Conferencia 3. Prueba de Hipótesis para una Población.

Tema 2. Inferencia Estadística.

Sumario:

- Hipótesis Estadística.
- Nivel de Significación.
- Criterio de aceptación/rechazo.
- Tipo de error. Ejemplos
- Prueba para la media de la distribución normal con varianza conocida y desconocida.
- Prueba para la proporción en la distribución binomial.

Bibliografía:

- Introductory Statistics Prem S. Mann, Christopher Jay Lacke, 2010 (Capítulo 9)
- Introductory Statistics. Barbara Illowsky, Susan Dean 2018 (Capítulo 9)

Introducción

Las mediciones que hemos estudiado no se realizan sobre poblaciones, sino sobre muestras extraídas aleatoriamente de las poblaciones. Por lo que siempre se corre el riesgo de tomar una decisión errónea pues las características de la muestra pueden diferir de aquellas que posee la población. Este riesgo debe hacerse cuantificable y para ello se sigue una teoría de trabajo llamada comprobación de hipótesis estadísticas.

Una hipótesis se contrasta comparando sus predicciones con la realidad: si coinciden, dentro del margen de error admisible, mantendremos la hipótesis, en caso contrario, la rechazaremos y buscaremos nuevas hipótesis capaces de explicar los datos observados. Para comprender mejor la idea detrás de las pruebas de hipótesis veamos primero un ejemplo.

Hipótesis Estadística.

Ejemplo 1: En la confección del examen de ingreso de matemática, los examinadores plantean que los estudiantes al terminar el grado 12 deben tener conocimientos básicos de Matemática. Después de un minucioso análisis junto con un estadístico se llega a la conclusión de que la probabilidad de que un estudiante de 12 grado tenga conocimientos básicos de matemática es de 0.8.

En este ejemplo los examinadores plantean una hipótesis, "*los estudiantes al terminar el grado 12 deben tener conocimientos básicos de Matemática*". Luego lo expresan en términos de la probabilidad de que el estudiante tenga los conocimientos, es decir formularon una hipótesis a partir del parámetro p de una distribución Bernoulli. Este ejemplo puede ayudar a entender las siguientes definiciones.

Definición 1 Hipótesis Estadística: Una hipótesis estadística es un supuesto acerca de la distribución de probabilidad de una o más variables aleatorias, o más particularmente acerca de uno o varios parámetros que caracterizan a dicha distribución.

Como una hipótesis es solo un supuesto, es necesario verificarlo. Para que los examinadores puedan verificar su hipótesis deben seleccionar una muestra de la población de estudiantes de 12 grado, aplicarles un examen de matemática que cubra los conocimientos básicos para ver cuántos son capaces de resolverlos y cuántos no. No obstante, aún después de realizar el experimento, los examinadores no pueden afirmar si su hipótesis es cierta o no, pues solo han examinado una muestra de la población. Es necesario aplicar algún procedimiento que les permita utilizar lo que han examinado en la muestra para decidir si su hipótesis es o no cierta. Las dócimas de hipótesis o pruebas de hipótesis son el procedimiento de Inferencia Estadística que brinda los métodos necesarios para decidir acerca de la validez de una hipótesis estadística, en base a los resultados observados en una muestra. Este problema se define de la siguiente forma.

Definición 2, Problema de Prueba de Hipótesis.
Sea una variable aleatoria $X \sim D(\theta)$ donde $D(\theta)$ es la distribución de X que depende de un parámetro θ. En el caso más extremo no conocemos nada de $D(\theta)$ y en situaciones más moderadas conocemos que pertenece a una distribución conocida, pero desconocemos el valor del parámetro θ.

En todo problema de prueba de hipótesis se tiene una partición del conjunto de valores de θ en dos subconjuntos disjuntos: $H = H_0 \cup H_1$el problema es decidir si θ pertenece a H_0 o a H_1.

