodos innovadores puede aumentar esa media.



En este caso, se establece un **grupo de investigación** que busca evidencias para concluir que el nuevo sistema aumenta la media del rendimiento.



La **hipótesis de investigación** es, entonces, que el nuevo sistema proporciona un rendimiento medio mayor.



Es decir, $\mu > 78$.

Como lineamiento general, una hipótesis de investigación se debe plantear como hipótesis alternativa.

- **$H_0: \mu \leq 78$.**
- **$H_A: \mu > 78$.**

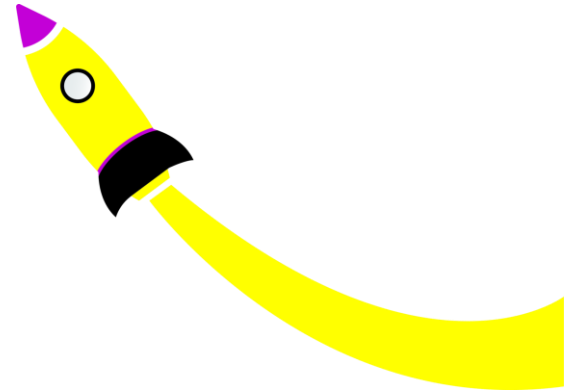


Hipótesis en afirmación



Cuando lo que realizamos es una afirmación, en este caso sería que quienes rinden un HC obtienen por lo menos 78 puntos en promedio, tratamos de corroborar que esa afirmación es correcta.

- $H_0: \mu \geq 78.$
- $H_a: \mu < 78.$





En toda situación en la que se desee probar la validez de una afirmación, la hipótesis nula se suele basar en la suposición de que la afirmación sea verdadera. Entonces, la **hipótesis alternativa** se formula de manera que rechazar **H₀** proporcione la evidencia estadística de que la suposición establecida es incorrecta.



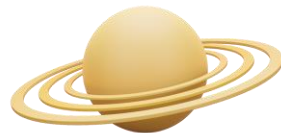


Hipótesis en Decisiones / Alternativa

Existen además otras formas de realizar el planteo de H_0 y H_a , como cuando se debe tomar una decisión.

Por ejemplo controlar la calidad de un determinado repuesto en donde debe medir obligatoriamente 10 cm.

- **H_0** : $\mu = 10$.
- **H_A** : $\mu \neq 10$.



Lo que determina solo dos alternativas.



Resumen

$$H_0: \mu \geq \mu_0$$

$$H_0: \mu \leq \mu_0$$

$$H_0: \mu = \mu_0$$

$$H_a: \mu < \mu_0$$

$$H_a: \mu > \mu_0$$

$$H_a: \mu \neq \mu_0$$

Metodología

01

Formular la hipótesis nula

02

Formular la hipótesis alternativa

03

Especificar el nivel de significación.

Metodología

04

Determinar el tamaño de la muestra.

05

Determinar el estadístico de prueba.

06

Establecer los valores críticos que dividen las zonas de rechazo y de no rechazo.

Metodología

07

Obtener los datos y calcular los estadísticos.

08

Determinar si el estadístico de prueba ha caído en la región de rechazo o en la de no rechazo.

09

Determinar la decisión estadística.

Metodología

10

Expresar la decisión estadística en términos del problema.



Nivel de significación



La **distribución muestral** del estadístico analizado, suele seguir una distribución estadística conocida, como la distribución normal estandarizada, la distribución t o la distribución chi cuadrado, éstas se utilizan como ayuda para determinar si la hipótesis nula es cierta.





Existen dos tipos de errores

Error de tipo I: probabilidad de que se rechace la hipótesis nula cuando es verdadera.



Error de tipo II: Es la probabilidad de aceptar la hipótesis nula cuando es falsa





Tamaño de la muestra



El tamaño de la muestra se determina al tomar en cuenta la importancia de **Error de Tipo I** y **Error de Tipo II** y al considerar las restricciones presupuestarias al efectuar el estudio.

$$n = \frac{(z_0 - z_1)^2 \sigma^2}{(\mu_0 - \mu_1)}$$





Tamaño de la muestra



Generalmente las muestras grandes, permiten detectar incluso diferencias pequeñas entre los valores hipotéticos los parámetros poblacionales. Para un nivel de Error I dado, aumentar el tamaño de la muestra reducirá Error II y así se incrementará el poder de la prueba para detectar que la hipótesis nula es falsa.



Estadístico de prueba.



Una vez definidas las **hipótesis nula** y **alternativa**, y el tamaño de la muestra se puede establecer la distribución a utilizar: normal, t-student ó chi cuadrado.





Zonas de rechazo y de no rechazo



Se representa como un área (como toda probabilidad en una función de densidad), que se ubica a la derecha, a la izquierda o a ambos lados (en este caso, con la mitad del Error de Tipo I en cada lado) según como se haya definido la Hipótesis alternativa.

