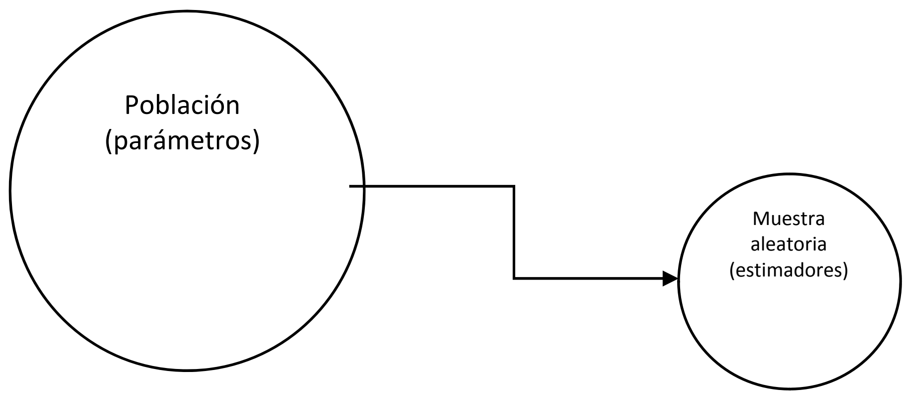
Unidad 4. Inferencia Estadística

Prof. Eliana Guzmán U.

Semestre A-2015

Inferencia Estadística

- La teoría de la Inferencia Estadística está conformada por aquellos métodos que permiten hacer generalizaciones, inferencias o sacar conclusiones sobre la población en estudio.



INFERENCIA ESTADÍSTICA

Existen:

- **Métodos clásicos**: las inferencias se basan en la información proporcionada por una muestra aleatoria, tomada de la población en estudio.

- **Métodos bayesianos**: se basan en conocimiento subjetivo previo acerca de la distribución de probabilidad con los parámetros desconocidos, junto con la información proporcionada por los datos muestrales.

PARÁMETROS

- Son las características medibles de una población.
- Los valores de los parámetros de la población, por lo general, se consideran valores verdaderos.

Áreas de la Inferencia Estadística

La Inferencia Estadística puede dividirse en dos áreas principales:

- **Estimación**: a partir del conocimiento de la distribución muestral, se infiere sobre la distribución de la población (valor de los parámetros poblacionales).

- **Pruebas de hipótesis**: no se intenta estimar un parámetro, sino tomar una decisión correcta respecto a una hipótesis que se plantea. Una vez más, se depende de la teoría del muestreo para obtener alguna medida de precisión para la decisión que se tome.

- Estimación:
  - Estimación puntual.
  - Estimación por Intervalos de Confianza.

Estimación

En este punto se va a tratar el tema de cómo hacer aproximaciones o estimaciones de <u>cuál es el valor que toman los parámetros</u> que describen el comportamiento de la población que se esté estudiando.

Recuerde que cada uno de estos comportamientos, se describe por una distribución de probabilidad teórica, ya sea para variables continuas o discretas.

Formas de Hacer la Estimación

- Estimación Puntual.
- Estimación por Intervalos.

ESTIMACIÓN PUNTUAL

- Consiste en estimar los parámetros poblacionales, haciendo uso de la información que aporta **una sola** muestra aleatoria.

- Estimador: es una variable aleatoria que se utiliza para obtener una estimación puntual.

- La estimación puntual de un parámetro poblacional, es un valor único del estimador, calculado a partir de la muestra.

ESTIMACIÓN PUNTUAL

- En general se tiene:

Población (Parámetros)	Muestra (Estimador)	Estimación Puntual
Θ	$\hat{\Theta}$	$\hat{\theta}$
μ	$\overline{X}$	$\overline{x}$
σ^2	S^2	s^2
P	$\hat{P}$	$\hat{p}$

- No se espera que un estimador estime sin error al parámetro poblacional, sino que no se aleje mucho del valor real.

