# PRUEBAS DE HIPÓTESIS

UNIDAD N° 2

# HIPÓTESIS ESTADÍSTICAS

Una hipótesis estadística es una aseveración o conjetura con respecto a una o más poblaciones

La verdad o falsedad de una hipótesis nunca se sabe con absoluta certidumbre a menos que examinemos toda la población (lo que es muy poco práctico)

Trabajamos con una muestra aleatoria para proporcionar evidencia que apoye o no la hipótesis, para de esta manera aceptarla o rechazarla

La aceptación de una hipótesis implica que los datos no dan suficiente evidencia para rechazarla

El rechazo implica que la evidencia muestral la refuta

El planteamiento de una hipótesis está influido por la estructura de la probabilidad de una conclusión errónea: *si el científico se interesa en apoyar con fuerza una opinión, desea llegar a la opinión en la forma de rechazo de una hipótesis*

## HIPÓTESIS NULA $H_0$

- CUALQUIER HIPÓTESIS QUE DESEAMOS PROBAR
- UNA HIPÓTESIS NULA CON RESPECTO A UN PARÁMETRO POBLACIONAL SIEMPRE SE ESTABLECERÁ DE MODO QUE ESPECIFIQUE UN VALOR EXACTO DEL PARÁMETRO

## HIPÓTESIS ALTERNATIVA $H_1$

- EL RECHAZO DE $H_0$ CONDUCE A LA HIPÓTESIS ALTERNATIVA
- LA HIPÓTESIS ALTERNATIVA PERMITE LA POSIBILIDAD DE VARIOS VALORES

Ejemplo: si $H_0$ es la hipótesis nula $p=0,5$ para una población binomial, la hipótesis alternativa $H_1$ sería una de las siguientes:

$$p > 0,5 \qquad p < 0,5 \qquad p \neq 0,5$$

# PRUEBA DE HIPÓTESIS

Esta técnica se usa para comprobar una afirmación acerca del valor de algún parámetro desconocido

Para implementar el procedimiento, el primer paso consiste en formular claramente las hipótesis del problema:

- ♣ Hipótesis nula ($H_0$)
- ♣ Hipótesis alternativa ($H_1$)

# PRUEBA DE UNA HIPÓTESIS ESTADÍSTICA

***Estadística de prueba:***

Elementos de la muestra sobre los que se basa nuestra decisión

***Región crítica:***

Elementos de la muestra que superan el valor esperado

***Región de aceptación:***

Son los valores menores o iguales al valor esperado

***Valor crítico:***

Es el último valor que se observa al pasar de la región de aceptación a la región crítica

# ERROR TIPO I Y ERROR TIPO II

El proceso de decisión puede conducir a dos conclusiones erróneas

## Error Tipo I

Se rechaza la hipótesis nula cuando ésta es verdadera

La probabilidad de cometer error tipo I, se llama ***nivel de significancia*** ($\alpha$)

## Error Tipo II

Se acepta la hipótesis nula cuando ésta es falsa

La probabilidad de cometer error tipo II, se denota por $\beta$, es imposible de calcular a menos que tengamos una hipótesis alternativa específica

# ERRORES DE TIPO I Y DE TIPO II

- SE ACLARA LA RELACIÓN ENTRE EL PROCESO DE DECISIÓN Y LOS ERRORES FACTIBLES EN LA SIGUIENTE TABLA, EXISTIENDO SÓLO DOS POSIBILIDADES: $H_0$ ES VERDADERA O FALSA. EL ESTADO REAL INDICA EL ESTADO CORRECTO DE LOS ASUNTOS RELATIVOS A $H_0$.

| **Decisión** | **Estado Real** | |
|---|---|---|
| | *$H_0$ es verdadera* | *$H_0$ es falsa* |
| Aceptar $H_0$ | Decisión correcta | Error de Tipo II |
| Rechazar $H_0$ | Error de Tipo I | Decisión correcta |

# NIVEL $\alpha$ Y EL PROCESO DE DECISIÓN

El nivel $\alpha$ que los científicos establecen al principio de un experimento es el nivel al cual desean limitar la probabilidad de cometer un Error de tipo I

Si se establece $\alpha$=0,05 está indicando que al reunir los datos, está dispuesto a limitar la probabilidad de no aceptar la $H_0$ cuando sea verdadera, a 5 veces en 100. Se limita la probabilidad de cometer un Error de Tipo I a 0,05

