Hipótesis Estadística

Índice

Objetivos

- Objetivo 1: Analizar el mínimo nivel de incertidumbre.
- Objetivo 2: Construir estadísticas de pruebas pertenecientes a una determinada distribución para poder concluir si se rechaza o no cierta hipótesis de interés.

1 Introducción

En esta unidad de aprendizaje analizaremos uno de los objetivos fundamentales en un estudio estadístico para determinar el **mínimo nivel de incertidumbre** a través de los conceptos básicos y aplicación de pruebas de hipótesis.

Esto se hace con el fin de encontrar herramientas útiles a la hora de tomar decisiones acerca de una **sospecha o duda razonable** que se tenga en un estudio de investigación. Nuestro interés se centrará entonces en la **construcción de estadísticas de prueba pertenecientes a una determinada distribución** para luego compararla con valores de la tabla en una distribución conocida y mirar si **se rechaza o no cierta hipótesis de interés**.

2 Conceptos Básicos

"Cuando se tengan experimentos con múltiples resultados basados en eventos aleatorios y la finalidad sea la toma de una decisión, conviene el planteamiento de una hipótesis"

En casos donde se tengan experimentos con múltiples resultados basados en eventos aleatorios y la finalidad sea la toma de una decisión, es de mayor interés el buen **planteamiento de una hipótesis**. Con base a esto, la estrategia natural es decidirse por uno de los experimentos en cuestión. Así el establecimiento de la regla de decisión delimitara la **región de rechazo** y con ello las **probabilidades de errores** a la hora de la toma de decisiones.

A continuación veremos los **conceptos básicos** involucrados en el estudio de las **pruebas de hipótesis** para entender mejor la aplicación de esta técnica de inferencia estadística.

2.1 Hipótesis Estadística

"Se trata de una afirmación o conjetura acerca de una distribución de una o más variables aleatorias"

Una <u>hipótesis estadística</u> es una **afirmación o conjetura** acerca de una distribución de una o más variables aleatorias.
Ejemplo 1: Si X sigue una distribución exponencial de parámetro λ una hipótesis podría ser la afirmación "$\lambda = 0.3$", análogamente si X sigue una distribución $N(\mu, \sigma^2)$ una hipótesis podría ser la afirmación "$\mu > 1$".

2.2 Tipos de Hipótesis

2.2.1 Hipótesis simple

Una <u>hipótesis</u> se dice <u>simple</u> si se especifica por completo la probabilidad en cuestión.

2.2.2 Hipótesis compuesta

Una hipótesis se dice compuesta si no se especifica por completo la probabilidad en cuestión.

Ejemplo 2: Si X sigue una distribución exponencial de parámetro λ una hipótesis podría ser la afirmación "$\lambda = 0.3$" es una hipótesis simple, en contraste si X sigue una distribución $N(\mu, \sigma^2)$ una hipótesis podría ser la afirmación "$\mu > 1$" se trata de una hipótesis compuesta.

"Esquema para la teoría de las pruebas de hipótesis"

En general para la **teoría de las pruebas de hipótesis** se maneja el siguiente esquema:

H_0: Hipótesis nula vs H_1: Hipótesis Alternativa

En donde las **hipótesis nula y alternativa** pueden ser simples o compuestas.

2.3 Prueba de Hipótesis

"La prueba de hipótesis ayuda a decidir si no se rechaza la hipótesis nula o si se rechaza en favor de la hipótesis alternativa"

Una prueba de hipótesis es una **regla para decidir** si no se rechaza la hipótesis nula o se rechaza en favor de la hipótesis alternativa.

Nótese que no está bien afirmar "se acepta la hipótesis nula" debido a que se hace es una conclusión con base a la información que se extrae de la muestra, en consecuencia lo que se tiene es que la muestra seleccionada no arroja información suficiente para rechazar la hipótesis nula, teniendo en cuenta que todo depende de la **calidad de la muestra** y de los **errores de muestreo**.