- La precisión de un estimador se incrementa con muestras grandes, pero no hay razón por la cual esperar que la estimación puntual de una muestra dada, debe ser exactamente igual que el parámetro poblacional que se supone estima.

ESTIMACIÓN PUNTUAL

Propiedades de los estimadores:

- Insesgado: se dice que un estimador $\hat{\Theta}$ es un estimador insesgado del parámetro Θ, si:

$$\mu_{\hat{\Theta}} = E(\hat{\Theta}) = \Theta$$

- Eficiente: si $\hat{\Theta}_1$ y $\hat{\Theta}_2$ son dos estimadores insesgados del mismo parámetro poblacional Θ, se seleccionaría el estimador cuya distribución muestral tuviera la varianza más pequeña. De aquí que, si $\sigma^2_{\hat{\Theta}_1} < \sigma^2_{\hat{\Theta}_2}$, se afirma que $\hat{\Theta}_1$ es un estimador más eficiente de Θ que $\hat{\Theta}_2$.

Estimación Puntual

- Consistente: cuando el tamaño de la muestra crece arbitrariamente, el valor que toma el estimador debe aproximarse al valor del parámetro desconocido.

  Si se verifican las condiciones:

  $$\lim_{n\to\infty} E(\hat{\Theta}) = \Theta$$

  $$\lim_{n\to\infty} Var(\hat{\Theta}) = 0$$

  entonces $\hat{\Theta}$ es consistente.

- Suficiente: el estimador debería aprovechar toda la información existente en la muestra aleatoria, que sirva para estimar al parámetro.

ESTIMACIÓN PUNTUAL DE LA MEDIA

- La distribución muestral de $\bar{X}$ se centra en μ y en la mayoría de las aplicaciones, la varianza de la media muestral, es menor que cualquier otro estimador de μ.
- Así la media muestral $\bar{X}$ se utilizará como una estimación puntual de la media poblacional μ.
- Recuerde que $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$, de tal manera que una muestra grande dará un valor de $\bar{X}$ que surge de una distribución muestral con una varianza pequeña. De aquí que $\bar{x}$ es probablemente una estimación muy precisa de μ cuando n es grande.

- Estimación:
  - Estimación puntual.
  - **Estimación por Intervalos de Confianza.**

Estimación por Intervalos

- Es improbable que incluso el estimador más eficaz, estime con exactitud al parámetro poblacional. Es cierto que la precisión aumenta con muestras grandes, pero no hay razón por la cual se debe esperar que una estimación puntual de una muestra dada sea exactamente igual al parámetro poblacional.
- Hay muchas situaciones en que es preferible determinar un intervalo dentro del cual se espera encontrar el valor del parámetro, lo cual recibe el nombre de estimación por intervalo.

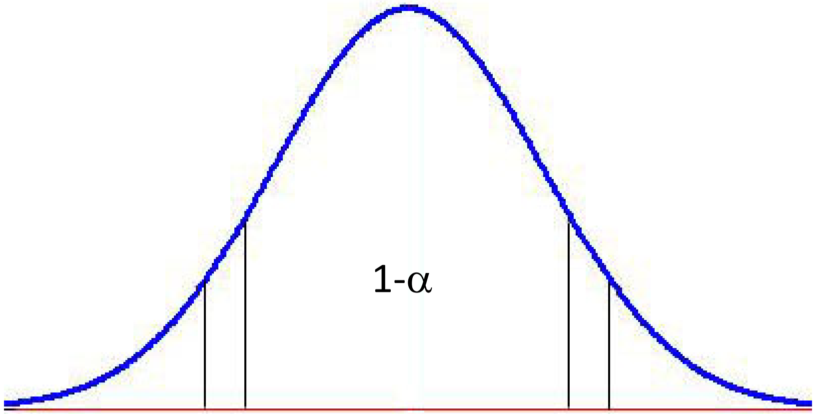
Estimación por Intervalos

- La estimación por intervalo del un parámetro poblacional Θ, es un intervalo de la forma $\bar{\theta}_I < \Theta < \bar{\theta}_S$, donde sus límites dependen del valor del estadístico $\hat{\Theta}$ para una muestra específica.