Si restringimos $\alpha$, disminuimos la probabilidad de cometer un Error de Tipo I, pero aumentamos la probabilidad de incurrir en un Error de Tipo II ($\beta$)

# ETAPAS DE UNA PRUEBA DE HIPÓTESIS

1) Identificación de la variable a estudiar

2) Formulación de la hipótesis nula y alternativa

3) Selección del criterio de decisión, o estadístico de prueba, y determinación de su distribución de probabilidades

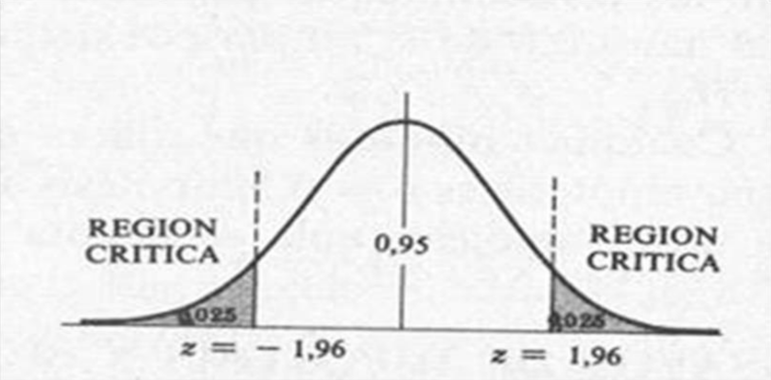
4) Selección del nivel de significación de la prueba, y cálculo de los puntos críticos –límites entre la zona de aceptación y de rechazo en la distribución del estadístico de prueba-

5) Cálculo de una estimación del estadístico de prueba, en base a los resultados del muestreo

6) Comparación entre la estimación del estadístico de prueba y los valores críticos, y decisión acerca del resultado de la prueba

# PRUEBAS EMPLEANDO DISTRIBUCIONES NORMALES

De acuerdo con una determinada hipótesis, la distribución muestral de un estadístico *S* es una distribución muestral con media $\mu_S$ y desviación estándar $\sigma_S$, por lo que la distribución de la variable estandarizada dada por $z = \frac{S-\mu_S}{\sigma_S}$ es la distribución normal estándar




Si observamos la imagen, se puede tener una confianza del 95% en que la hipótesis es verdadera, entonces la puntuación *z* del estadístico muestral real *S* estará entre -1.96 y 1.96 (el área bajo la curva normal entre estos dos valores es el 95%)

El 0.05, que es el total del área sombreada, es el nivel de significancia de la prueba. Esta cantidad representa la probabilidad de estar equivocado al rechazar la hipótesis (probabilidad de cometer error tipo I)

El conjunto de puntuaciones *z* que queda fuera del intervalo -1.96 a 1.96 constituye lo que se llama *región crítica de la hipótesis, región de rechazo de la hipótesis* o *región de significancia*. Al conjunto de puntuaciones *z* que queda dentro del intervalo -1.96 a 1.96 se le llama *región de aceptación de la hipótesis*

# TABLA PARA VALORES DE SIGNIFICANCIA MÁS UTILIZADOS

| Nivel de significancia $\alpha$ | 0.10 | 0.05 | 0.01 | 0.005 | 0.002 |
|---|---|---|---|---|---|
| Valores críticos de z para pruebas de una cola | -1.28 o 1.28 | -1.645 o 1.645 | -2.33 o 2.33 | -2.58 o 2.58 | -2.88 o 2.88 |
| Valores críticos de z para pruebas de dos colas | -1.645 y 1.645 | -1.96 y 1.96 | -2.58 y 2.58 | -2.81 y 2.81 | -3.08 y 3.08 |

Con excel es posible calcular el valor crítico de z que se desee, sólo se debe tener en cuenta, que si se trabaja a dos colas, se debe dividir a la mitad la región crítica:

<u>Para $\alpha = 0{,}04$</u>

A una cola: =DISTR.NORM.ESTAND.INV(0,04) = -1,75 (región crítica hacia la izquierda)

=DISTR.NORM.ESTAND.INV(1 - 0,04) = 1,75 (región crítica hacia la derecha)

A dos colas: =DISTR.NORM.ESTAND.INV(0,04/2) = -2,05 (tener en cuenta que la curva es simétrica, por lo que la región queda comprendida entre -2,05 a 2,05)

