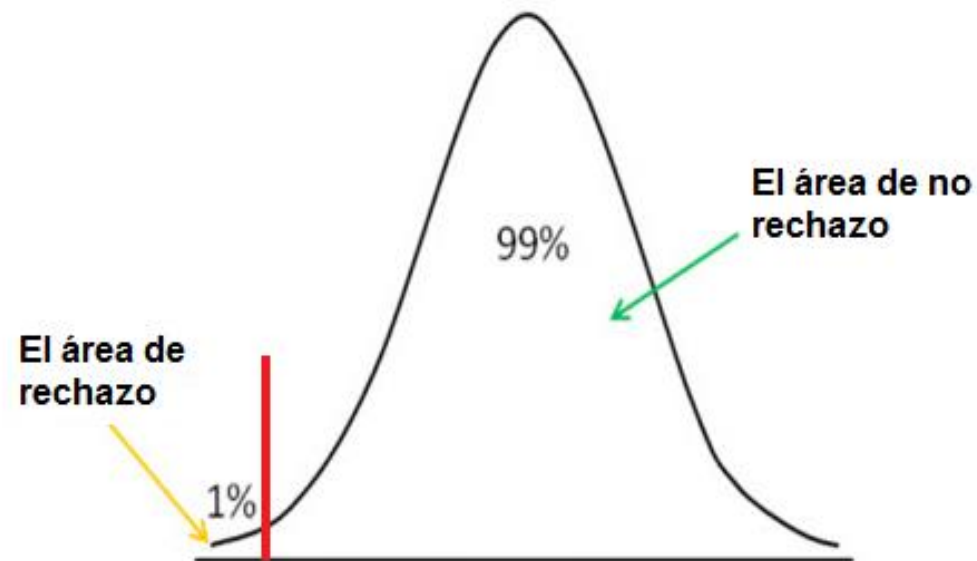


Prueba de Hipótesis

¿Qué es una hipótesis?

En estadística, es una afirmación respecto a alguna característica de una población. Donde se comparan las predicciones con la realidad que observamos. Si existe coincidencia entonces se acepta la hipótesis. En caso contrario, la rechaza.



Pasos para formular una prueba de hipótesis

Las hipótesis son afirmaciones y se puede usar estadística para probarlas o refutarlas. Donde se requiere probar las afirmaciones y verificarlas. Los pasos que se emplean para formular una prueba de hipótesis son los siguientes:



1. Definir la hipótesis

En una prueba de hipótesis se supondrá siempre una pregunta a la que se esta buscando responder. La formulación de la pregunta sigue una estructura concreta y se construye a través de dos elementos. Una hipótesis nula (H_0) y una hipótesis alternativa (H_1) .

- ❖ **Hipótesis nula (H_0)** : Es la hipótesis que sostiene que la suposición inicial que se tiene respecto a un parámetro. La hipótesis nula es aquella hipótesis que se pretende rechazar. Normalmente va asociada con algún signo comparativo (= , \leq , \geq)
- ❖ **Hipótesis alternativa (H_1)** : Es la hipótesis de la investigación que se pretende probar que es cierta. Es una suposición previa que tiene el investigador y para intentar demostrar que es verdadera llevará a cabo la prueba de hipótesis. Normalmente se asocia con el signo comparativo contrario al de la hipótesis nula(\neq , $<$, $>$)

Formular Hipótesis

La hipótesis nula es: La media de la población de edades de todos los estudiante de 4Geeks es igual a 25 años. Formalmente, esto se escribe: $H_0: \mu = 25$

La hipótesis alternativas es:

La media poblacional es menor que el objetivo. unilateral: $\mu < 25$

La media poblacional es mayor que el objetivo. unilateral: $\mu > 25$

La media poblacional difiere del objetivo. a dos lados: $\mu \neq 25$

Como se tiene que asegurar que las edades no sean mayores o menores a 25 años, se elige la hipótesis alternativa de dos lados, la cual establece que la media de la población de todas las edades no es igual 25 años. Formalmente se escribe como $H_1: \mu \neq 25$

2. Elegir el nivel de significancia

El nivel de significancia (α , alfa) hace referencia al grado de error que se admite para aceptar o rechazar la hipótesis. Es la probabilidad de rechazar una hipótesis que en realidad es verdadera.