Estimación por Intervalos

- Los intervalos de confianza se construyen usando un coeficiente o nivel de confianza ($1-\alpha$), que no es más que la probabilidad de que el parámetro que se está estimando, se encuentre dentro de dicho intervalo. Cuanto más amplio sea el intervalo de confianza, se tiene una mayor confianza de que el intervalo dado contenga al parámetro desconocido.

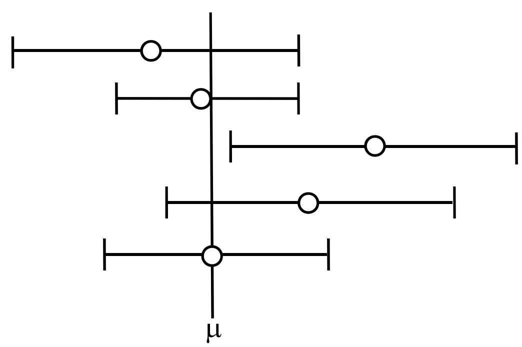
Estimación por Intervalos



Lo ideal sería tener un intervalo corto con un mayor grado de confianza. Algunas veces las restricciones en el tamaño de la muestra, impiden tener intervalos pequeños sin sacrificar algo del grado de confianza.

ESTIMACIÓN POR INTERVALOS

- Se puede establecer un intervalo de estimación para la media μ, si la muestra se selecciona de una **población normal** o si **n es grande** ($n \geq 30$), considerando la distribución muestral de $\bar{X}$.



Intervalos de confianza a estudiar para la media

- Intervalo de Confianza para μ, si se conoce σ.
- Intervalo de Confianza para μ, si no se conoce σ.
- Intervalo de Confianza para μ, de muestra grande.

INTERVALO DE CONFIANZA PARA μ, SI SE CONOCE σ.

- Se selecciona una muestra aleatoria de tamaño n de población cuya varianza σ^2 se conoce y la media $\bar{x}$ se calcular para obtener el siguiente intervalo de confianza del (1-α)100% para μ:

$$\bar{x} - Z_{\alpha/2} \sigma/\sqrt{n} < \mu < \bar{x} + Z_{\alpha/2} \sigma/\sqrt{n}$$

donde $Z_{\alpha/2}$ es el valor de Z a la derecha del cual se tiene un área de α/2.

INTERVALO DE CONFIANZA PARA μ, SI SE CONOCE σ.

- **Error**: si μ es realmente el valor del centro del intervalo, entonces estima a μ sin error. La mayor parte de las veces, sin embargo, no será exactamente igual a μ y la estimación puntual es errónea. El tamaño de este error (e) será el valor absoluto de la diferencia entre y μ y se puede tener una confianza del (1-α)100% que esta diferencia no excederá de $z_{\alpha/2}\,\sigma/\sqrt{n}$.

Intervalo de Confianza para μ, si se conoce σ.

- **Tamaño de la muestra**: si se utiliza a como una estimación de μ, se puede tener una confianza del $(1-\alpha)100\%$ de que el error no excederá una cantidad específica e cuando el tamaño de la muestra es:

$$n = \left(\frac{Z_{\alpha/2}\sigma}{e} \right)^2$$ redondeada al entero superior más cercano

Intervalo de Confianza para μ, si no se conoce σ.

- Si $\bar{x}$ y s son la media y la desviación estándar de una muestra aleatoria tomada de una población normal con varianza desconocida σ^2, un intervalo de confianza del $(1-\alpha)100\%$ para μ es:

$$\bar{x} - t_{\alpha/2} \, {}^{s}\!/\!\sqrt{n} < \mu < \bar{x} + t_{\alpha/2} \, {}^{s}\!/\!\sqrt{n}$$

donde $t_{\alpha/2}$ es el valor de la Distribución t de Student con n-1 grados de libertad, que deja un área de $\alpha/2$ a la derecha.