En el ejemplo que venimos siguiendo al analizar la muestra de estudiantes de 12 grado tenemos que existen unos que tiene los conocimientos y otros no; por tanto, estamos en presencia de la variable aleatoria: X: Sabe matemática básica que sigue una distribución binomial $X \sim Bin(n, p)$, donde n es el tamaño de la muestra analizada y el parámetro p es la probabilidad de que un estudiante de 12 grado sepa matemática básica. Por tanto $p \in [0,1]$, teniendo en cuenta que la hipótesis de los examinadores es que esa probabilidad sea mayor que 0.8, la partición de $[0, 1]$ sería:

$$H_0 = [0, 0.8] \; y \; H_1 = \;]0.8]$$

ya que el problema es determinar si:

$$p \leq \; 0.8 \; o \; p > \; 0.8$$

Luego, el problema de prueba de hipótesis se plantea de la siguiente manera,

$$H_0 \colon \theta \; \in \; H_0 \; vs \; H_1 \colon \theta \; \in \; H_1$$

Donde,

- H_0: Se denomina **hipótesis nula**. H_0 Representa la hipótesis que se mantiene a no ser que los datos indiquen falsedad. Esta hipótesis nunca se considera probada, aunque puede ser rechazada por los datos.
- H_1: Se denomina **hipótesis alternativa**. En esta hipótesis se plantea el interés del investigador y nunca se incluye la igualdad.

Regla de Decisión

Se desea disponer de una regla que permita decidir cuál hipótesis se rechaza o se acepta:

Definición 3 (Regla de decisión)
Una regla de decisión es un procedimiento probabilístico que depende de observaciones realizadas (resultados) sobre experimentos estrechamente relacionados con el problema en estudio, el cual permite decidir si se rechaza o no una hipótesis previamente formulada.

Una regla de decisión puede ser un procedimiento tan simple como lanzar una moneda para decidir que hipótesis es la correcta o puede ser más complicada e implicar la realización de cálculos. Digamos en el caso del ejemplo tenemos dos reglas de decisión:

$$\varphi_1(X) = \begin{cases} aceptar\ H_0 & si\ sale\ cara \\ rechazar\ H_0 & si\ sale\ cruz \end{cases} \qquad \varphi_2(X) = \begin{cases} aceptar\ H_0 & si\ T(X) \leq 500 \\ rechazar\ H_0 & si\ T(X) > 500 \end{cases}$$

Donde $T(X)$ es la cantidad de estudiantes de 12 grado que respondió correctamente el examen de una muestra de 1000 estudiantes. Se pueden sustituir los valores $aceptar\ H_0$ y $rechazar\ H_0$ por los valores 0 y 1 respectivamente.
Si bien es muy importante definir el experimento a realizar para probar la hipótesis, también es importante definir una buena regla de decisión que se base en este. Digamos que en vez de tener o no conocimientos acerca de una asignatura lo que quisiéramos saber es la eficacia de un medicamento, podríamos usar ambas reglas de decisiones con algunas modificaciones (sea $T(X)$ cantidad de pacientes que respondió favorablemente al tratamiento). Sin dudas confiaríamos más en la decisión si la regla utilizada fuese $\varphi_2(X)$ porque para esta regla nos basamos en resultados observados en la muestra, mientras que en $\varphi_1(X)$ utilizamos un procedimiento totalmente aleatorio.

Región Crítica

Siendo la regla de decisión un instrumento para decidir en base a las observaciones, si se rechaza o no la hipótesis nula, en ella deben quedar perfectamente especificados para cuales valores de las observaciones rechazamos la hipótesis nula y para cuáles no. De esta forma el espacio muestral queda dividido en dos regiones; una en la que se rechaza H_0 y otra que se acepta H_0.

Definición 4 (Región crítica)
La región crítica asociada a una regla de decisión es el conjunto de valores de la muestra para los cuales se rechaza la hipótesis nula.
Digamos que para el ejemplo 1 se tiene la siguiente regla de decisión:

$$\varphi_2(X) = \begin{cases} 0 & \bar{p} < 0.8 \\ 1 & \bar{p} \geq 0.8 \end{cases}$$

Donde $\bar{p} = \frac{m}{n}$ es la proporción de estudiantes en la muestra que aprobaron el examen (m cantidad de aprobados en una muestra de estudiantes de tamaño n)
La región crítica asociada a esta regla es el conjunto de valores posibles de $\bar{p}$ que son mayores que 0.8 o sea

$$RC = \{\bar{p} :\ \bar{p} \in [0.8, 1)\}$$

El investigador siempre debe tener presente que la decisión que toma se basa solo en la investigación que, de una muestra o subconjunto de la población, imprimiéndole a los resultados un carácter probabilístico, en el sentido de que nunca sabrá si la respuesta asumida como la correcta es realmente la correcta.