Obtener los datos y calcular los estadísticos



Este paso está reservado a la efectiva realización de la **investigación muestral**. Es decir que en este momento es cuando se realiza el estudio tendiente a obtener los valores muestrales y calcular los estadísticos.



Distribución Normal

Se asignan probabilidades a los eventos después del experimento. Es decir, la **asignación de probabilidades** está basada en el conocimiento de la ocurrencia de eventos que están en dependencia con el evento de estudio.

$$z_1 = \frac{\bar{x} - \mu_0}{\frac{S_x}{\sqrt{n}}} \quad z_1 = \frac{S_x - \sigma_x}{\frac{S_x}{\sqrt{2n}}}$$



t-Student

$$t_1 = \frac{\bar{x} - \mu_0}{\frac{S_x}{\sqrt{n-1}}}$$



Chi Cuadrado

$$\frac{n S_x^2}{\sigma_x^2} = \frac{\sum (x_i - \bar{x})^2}{\sigma_x^2} = \chi_{vgl}^2$$



Determinar la decisión estadística

Se determina la decisión de la prueba de hipótesis

- si $z_1 > z_c$ entonces z_1 cae en la "**zona de rechazo**" y se considera que las diferencias entre z_1 y z_c son significativas entonces Rechazo la Hipótesis nula.
- si $z_1 \leq z_c$ entonces z_1 cae en la "**zona de no rechazo**" y se considera que las diferencias entre z_1 y z_c no son significativas entonces No Rechazo la Hipótesis nula.



Pruebas de una cola

Cuando nos encontramos frente a una prueba de hipótesis del tipo: " $H_0: \mu \leq x$ " y " $H_1: \mu > x$ " o " $H_0: \mu \geq x$ " y " $H_1: \mu < x$ ", la denominamos **prueba de una cola**.

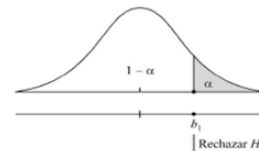
Tipos de pruebas de hipótesis

- Prueba unilateral

Cola a la derecha

$$H_0: \theta \leq \theta_0$$

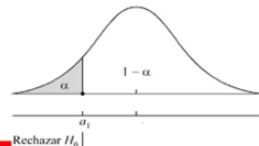
$$H_1: \theta > \theta_0$$



Cola a la izquierda

$$H_0: \theta \geq \theta_0$$

$$H_1: \theta < \theta_0$$



Pruebas de una cola



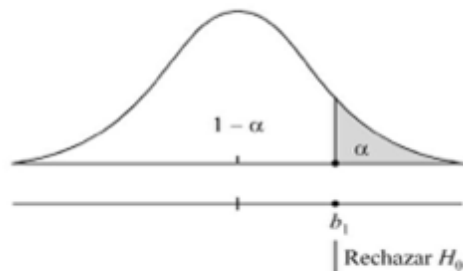
Tipos de pruebas de hipótesis

- Prueba unilateral

Cola a la derecha

$$H_0 : \theta \leq \theta_0$$

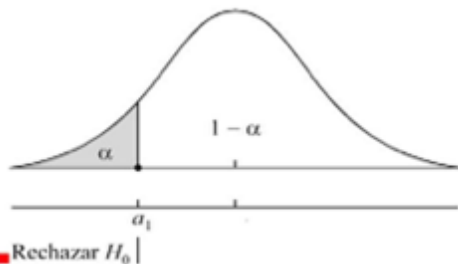
$$H_1 : \theta > \theta_0$$



Cola a la izquierda

$$H_0 : \theta \geq \theta_0$$

$$H_1 : \theta < \theta_0$$





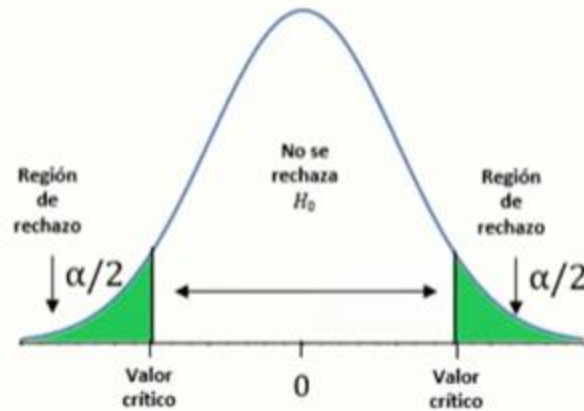
Pruebas de dos colas

Cuando nos encontramos frente a una prueba de hipótesis del tipo: " $H_0: \mu = x$ " y " $H_1: \mu \neq 1x$ ", la denominamos **prueba de dos colas.**



Pruebas de dos colas

Prueba de dos colas:



$$H_0: \mu = \mu_0$$
$$H_1: \mu \neq \mu_0$$

HENRY