Intervalo de Confianza para μ, de muestra grande

- Incluso cuando no se puede suponer la normalidad de la población, con σ desconocida y n≥30, s podría reemplazar a σ y se podría utilizar el intervalo de confianza:

$$\bar{x} - Z_{\alpha/2} \, s/\sqrt{n} < \mu < \bar{x} + Z_{\alpha/2} \, s/\sqrt{n}$$

donde $Z_{\alpha/2}$ es el valor de Z a la derecha del cual se tiene un área de $\alpha/2$.

Estimación de la diferencia entre dos medias

- Si se tienen dos poblaciones con medias μ_1 y μ_2 y varianzas σ_1^2 y σ_2^2 respectivamente, el estimador puntual de la diferencia entre μ_1 y μ_2 lo da el estadístico $\bar{X}_1 - \bar{X}_2$.
- Intervalos de confianza a estudiar para la diferencia de medias:
  - Intervalo de Confianza para μ_1 - μ_2 conociendo σ_1^2 y σ_2^2.
  - Intervalo de Confianza para μ_1 - μ_2, con $\sigma_1^2=\sigma_2^2$ pero desconocidas.
  - Intervalo de Confianza para μ_1 - μ_2 con $\sigma_1^2\neq\sigma_2^2$ y desconocidas.

Intervalo de Confianza para μ_1 - μ_2 conociendo σ_1^2 y σ_2^2:

- Si $\bar{x}_1$ y $\bar{x}_2$ son las medias de muestras aleatorias independientes de tamaños n_1 y n_2 de poblaciones con varianzas conocidas σ_1^2 y σ_2^2, respectivamente, un intervalo de confianza del $(1-\alpha)100\%$ para μ_1 - μ_2 es:

$$(\bar{x}_1 - \bar{x}_2) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

donde $Z_{\alpha/2}$ es el valor de Z que tiene a su derecha un área bajo la curva igual a $\alpha/2$.

Intervalo de Confianza para μ_1 - μ_2, con $\sigma_1^2=\sigma_2^2$ pero desconocidas

- Si $\bar{x}_1$ y $\bar{x}_2$ son las medias de muestras aleatorias independientes de tamaños n_1 y n_2, respectivamente, de poblaciones aproximadamente normales con varianzas iguales pero desconocidas, un intervalo de confianza del $(1-\alpha)100\%$ para μ_1 - μ_2 es:

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Intervalo de Confianza para $\mu_1 - \mu_2$, con $\sigma_1^2 = \sigma_2^2$ pero desconocidas

donde S_p es la estimación común de la desviación estándar poblacional y $t_{\alpha/2}$ es el valor de la distribución t con $v = n_1 + n_2 - 2$ grados de libertad, que tiene a su derecha un área bajo la curva igual a $\alpha/2$.

$$S_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Intervalo de Confianza para μ_1 - μ_2 con $\sigma_1^2 \neq \sigma_2^2$ y desconocidas

- Si $\bar{x}_1$ y s_1^2 y $\bar{x}_2$ y s_2^2 son las medias y varianzas de muestras aleatorias pequeñas e independientes de tamaños n_1 y n_2, respectivamente, de poblaciones aproximadamente normales con varianzas diferentes y desconocidas, un intervalo de confianza del $(1-\alpha)100\%$ para $\mu_1 - \mu_2$ es:

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Si las muestras n_1 y n_2 son grandes (mayores o iguales a 30), no es necesario conocer que la población sigue un comportamiento normal.

Intervalo de Confianza para μ_1 - μ_2 con $\sigma_1^2 \neq \sigma_2^2$ y desconocidas

donde $t_{\alpha/2}$ es el valor de la distribución t con v grados de libertad, que tiene a su derecha un área bajo la curva igual a $\alpha/2$.