Errores en una prueba de hipótesis

En términos del ejemplo 1, puede que aceptemos H_1 sin que esta sea cierta y los estudiantes en realidad no tengan conocimientos básicos de matemática. Pero podemos aceptar H_0 cuando los estudiantes tienen los conocimientos. Esto es lo que definimos como Error de tipo 1 y Error de tipo 2.
Error de tipo I: Es el error que se comete cuando se rechaza H_0 siendo cierta.
Error de tipo II: Es el error que se comete cuando se acepta H_0 siendo falsa.
Luego, una buena regla de decisión debe tratar de minimizar la probabilidad de cometer ambos tipos de errores.

Supongamos que estamos en un juicio y tenemos un acusado y nos interesa saber si es culpable, por tanto, planteamos las hipótesis siguientes:

H_0: acusado es inocente
H_1: acusado es culpable

Está claro que siempre es mejor dejar a un presunto culpable libre, que condenar a un inocente. El error de tipo 1 es el equivalente a condenar un inocente. Observando la tabla siguiente podemos ver los tipos de errores y las probabilidades de cometerlos.

	$\boldsymbol{H_0}$ **cierta**	$\boldsymbol{H_0}$ **falsa**
Aceptar $\boldsymbol{H_0}$	No hay error $1-\alpha =$ nivel de confianza	Error de tipo II β
Rechazar $\boldsymbol{H_0}$	Error de tipo I $\alpha =$ nivel de significación	No hay error $1-\beta =$ potencia $= p$

Por tanto

$$P(Rechazar H_0|H_0\ es\ verdadera\ ) = \alpha$$

$$P(Rechazar H_0|H_0\ es\ falsa) = 1-\beta = p$$

$$P(Aceptar\ H_0|H_0\ es\ falsa) = \beta$$

Los dos tipos de error dependen uno del otro. No podemos reducir los valores de α y β de forma simultánea, para una prueba de hipótesis con tamaño de muestra fijo. Reducir el valor de α elevará el valor de β y viceversa. Este problema podemos resolverlo aumentando mucho el tamaño de la muestra, pero esta solución no siempre es factible y en muchos casos encarece muchísimos los costos de los estudios, en la práctica es imposible minimizar ambos errores simultáneamente. Dada la imposibilidad de disminuir simultáneamente la probabilidad de cometer ambos errores, la búsqueda de una buena regla de decisión se realiza de la siguiente manera:

1. Se fija la probabilidad de cometer error de tipo 1 mediante el nivel de significación.
$$P\ (Error\ de\ tipo\ I) = P(Rechazar H_0|H_0\ es\ verdadera\ ) = \alpha$$
2. Dentro de todas las reglas de decisión que tengan nivel de significación menor o igual que α se selecciona aquella que hace mínima la probabilidad de cometer el error de tipo II.
En todas las dócimas paramétrica que vamos a estudiar se cumple la probabilidad de cometer el error de tipo II es mínima.

Los valores de α más usados en la práctica son $0.05, 0.01$ y 0.1. Nos interesamos particularmente en los casos en que $D(\theta) = N\ (\mu, \sigma^2) con\ \theta = \mu$ ó $\theta\ = \sigma^2$ y $D(\theta) = B(n,p)\ con\ \theta = p$

Dócimas paramétricas

Llamémosle θ a un parámetro de una población y θ_0 al valor prefijado de θ a partir del cual definimos las hipótesis, podemos entonces definir los planteamientos siguientes:

a.	b.	c.
$H_0: \theta \leq \theta_0$ $H_0: \theta > \theta_0$	$H_0: \theta \geq \theta_0$ $H_0: \theta < \theta_0$	$H_0: \theta = \theta_0$ $H_0: \theta \neq \theta_0$

Basados en el criterio anterior se pueden encontrar las reglas de decisión adecuadas para los problemas de pruebas de hipótesis asociados a los parámetros μ y σ^2 de la distribución Normal, y el parámetro p de la distribución Binomial.