2.4 Tipos de Errores

Al realizar una **prueba de hipótesis** se pueden cometer errores.

- Al rechazo de la hipótesis nula cuando ésta es verdadera se le conoce como *error tipo I*, y a la probabilidad de cometer este primer tipo de error se le denota por la letra α.
- En cambio, al no rechazo de la hipótesis nula cuando ésta es falsa recibe el nombre de *error tipo II*, y a la probabilidad de cometer este segundo tipo de error se le denota por la letra β.

Estas definiciones de errores se resumen en la siguiente tabla:

	H_0 Cierta	H_0 Falsa
Rechazar H_0	Error tipo I con probabilidad α	Decisión correcta
No rechazar H_0	Decisión correcta	Error tipo II con probabilidad β

Tabla 1: *Tipos de errores*

2.5 Región Crítica y Nivel de Significación

"Región critica es la región de rechazo de H_0 y el tamaño de la región critica es la probabilidad de cometer el error tipo I"

Se le llama región critica a la región de rechazo de H_0, y a la probabilidad de cometer el *error tipo I*, esto es α, se le llama tamaño de la región critica. A esta probabilidad se le conoce también con el nombre de nivel de significancia.

3 Pruebas para la Media de una Distribución Normal

Sea $\boldsymbol{x_1, x_2, \ldots, x_n}$ una muestra aleatoria de población media $\boldsymbol{\mu}$ desconocida y varianza $\boldsymbol{\sigma^2}$ conocida. Sabemos que $\boldsymbol{\bar{x}}$ tiene distribución $\boldsymbol{N(\mu, \frac{\sigma^2}{\sqrt{n}})}$ por tanto:

$$\frac{\bar{x}-\mu}{\sigma^2/\sqrt{n}} \sim N(0,1)$$

Sea $\boldsymbol{\mu_0}$ un número real particular. Deseamos probar las hipótesis $\boldsymbol{H_0: \mu = \mu_0}$ contra $\boldsymbol{H_1: \mu \neq \mu_0}$ El problema es encontrar una regla para decidir cuándo rechazar $\boldsymbol{H_0}$ en favor de $\boldsymbol{H_1}$ con base en los datos de la muestra $\boldsymbol{x_1, x_2, \ldots, x_n}$. Cuando $\boldsymbol{H_0}$ es cierta, esto es, cuando $\boldsymbol{\mu}$ es efectivamente $\boldsymbol{\mu_0}$, tenemos que $\boldsymbol{X \sim N(\mu_0, \frac{\sigma^2}{\sqrt{n}})}$ y por lo tanto:

$$\frac{\bar{x}-\mu_0}{\sigma^2/\sqrt{n}} \sim N(0,1)$$

La estadística $\boldsymbol{Z = \frac{\bar{x}-\mu_0}{\sigma^2/\sqrt{n}}}$ es una medida natural de la distancia entre $\boldsymbol{\bar{x}}$,un estimador de $\boldsymbol{\mu}$, y su valor esperado $\boldsymbol{\mu_0}$ cuando $\boldsymbol{H_0}$ es cierta. Es entonces razonable rechazar $\boldsymbol{H_0}$ cuando la variable $\boldsymbol{Z}$ sea grande. Es por ello que tomamos como criterio de decisión rechazar $\boldsymbol{H_0}$ cuando $\boldsymbol{|Z| \geq k}$, para cierta constante $\boldsymbol{k}$.

¿Cómo encontramos el número $\boldsymbol{k}$? En una tabla de la distribución normal podemos encontrar un valor $\boldsymbol{z_{\alpha/2}}$ tal que $\boldsymbol{P(|Z| \geq z_{\alpha/2}) = \alpha}$, en donde $\boldsymbol{\alpha}$ lo determina la persona que lleva a cabo la prueba de hipótesis, típicamente $\boldsymbol{\alpha = 0.1}$. Véase el Gráfico 1. Este valor $\boldsymbol{z_{\alpha/2}}$ es precisamente la constante $\boldsymbol{k}$ buscada pues con ello se logra que la región de rechazo sea de tamaño $\boldsymbol{\alpha}$.