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\left[(s_1^2/n_1)^2/(n_1-1)\right] + \left[(s_2^2/n_2)^2/(n_2-1)\right]}$$

ESTIMACIÓN DE LA PROPORCIÓN

- Un estimador puntual de la proporción p, en un experimento binomial, está dado por el estadístico $\hat{P} = X / n$, donde X representa el número de éxitos en n intentos. Por lo tanto, la proporción muestral $\hat{p} = x / n$ se utiliza como estimación puntual del parámetro p.
- Si no se espera que la proporción desconocida p se acerque demasiado a cero o a uno, se puede establecer un intervalo de confianza para p considerando la distribución muestral de $\hat{p}$.

INTERVALO DE CONFIANZA PARA P DE UNA MUESTRA GRANDE

- Si $\hat{p}$ es la proporción de éxitos en una muestra aleatoria de tamaño n, y $\hat{q} = 1 - \hat{p}$, un intervalo de confianza aproximado de (1-α)100% para el parámetro binomial p es:

$$\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

donde $Z_{\alpha/2}$ es el valor de Z que tiene a su derecha un área bajo la curva igual a α/2.

- Cuando n es pequeño y se cree que la proporción desconocida p se acerca a 0 o a 1, este intervalo no es confiable y, por lo tanto, no debe utilizarse. Para estar seguro se requiere que ambos n$\hat{p}$ y n$\hat{q}$ sean mayores o iguales a 5.

INTERVALO DE CONFIANZA PARA P DE UNA MUESTRA GRANDE

- Error: Si se utiliza a $\hat{p}$ como una estimación puntual de p, se puede tener una confianza del (1-α)100% de que el error no excederá de:

$$e \leq Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

INTERVALO DE CONFIANZA PARA P DE UNA MUESTRA GRANDE

- Tamaño de n: Si $\hat{p}$ se utiliza como una estimación puntual de p, se puede tener una confianza del (1-α)100% de que el error será menor que una cantidad específica e cuando el tamaño de la muestra sea aproximadamente:

$$n = \frac{Z^2_{\alpha/2}\hat{p}\hat{q}}{e^2}$$

- Cuando no se conoce una estimación de p:

$$n = \frac{Z^2_{\alpha/2}}{4e^2}$$

ESTIMACIÓN DE LA DIFERENCIA ENTRE DOS PROPORCIONES

- Para comparar dos poblaciones con comportamiento Binomial, se utilizará un Intervalo de Confianza para $p_1 - p_2$ de una muestra grande.

ESTIMACIÓN DE LA DIFERENCIA ENTRE DOS PROPORCIONES

- Si $\hat{p}_1$ y $\hat{p}_2$ son las proporciones de éxito en muestras aleatorias de tamaños n_1 y n_2 de dos poblaciones binomiales, un intervalo de confianza del $(1-\alpha)100\%$ para la diferencia entre los dos parámetros binomiales, p_1-p_2, es:

$$(\hat{p}_1 - \hat{p}_2) - Z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + Z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

- donde $Z_{\alpha/2}$ es el valor de Z que tiene a su derecha un área bajo la curva igual a $\alpha/2$.
- Se requiere que: $n_1\hat{p}_1$, $n_1\hat{q}_1$, $n_2\hat{p}_2$ y $n_2\hat{q}_2$ sean mayores o iguales a 5.

Pruebas de Hipótesis

Fundamentos de las pruebas de hipótesis

- Las pruebas de hipótesis consisten en la formulación de un procedimiento de decisión basado en datos experimentales, que pueda producir una conclusión acerca del sistema en estudio.
- Una **hipótesis estadística** es una afirmación o conjetura acerca de una o más poblaciones.
- Se plantean las conjeturas o hipótesis, se utilizan datos experimentales (muestra aleatoria) y se toman decisiones con base en ellos.

FUNDAMENTOS DE LAS PRUEBAS DE HIPÓTESIS

- Nunca se sabe con absoluta certeza la verdad o falsedad de una hipótesis estadística, a no ser que se examine la población entera, pero como se sabe que esto no es posible en la mayoría de los casos prácticos, se toma una muestra aleatoria de la población de interés y se utilizan dichos datos para proporcionar evidencias que confirmen o no la hipótesis. La evidencia de la muestra que es inconsistente con la hipótesis planteada, conduce a un rechazo de la misma, mientras que la evidencia que apoya la hipótesis, conduce a su aceptación.