Si el nivel de significación es $\alpha=0.05$, significa que la probabilidad de rechazar una hipótesis cuando es cierta es del 5%. En otras palabras, la probabilidad de hacer la estimación de un parámetro estadístico y equivocarse con un error mayor al margen de error es del 5%.

$$\alpha = 0,05$$

$$1 - \alpha = 0,95$$

3. Verificar la hipótesis

Una vez planteada la hipótesis nula, la hipótesis alternativa y el nivel de significación, se debe elegir el estadístico adecuado para la prueba de hipótesis. Hay muchas formas de hacerlo, pero las más utilizadas en el mundo real son el contraste de hipótesis para la media y el contraste de hipótesis para la proporción.

Para verificar la hipótesis, primero se calcula el Z-score, que es el valor que describe la posición de un valor individual dentro de una distribución de datos. El Z-score indica cuántas desviaciones estándar está un valor de la media de la distribución.

Prueba de hipótesis para la media

Es la comparación de dos medias, la que se calcula y la poblacional para compararlas, se necesita previamente conocer la media y la desviación estándar de la población. La fórmula que caracteriza este cálculo es:

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

donde \bar{x} y σ son la media y la desviación estándar de la población, μ es la media que se calcula de la muestra y n es el tamaño de la muestra.

Prueba de hipótesis para la proporción

Es la comparación de dos proporciones, la que se calcula y la poblacional para compararlas, se necesita previamente conocer la proporción de la población. La fórmula que caracteriza este cálculo es:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

donde p es la proporción propuesta, de la población, \hat{p} es la proporción que se calcula de la muestra, n es el tamaño de la muestra y la raíz cuadrada incluyendo lo de dentro es la desviación estándar de la proporción.

Valor - p

A partir del Z-score, se calcula el p-valor, que es una medida que se usa en estadística para ayudar a decidir si los resultados de una prueba son significativos o no. Es la probabilidad de obtener los resultados del tu experimento entorno a la veracidad de la hipótesis nula.

Si el p-valor es muy pequeño, significa que es muy poco probable que la hipótesis nula sea cierta. Es decir Si el p-valor es menor al el nivel de significancia (conocido como α o alfa), entonces se puede rechazar la hipótesis nula. Donde el resultado es estadísticamente significativo. En cambio, si el p-valor es grande mayor a al nivel de significancia, no se puede rechazar la hipótesis nula.

4. Decisión. ¿Se acepta la hipótesis?

Dependiendo del objeto de la hipótesis y el proceso de verificación, se debe analizar el resultado y si se encuentra dentro del intervalo de confianza o no. El el objeto de estudio es la hipótesis alternativa y en función del resultado se puede aceptar o rechazar la hipótesis nula.

Intervalo de confianza: Par de números entre los cuales se estima que estará cierto valor desconocido respecto a un parámetro poblacional con un determinado nivel de confianza

Otras técnicas para hacer hipótesis.

- **T-Test:** Permite comparar dos grupos o categorías numéricas con un tamaño muestral reducido. Este caso es el contraste de hipótesis para la media **en el caso de que se desconozca la varianza con muestra reducida**. Sin embargo, también se puede usar cuando se conoce.
- **Z-Test:** Permite comparar dos grupos o categorías numéricas con un tamaño muestral amplio. Este caso es el contraste de hipótesis para la media **en el caso de que se desconozca la varianza con muestra amplia**. Sin embargo, también se puede usar cuando se conoce.
- **ANOVA Test:** Permite comparar dos o más grupos o categorías numéricas.
- **Chi-Cuadrado Test:** Analiza la relación entre dos variables categóricas.
- **Correlation Test:** Analiza la relación entre dos variables numéricas.