Prueba de Hipótesis para $\boldsymbol{\mu}$

En el caso de la media podemos ver una representación gráfica en la figura 1. A las pruebas de tipo y b se les llama pruebas de una cola y a las pruebas de tipo c se les llama pruebas de dos colas. En cada caso la probabilidad de rechazar H_0 es el área bajo la curva sombreada en azul. C y C' son los valores cortes de la región critica, o sea los estadísticos que utilizaremos para comparar en la región critica.

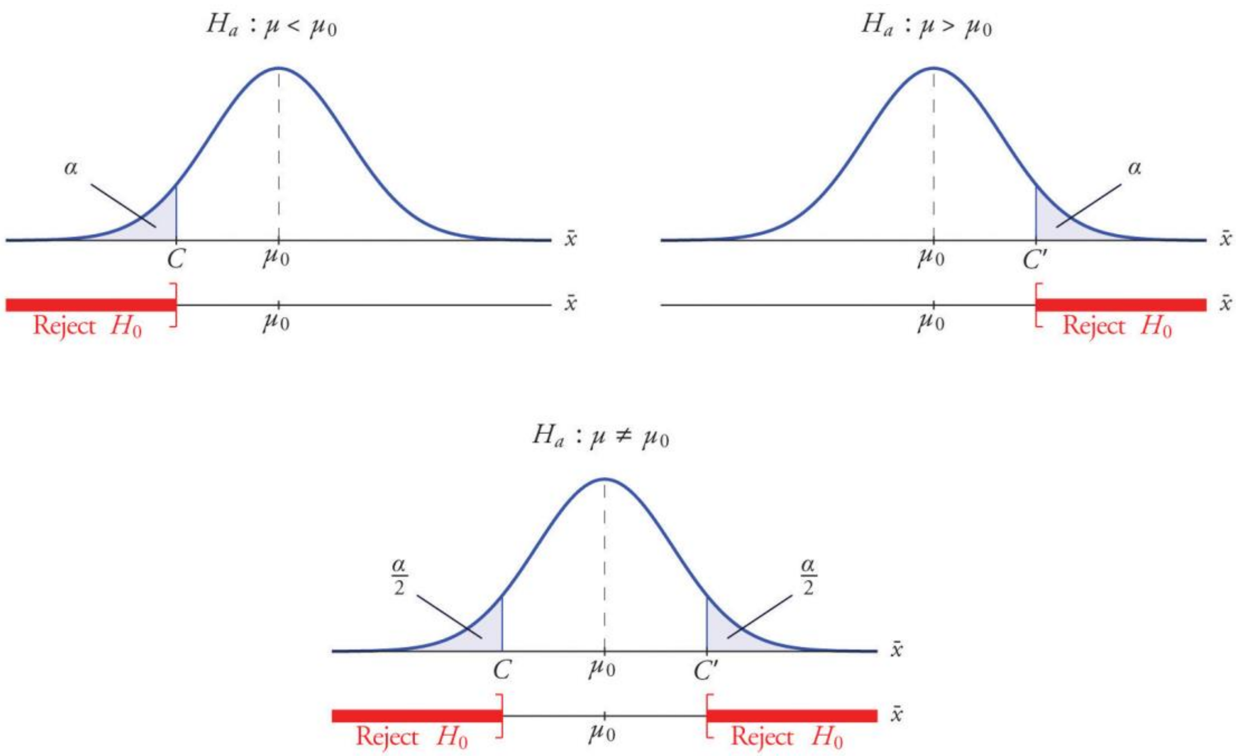

Figura 1. Representación gráfica de las pruebas de hipótesis ara la media ce una distribución normal.

Prueba de Hipótesis para μ con σ^2 conocida.

1. Hipótesis

$H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu \geq \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$

2. Fijar α
3. Estadígrafo

$$Z = \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n}$$

4. Región Critica (RH_0)

$Z > Z_{1-\alpha}$	$Z < -Z_{1-\alpha}$	$\lvert Z \rvert > Z_{1-\frac{\alpha}{2}}$

5. Conclusión

Prueba de Hipótesis para μ con σ^2 desconocida.