FUNDAMENTOS DE LAS PRUEBAS DE HIPÓTESIS

- La aceptación de una hipótesis implica tan solo que los datos no proporcionan evidencias suficientes para refutarla. Por otro lado, el rechazo implica que la evidencia de la muestra la refuta.
- Si el científico se interesa en respaldar con fuerza un argumento, espera probar dicho argumento en la forma de rechazo de una hipótesis.

Fundamentos de las pruebas de hipótesis

- Por ejemplo, si se desea mostrar que el tomar café aumenta el riesgo de padecer cáncer, la hipótesis a probar deberá ser “no hay aumento en el riesgo de padecer de cáncer debido a la ingestión de café”; como resultado el argumento se alcanza vía rechazo.

Formulación de hipótesis. Hipótesis simple y compuesta

- La estructura de la prueba de hipótesis se formulará utilizando el término hipótesis nula. Esto se refiere a cualquier hipótesis que se desee probar y se representa por H_0.
- El rechazo de H_0 da como resultado la aceptación de una hipótesis alternativa, que se representa por H_1.

Formulación de hipótesis. Hipótesis simple y compuesta

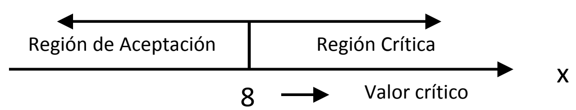
- Una hipótesis nula referida a un parámetro poblacional, siempre será establecida en forma tal que especifique un valor exacto del parámetro, mientras que la hipótesis alternativa admite la posibilidad de varios valores.
- Por ejemplo: Si H_0: $p=0.5$, la hipótesis alternativa puede ser: H_1: $p < 0.5$, ó

  H_1: $p > 0.5$, ó

  H_1: $p \neq 0.5$.

PRUEBA DE UNA HIPÓTESIS ESTADÍSTICA

- La decisión se basa en un **estadístico de prueba**, por ejemplo X: cantidad de éxitos, si se está trabajando con una Distribución Binomial.
- Y en el establecimiento de la **región crítica**, que es la zona donde se rechaza H_0.



Tipos de errores que se pueden cometer en las Pruebas de Hipótesis

- Este procedimiento que se acaba de describir podría conducir a cualquiera de dos conclusiones erróneas, que se resumen en la siguiente tabla:

	H_0 es verdadera	H_0 es falsa
Se acepta H_0	Decisión correcta	Error tipo II (β)
Se rechaza H_0	Error tipo I (α)	Decisión correcta

- Nivel de significancia (α): es la probabilidad de cometer un error tipo I.

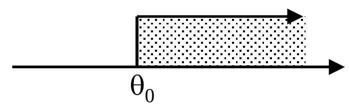
Tipos de errores que se pueden cometer en las Pruebas de Hipótesis

- Para una muestra de tamaño fijo, la disminución en la probabilidad de cometer un error, casi siempre resulta en un incremento en la probabilidad de cometer el otro error.
- Afortunadamente, la probabilidad de cometer ambos tipos de errores (α y β), solo puede reducirse, incrementando el tamaño de la muestra.

Tipos de pruebas de Hipótesis

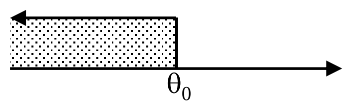
- Pruebas de una cola: donde H_1 es unilateral, por ejemplo:

$$H_0: \theta = \theta_0$$
$$H_1: \theta > \theta_0$$



ó

$$H_0: \theta = \theta_0$$
$$H_1: \theta < \theta_0$$



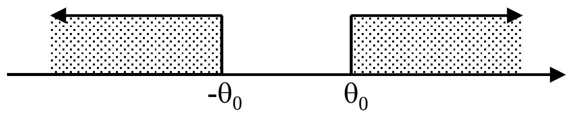
Donde θ es un parámetro poblacional.