# PRUEBAS DE UNA Y DOS COLAS

## PRUEBA DE UNA SOLA COLA

- UNA PRUEBA DE CUALQUIER HIPÓTESIS ESTADÍSTICA, DONDE LA ALTERNATIVA ES ***UNILATERAL***, COMO

$H_0\text{:}\ \ \theta = \theta_0$ O QUIZÁS $H_0\text{:}\ \ \theta = \theta_0$

$H_1\text{:}\ \ \theta > \theta_0$ $\qquad\qquad H_1\text{:}\ \ \theta < \theta_0$

- POR LO GENERAL, LA REGIÓN CRÍTICA PARA LA HIPÓTESIS ALTERNATIVA $\theta > \theta_0$ YACE EN LA COLA DERECHA DE LA DISTRIBUCIÓN DE LA ESTADÍSTICA DE PRUEBA, MIENTRAS QUE LA REGIÓN CRÍTICA PARA LA HIPÓTESIS ALTERNATIVA $\theta < \theta_0$ YACE POR COMPLETO EN LA COLA IZQUIERDA
- EN CIERTO SENTIDO, LA DESIGUALDAD APUNTA EN LA DIRECCIÓN DONDE SE ENCUENTRA LA REGIÓN CRÍTICA

## PRUEBA DE DOS COLAS

- UNA PRUEBA DE CUALQUIER HIPÓTESIS ALTERNATIVA DONDE LA ALTERNATIVA ES ***BILATERAL***, COMO

$$H_0\text{:}\ \ \theta = \theta_0$$
$$H_1\text{:}\ \ \theta \neq \theta_0$$

- SE LLAMA PRUEBA DE DOS COLAS, PUES LA REGIÓN SE DIVIDE EN DOS PARTES, QUE A MENUDO TIENEN PROBABILIDADES IGUALES QUE SE COLOCAN EN CADA COLA DE LA DISTRIBUCIÓN DE LA ESTADÍSTICA DE PRUEBA
- LA HIPÓTESIS ALTERNATIVA $\theta \neq \theta_0$ ESTABLECE QUE $\theta < \theta_0$ O $\theta > \theta_0$

## Pruebas de una y dos colas

La hipótesis nula, $H_0$, siempre se establecerá con el uso del signo de igualdad para que se especifique un solo valor. De esta manera se puede controlar la probabilidad de cometer un error tipo I

Si se establece una prueba de una cola o de dos colas dependerá de la conclusión que se extraiga si $H_0$ se rechaza

La posición de la región crítica se puede determinar sólo después que se establece $H_1$

Son deseables ciertos principios para determinar cuál hipótesis se establecerá como $H_0$ y cuál como $H_1$

Si la afirmación sugiere una sola dirección como *mayor que, menor que, superior a, inferior a*, etc., entonces $H_1$ se establece con el signo de desigualdad ($<$ ó $>$) que corresponda a la dirección sugerida

Si la afirmación sugiere una dirección compuesta (igualdad y dirección) como *al menos, mayor o igual que, a lo más, no mayor que,* etc., entonces toda esta dirección compuesta ($\leq$ ó $\geq$) se expresa como $H_0$, pero con el uso únicamente con el signo igual, y $H_1$ se da en la dirección opuesta

Si no se sugiere ninguna dirección en la afirmación, entonces $H_1$ se establece con el signo de *diferente* ($\neq$)

# PRUEBAS ESPECIALES

Cuando las muestras son grandes, las distribuciones muestrales de muchos estadísticos tienen una distribución normal (o por lo menos aproximadamente normal), y en estas pruebas se puede emplear la correspondiente puntuación $z$

Sólo trabajaremos con algunos de los estadísticos de interés práctico. En cada uno de estos casos, el resultado es válido para poblaciones infinitas o cuando el muestreo se hace con reposición, caso contrario, deben modificarse las fórmulas

# PRUEBAS ESPECIALES

***Media***

Aquí $S = \bar{X}$, la media muestral; $\mu_S = \mu_{\bar{X}} = \mu$, la media poblacional, y $\sigma_S = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}$, donde $\sigma$ es la desviación estándar poblacional y N es el tamaño de la muestra. La puntuación z está dada por:

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}}$$

Si es necesario, para estimar $\sigma$ se emplea la desviación muestral $s$

***Proporciones***

Aquí $S = P$, la proporción de «éxitos» en una muestra; $\mu_S = \mu_p = p$, donde $p$ es la proporción poblacional de éxitos y N es el tamaño de la muestra, y $\sigma_S = \sigma_p = \sqrt{pq/N}$, donde $q = 1 - p$. La puntuación z está dada por:

$$z = \frac{P - p}{\sqrt{pq/N}}$$

En el caso de $P = X/N$, donde $X$ es la cantidad de éxitos obtenidos realmente en una muestra, la puntuación z se transforma en:

$$z = \frac{X - Np}{\sqrt{Npq}}$$

Es decir, $\mu_X = \mu = Np, \ \sigma_X = \sigma = \sqrt{Npq} \ y \ S = X$

# EJEMPLO

- ***UNA MUESTRA ALEATORIA DE 64 BOLSAS DE PALOMITAS DE MAÍZ PESAN, EN PROMEDIO 5.23 ONZAS CON UNA DESVIACIÓN ESTÁNDAR DE 0.24 ONZAS. PRUEBE LA HIPÓTESIS DE QUE $\mu = 5,5$ ONZAS, CONTRA AL HIPÓTESIS ALTERNATIVA, $\mu < 5,5$ ONZAS EN EL NIVEL DE SIGNIFICANCIA DE 0.05***

***SOLUCIÓN***

- SE TRATA DE UNA DISTRIBUCIÓN MUESTRAL DE MEDIAS CON DESVIACIÓN ESTÁNDAR DESCONOCIDA, PERO COMO EL TAMAÑO DE MUESTRA ES MAYOR A 30 SE PUEDE TOMAR LA DESVIACIÓN MUESTRAL COMO UN ESTIMADOR PUNTUAL PARA LA POBLACIONAL
- *DATOS:*

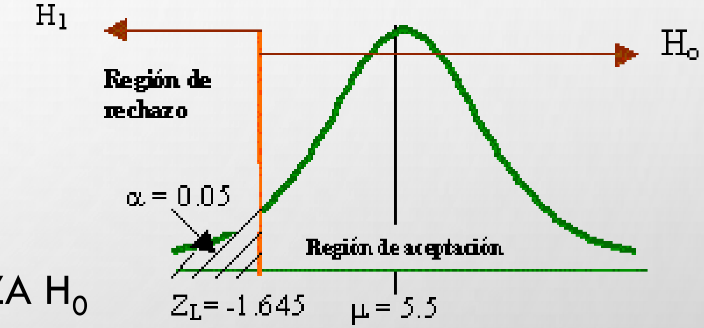
$\mu = 5{,}5\ onzas \quad s = 0{,}24\ onzas \quad \bar{x} = 5{,}23\ onzas \quad n = 64 \quad \alpha = 0{,}05$

- *ENSAYO DE HIPÓTESIS*

$H_0\colon\ \mu = 5{,}5\ onzas \quad H_1\colon\ \mu < 5{,}5\ onzas$

- *REGLA DE DECISIÓN*

SI $z_R \geq -1{,}645$ NO SE RECHAZA $H_0$, SI $z_R < -1{,}645$ SE RECHAZA $H_0$




- CÁLCULOS

$$z_R = \frac{\bar{x}_R - \mu}{\sigma/\sqrt{n}} = \frac{5{,}23 - 5{,}5}{0{,}24/\sqrt{64}} = -9$$

- *JUSTIFICACIÓN Y DECISIÓN*

COMO –9 < -1,645 POR LO TANTO SE RECHAZA $H_O$ Y SE CONCLUYE CON UN NIVEL DE SIGNIFICANCIA DEL 0.05 QUE LAS BOLSAS DE PALOMITAS PESAN EN PROMEDIO MENOS DE 5.5 ONZAS

# VALOR *P* EN PRUEBAS DE HIPÓTESIS

El valor *p* es la probabilidad de obtener un estadístico muestral tan extremo o más extremo que el obtenido, suponiendo que la hipótesis nula sea verdadera

Para probar una hipótesis empleando este método se establece un valor $\alpha$; se calcula el valor $p$ y si el valor $p \leq \alpha$, se rechaza $H_0$

En caso contrario no se rechaza $H_0$

En pruebas para medias empleando muestras grandes (n>30), el valor de *p* se calcula como sigue:

1. Para $H_0$: $\mu = \mu_0$ y $H_1$: $\mu < \mu_0$, valor $p = P(Z < el\ estadístico\ de\ prueba\ elegido)$
2. Para $H_0$: $\mu = \mu_0$ y $H_1$: $\mu > \mu_0$, valor $p = P(Z > el\ estadístico\ de\ prueba\ elegido)$
3. Para $H_0$: $\mu = \mu_0$ y $H_1$: $\mu \neq \mu_0$,

   valor $p = P(Z < -|el\ estadístico\ de\ prueba\ elegido|) + P(Z > |el\ estadístico\ de\ prueba\ elegido|)$

El estadístico calculado es $\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$, donde $\bar{x}$ es la media de la muestra, $s$ es la desviación estándar de la muestra y $\mu_0$ es el valor que se ha especificado para $\mu$ en la hipótesis nula. $\sigma$ no se conoce, se estima a partir de la muestra y se usa $s$

# EJEMPLO

- *UNA MUESTRA ALEATORIA DE 64 BOLSAS DE PALOMITAS DE MAÍZ PESAN, EN PROMEDIO 5.23 ONZAS CON UNA DESVIACIÓN ESTÁNDAR DE 0.24 ONZAS. PRUEBE LA HIPÓTESIS DE QUE $\mu = 5,5$ ONZAS, CONTRA AL HIPÓTESIS ALTERNATIVA, $\mu < 5,5$ ONZAS EN EL NIVEL DE SIGNIFICANCIA DE 0.05*
- *RESOLVER CALCULANDO EL VALOR DE P*

**SOLUCIÓN**

- *DATOS:*

$\mu = 5{,}5\ onzas \quad s = 0{,}24\ onzas \quad \bar{x} = 5{,}23\ onzas \quad n = 64 \quad \alpha = 0{,}05$

- *ENSAYO DE HIPÓTESIS*

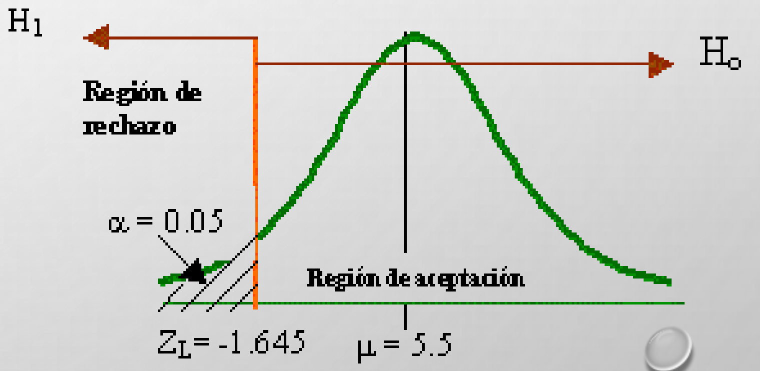
$H_0\colon\ \mu = 5{,}5\ onzas \quad H_1\colon\ \mu < 5{,}5\ onzas$

- *REGLA DE DECISIÓN*

SI $p > 0{,}05$ NO SE RECHAZA $H_0$, SI $p \leq 0{,}05$ SE RECHAZA $H_0$

- CÁLCULOS

$$z_R = \frac{\bar{x}_R - \mu}{\sigma/\sqrt{n}} = \frac{5{,}23 - 5{,}5}{0{,}24/\sqrt{64}} = -9$$




P = DISTR.NORM.ESTAND(-9) = $1{,}12859 \cdot 10^{-19}$

- *JUSTIFICACIÓN Y DECISIÓN*

COMO EL VALOR DE P ES $1{,}12859 \cdot 10^{-19}$ Y ES MENOR AL VALOR DEL NIVEL DE SIGNIFICANCIA DE 0.05 POR LO TANTO SE RECHAZA $H_0$, Y SE CONCLUYE QUE LAS BOLSAS DE PALOMITAS PESAN EN PROMEDIO MENOS DE 5.5 ONZAS

# UNA SOLA MUESTRA: PRUEBAS CON RESPECTO A UNA SOLA MEDIA (VARIANZA CONOCIDA)

Se debe considerar la hipótesis:

$$H_0\colon \mu = \mu_0$$

$$H_1\colon \mu \neq \mu_0$$

La estadística de prueba apropiada se debe basar en la variable aleatoria $\bar{X}$ (con distribución aproximadamente normal, por el teorema del límite central)