1. Hipótesis

$H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu \geq \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$

2. Fijar α
3. Estadígrafo

$$T = \frac{\bar{X} - \mu_0}{S}\sqrt{n}$$

4. Región Critica (RH_0)

$T > t_{1-\alpha}(n-1)$	$T < -t_{1-\alpha}(n-1)$	$\lvert T\rvert > t_{1-\frac{\alpha}{2}}(n-1)$

5. Conclusión

Prueba de Hipótesis para σ^2

Prueba de Hipótesis para σ^2de la distribución Normal

1. Hipótesis

$H_0: \sigma^2 \leq \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$	$H_0: \sigma^2 \geq \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$

2. Fijo α
3. Estadígrafo

$$U = \frac{(n-1)S^2}{\sigma_0^2}$$

4. Región Critica (RH_0)

$U > \chi^2_{1-\alpha}(n-1)$	$U < \chi^2_{\alpha}(n-1)$	$U < \chi^2_{\frac{\alpha}{2}}(n-1)\ o$ $U > \chi^2_{1-\frac{\alpha}{2}}(n-1)$

5. Conclusión

Prueba de Hipótesis para proporciones

Prueba de Hipótesis para proporciones (para el parámetro p de la distribución $B(n,p)$)

1. Hipótesis

$H_0: p \leq p_0$ $H_1: p > p_0$	$H_0: p \geq p_0$ $H_1: p < p_0$	$H_0: p = p_0$ $H_1: p \neq p_0$

2. Fijar α
3. Estadígrafo

$$\bar{Z} = \frac{\bar{p} - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} \ \ y\ n > 30$$

4. Región Critica (RH_0)

$\bar{Z} > Z_{1-\alpha}$	$\bar{Z} < -Z_{1-\alpha}$	$\lvert\bar{Z}\rvert > Z_{1-\frac{\alpha}{2}}$

5. Conclusión

En el ejemplo 1 que hemos venido analizando, tenemos que

$$H_0: p \leq\ 0.8$$
$$H_1: p >\ 0.8$$

Tenemos entonces que Se rechaza $H_0\ si\ \bar{Z} < Z_\alpha$

Tomamos $\alpha\ =\ 0.05$ y supongamos que se le aplicó el examen a 500 estudiantes de 12 grado, de los cuales aprobaron 450. Entonces

$$\bar{p} = \frac{450}{500} = 0.9$$

Tenemos que $n\ =\ 500$ y $p_0 =\ 0.8$, luego

$$\bar{Z}\ = \frac{\bar{p} - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{0.9-0.8}{\sqrt{0.8(1-0.8)}}\sqrt{500} = \frac{0.1}{\sqrt{0.8(1-0.8)}}\sqrt{500} = \frac{0.1}{\sqrt{0.8*0.2}}\sqrt{500}$$
$$= \frac{0.1}{\sqrt{0.16}}22.36 = \frac{0.1}{0.4}22.36 = 0.25*22.36 = 5.59$$

Se sabe que

$$Z_{1-\alpha} = Z_{1-0.05} = Z_{0.95} = 1.65$$

Por consiguiente

$$\bar{Z} = 5.59 >\ 1.65 = Z_{0.95}$$

Se rechaza H_0 con un nivel de significación del 5%. En este caso decimos que es posible afirmar con una probabilidad de error menor o igual que 0.05 que la proporción de estudiantes de 12 grado con conocimientos básicos de matemática es mayor que 0.8 (más del 80% de estudiantes de 12 grado tienen conocimientos avanzados de geometría y trigonometría).

p-value

Al área bajo la curva de las colas de la distribución se le llama también $p - value$ o $p - valor$.

Definición 5. p-valor:

Asumiendo que la hipótesis nula es verdadera el p-valor puede ser definido como la probabilidad de que un estadístico muestral (como la media de la muestra) este lo más alejado posible del valor de la hipótesis obtenido de la muestra en consideración. Esto indica que el p-valor es el menor nivel de significación bajo el cual la hipótesis nula es rechazada.

Este valor es extremadamente útil cuando se realizan pruebas de hipótesis utilizando sistemas, pues en estos casos no escogemos el nivel de significación expresamente, la computadora calcula el p-valor y si este es menor que un α determinado por nosotros entonces se cumple la región critica.