TIPOS DE PRUEBAS DE HIPÓTESIS

- Pruebas de dos colas: donde H_1 es bilateral

$H_0: \theta = \theta_0$ La región crítica se divide en 2 partes,

$H_1: \theta \neq \theta_0$ ya que H_1 establece que $\theta < -\theta_0$ y $\theta > \theta_0$

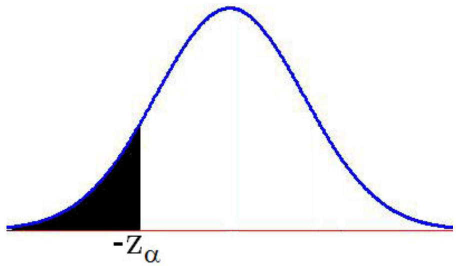


Donde θ es un parámetro poblacional.

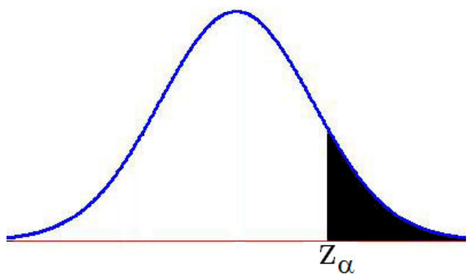
REGIÓN CRÍTICA

Para todas las pruebas de hipótesis la región crítica será:

- Si se tiene una prueba de una sola cola la región crítica es:
  - Si H_1: $\theta < \theta_0 \rightarrow z < -z_\alpha$

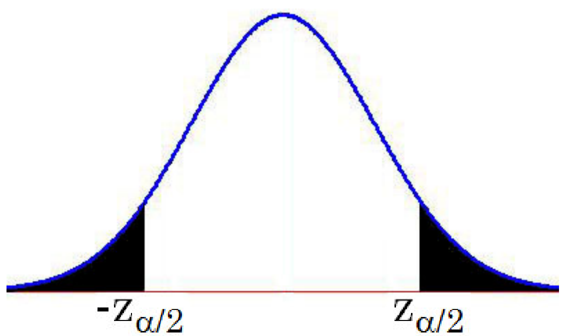


  - Si H_1: $\theta > \theta_0 \rightarrow z > z_\alpha$



Región Crítica

- Si se tiene una prueba de dos colas (H_1: $\theta \neq \theta_0$), la región crítica es $z < -z_{\alpha/2}$ y $z > z_{\alpha/2}$.



Pruebas a estudiar

- En cada prueba el procedimiento es el mismo, lo único que cambia es el estadístico de prueba a utilizar y el establecimiento de la región crítica.
- En cada prueba es muy importante plantear correctamente las hipótesis a utilizar, ya que un error en dicho planteamiento, cambia totalmente las decisiones a tomar.

PRUEBAS A ESTUDIAR

<u>Guías para establecer las hipótesis</u>:

Leer cuidadosamente el problema para determinar la afirmación que desea probarse:

- Si esta afirmación sigue una sola dirección como, por ejemplo, más que, menos que, inferior a, superior a, etc. Entonces en H_1 se establecerá el signo de desigualdad ($<$ ó $>$) correspondiente a la dirección sugerida.

PRUEBAS A ESTUDIAR

- Si la afirmación sugiere una dirección compuesta ($=, \leq, \geq$) como, por ejemplo, al menos, a lo sumo, igual o mayor que, no mayor que, etc., entonces esta dirección compuesta completa ($\leq, \geq$) se expresa como H_0 pero usando solo el signo de igualdad y H_1 se expresa en la dirección opuesta.
- Si no se sugiere ninguna dirección en la afirmación, entonces H_1 se establece utilizando el signo diferente que ($\neq$).