A la variable aleatoria $\boldsymbol{Z}$ se le llama la estadística de la prueba, y la prueba se denomina prueba de dos colas pues la región de rechazo consta de las dos colas de la distribución normal que se muestran en el Gráfico 1.

Llevar a cabo esta prueba de hipótesis consiste en usar los datos de la muestra para encontrar el valor de $\boldsymbol{Z}$, $\boldsymbol{|Z| \geq z_{\alpha/2}}$, entonces se rechaza $\boldsymbol{H_0}$, en caso contrario no se rechaza $\boldsymbol{H_0}$. Similarmente se definen las pruebas a cola inferior y superior para valores extremos.

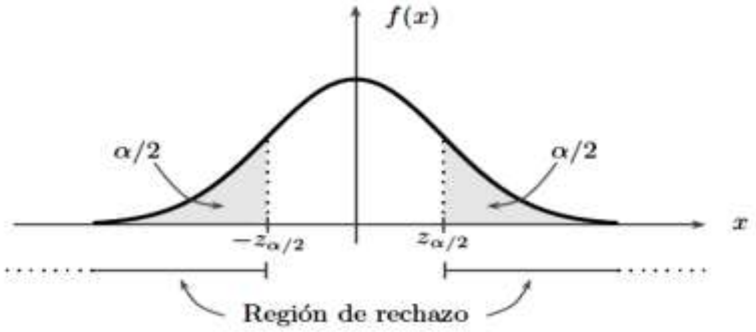

Gráfico 1: Región de rechazo a dos colas

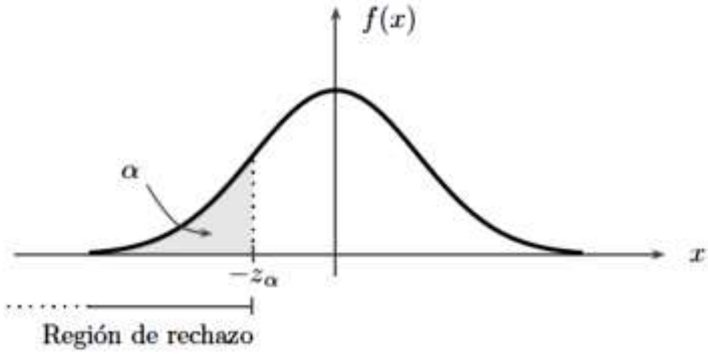

Gráfico 2: *Región de rechazo a cola inferior*

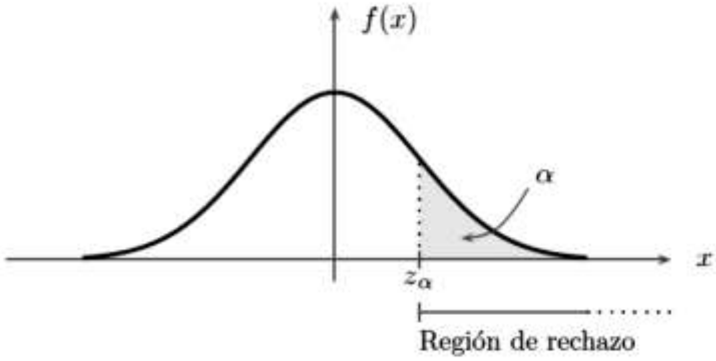

Gráfico 3: *Región de rechazo a cola superior*

3.1 Prueba de Dos Colas para μ

"Suponemos dos zonas de rechazo para probar esta hipótesis"

Para probar esta hipótesis en primera instancia se tiene en cuenta que existen dos zonas de rechazo en ambas colas como se muestra en el Gráfico 1 y en segunda instancia se calcula el estadístico de prueba Z y se compara con los valores críticos Z en la tabla, así tenemos que el valor en mención viene dado por:

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

En donde $\bar{x}$ es el valor de la media muestral, μ_H es el valor de la media poblacional bajo la hipótesis nula y $\sigma^2/\sqrt{n}$ es el error estándar de la distribución muestral.
Cuando se desconoce σ se utiliza su respectiva estimación, la desviación estándar muestral s y Z se vuelve:

$$Z = \frac{\bar{x} - \mu_H}{s/\sqrt{n}}$$

Ejemplo 3: Se supone que una empresa embotelladora de bebidas gaseosas de que la media poblacional es de 16 onzas y seleccionan un nivel de significancia del 5%. Debido al planteamiento del problema el conjunto de hipótesis queda como sigue:
$H_0: \mu = 16$
$H_1: \mu \neq 16$
Si la empresa embotelladora toma una muestra de $n = 50$ botellas con una media de $\bar{x} = 16.357$ onzas y una desviación estándar de $s = 0.866$ onzas, tenemos que

$$Z = \frac{16.357 - 16}{\frac{0.866}{\sqrt{50}}} = 2.91$$

Ahora comparando Z con los valores críticos de z de la tabla que son ± 1.96. La regla de decisión sería: no se rechaza la hipótesis nula sí $-1.96 \leq Z \leq 1.96$. Se rechaza si $Z < -1.96$ o $Z > 1.96$.

Luego como $Z = 2.91 > 1.96$ se rechaza la hipótesis nula a un nivel de significancia del 5% en favor de la hipótesis alternativa.

3.2 Pruebas de Una Cola para μ

En contraste con el anterior caso, en este solo se está interesado en una de las dos colas de la distribución. Como se muestra en los Gráficos 2 y 3 respectivamente.
Ejemplo 4: En una reunión informativa para una oficina corporativa, el gerente de un importante hotel, reporto que el número promedio de habitaciones alquiladas por noche es de por lo menos 212. Uno de los funcionarios corporativos cree que esta cifra puede estar sobreestimada. Una muestra de 150 noches produce una media de 201.3 habitaciones y una desviación estándar de 45.5 habitaciones. Si estos resultados sugieren que el gerente ha inflado su reporte, será amonestado severamente. A un nivel del 1% ¿Cuál es el destino del gerente?
La afirmación del gerente de que $\mu \geq 212$ lleva el signo igual y por tanto se toma la siguiente hipótesis nula

$$H_0: \mu \geq 212 \text{ vs } H_1: \mu < 212$$

Luego

$$Z = \frac{201.3 - 212}{\frac{45.5}{\sqrt{150}}} = -2.88$$

La regla de decisión es: No rechazar H_0 sí $Z \geq -2.33$. Rechazar sí $Z < -2.33$. Entonces el valor $Z = -2.88$ claramente está en la zona de rechazo lo que indica que el gerente podría estar en serios problemas.

4 Resumen

- En casos donde se tengan experimentos con múltiples resultados basados en eventos aleatorios y la finalidad sea la toma de una decisión, es de mayor interés, el buen planteamiento de una hipótesis.
- Los conceptos básicos involucrados en el estudio de las pruebas de hipótesis son: la hipótesis estadística, los tipos de hipótesis, la prueba de hipótesis, los tipos de errores y la región crítica y nivel de significancia.

5 Referencias Bibliográficas

- Miranda, I. E., Palacín, F., Sánchez, M. L., Márquez, M., Chía, A. R., Navas, A. S., y otros. (3ra. Edición 2006). Estadística Descriptiva y Probabilidad. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.
- Montgomery, D., & R., R. (2da. Edición 2008). Probabilidad y Estadística Aplicada a la Ingeniería. México: Limusa Wiley.
- Walpole, R., Myers, R., & Myers, S. y. (2007). Probabilidad y Estadística para Ingeniería y Ciencias. México: Pearson.