ETAPAS A SEGUIR EN LAS PRUEBAS DE HIPÓTESIS

1. Establecer la hipótesis nula H_0: $\theta = \theta_0$
2. Seleccionar una hipótesis alternativa apropiada H_1 de una de las alternativas: $\theta < \theta_0$, $\theta > \theta_0$, ó $\theta \neq \theta_0$.
3. Seleccionar un nivel de significancia de tamaño α.
4. Seleccionar el estadístico de prueba apropiado y establecer la región crítica.
5. Calcular el valor del estadístico de prueba a partir de los datos muestrales.
6. Decidir: **rechazar H_0** si el estadístico de prueba se ubica dentro de la región crítica, de otra forma, **no rechazar H_0**.

PRUEBAS DE HIPÓTESIS A ESTUDIAR

- Prueba de hipótesis para la media (μ) con σ^2 conocida.
- Prueba de hipótesis para la media (μ) con σ^2 desconocida.
- Prueba de hipótesis para las proporciones (p):
  - Si n no es grande.
  - Si n es grande.

PRUEBA DE HIPÓTESIS PARA LA MEDIA (μ) CON σ^2 CONOCIDO

- Se siguen las mismas seis etapas que para cualquier prueba de hipótesis, usando el estadístico de prueba:

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

PRUEBA DE HIPÓTESIS PARA LA MEDIA (μ) CON σ^2 DESCONOCIDO

- Se siguen las mismas seis etapas que para cualquier prueba de hipótesis, usando el estadístico de prueba:

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

con $\nu = n - 1$ grados de libertad

PRUEBA DE HIPÓTESIS PARA LAS PROPORCIONES (P)

- Consiste en probar la hipótesis de que la proporción de éxitos en un experimento Binomial p, es igual que un valor especificado.
- Se usa como estadístico de prueba la variable Binomial X con $p=p_0$.
- Se siguen las seis etapas de la pruebas de hipótesis, solo que no se calcula un estadístico de prueba sino el valor de P.
- La región crítica siempre será $P \leq \alpha$, sin importar como sea H_1.

PRUEBA DE HIPÓTESIS PARA LAS PROPORCIONES (P)

Existen dos casos:

1. Si n no es grande: Debido a que X es una v.a. Binomial discreta, es poco probable que pueda determinarse una región crítica cuyo tamaño sea exactamente igual a α. Por esta razón es preferible, al tratar con muestras pequeñas, basar las decisiones en el valor denominado **P**.

PRUEBA DE HIPÓTESIS PARA LAS PROPORCIONES (P)

Existen entonces tres casos:

a. Si:

H_0: $p = p_0$

H_1: $p < p_0$

$\mathbf{P} = P(X \leq x)$ cuando $p=p_0$

donde x es el número de éxitos en la muestra de tamaño n

Si $P \leq \alpha$, la prueba es significativa en el nivel α y se rechaza H_0 a favor de H_1.

PRUEBA DE HIPÓTESIS PARA LAS PROPORCIONES (P)

b. Si:

H_0: $p = p_0$
H_1: $p > p_0$

$\left\{ \begin{array}{l} \mathbf{P} = P(X \geq x) \text{ cuando } p=p_0 \\ \text{donde x es el número de éxitos en la muestra de tamaño n} \end{array} \right.$

Si $P \leq \alpha$, la prueba es significativa en el nivel α y se rechaza H_0 a favor de H_1.

PRUEBA DE HIPÓTESIS PARA LAS PROPORCIONES (P)

c. Si:

H_0: $p = p_0$

H_1: $p \neq p_0$

P $= 2P(X \leq x)$ cuando $p=p_0$ si $x<np_0$

P $= 2P(X \geq x)$ cuando $p=p_0$ si $x>np_0$

donde x es el número de éxitos en la muestra de tamaño n

Si $P \leq \alpha$, la prueba es significativa en el nivel α y se rechaza H_0 a favor de H_1.

PRUEBA DE HIPÓTESIS PARA LAS PROPORCIONES (P)

2. Si n es grande y p_0 no se acerca a cero o uno, se utiliza la aproximación normal y el valor z para probar que H_0: $p = p_0$ está dado por:

$$z = \frac{x - np_0}{\sqrt{np_0 q_0}}$$

el cual es un valor de la variable aleatoria normal estándar Z.