Es conveniente estandarizar la variable aleatoria: $z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

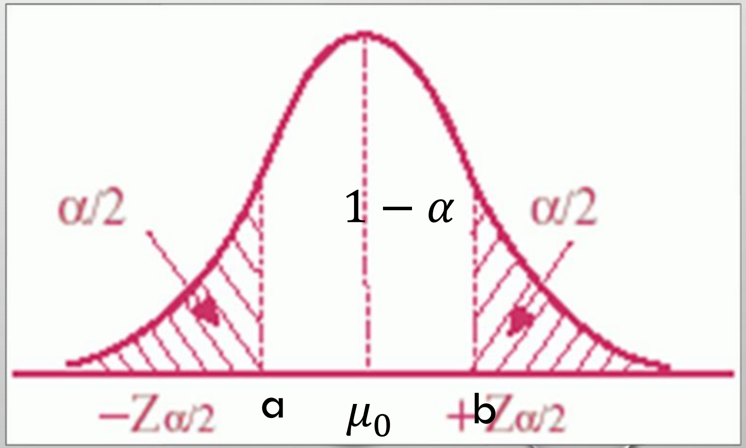
La región de aceptación está dada por: $P\left(-z_{\frac{\alpha}{2}} < \frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} < z_{\frac{\alpha}{2}}\right) = 1-\alpha$

Si $-z_{\frac{\alpha}{2}} < z < z_{\frac{\alpha}{2}}$, no se rechaza $H_0$. El rechazo de $H_0$, implica la aceptación de la hipótesis alternativa $\mu \neq \mu_0$

La región de aceptación en términos del promedio calculado $\bar{x}$: rechazar $H_0\ si\ \bar{x} > b$ ó $\bar{x} < a$, donde

$$a = \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad \mathrm{b} = \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Región crítica para la hipótesis alternativa $\mu \neq \mu_0$



# RELACIÓN CON LA ESTIMACIÓN DEL INTERVALO DE CONFIANZA

La estimación del intervalo de confianza incluye el cálculo de límites para los que es «razonable» que el parámetro en cuestión se encuentre dentro de ellos

Para el caso de una sola media poblacional μ con $\sigma^2$ conocida, la estructura de l aprueba de hipótesis y de la estimación del intervalo de confianza se basan en la variable aleatoria

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

Resulta que la prueba de $H_0\colon \mu = \mu_0$ contra $H_1\colon \mu \neq \mu_0$ a un nivel de significancia $\alpha$ *es equivalente a calcular un intervalo de confianza de 100(1-*$\alpha$*)% sobre* $\mu$ *y rechazar* $H_0$ *si* $\mu_0$ *no está dentro del intervalo de confianza.* Si $\mu_0$ está dentro del intervalo de confianza, la hipótesis no se rechaza

$$\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

# PRUEBAS PARA DIFERENCIAS MUESTRALES: DIFERENCIAS ENTRE MEDIAS

Sean $\bar{X}_1$ y $\bar{X}_2$ las medias muestrales de muestras grandes de tamaños $N_1$ y $N_2$ obtenidas de poblaciones cuyas medias son $\mu_1$ y $\mu_2$ y cuyas desviaciones estándar son $\sigma_1$ y $\sigma_2$, respectivamente

Considérese que la hipótesis nula de *no hay diferencia* entre las dos medias poblacionales (es decir $\mu_1 = \mu_2$)

La distribución muestral de las diferencias entre las medias es aproximadamente normal con media y desviación estándar dadas por:

$$\mu_{\bar{X}1-\bar{X}2} = 0 \quad y \quad \sigma_{\bar{X}1-\bar{X}2} = \sqrt{\frac{{\sigma_1}^2}{N_1} + \frac{{\sigma_2}^2}{N_2}}$$

Donde si es necesario, se pueden usar las desviaciones estándar muestrales como estimaciones de $\sigma_1$ y $\sigma_2$

Empleando la variable estandarizada, o puntuación $z$, dada por:

$$z = \frac{\bar{X}_1 - \bar{X}_2 - 0}{\sigma_{\bar{X}1-\bar{X}2}} = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}1-\bar{X}2}}$$

Se puede probar la hipótesis nula contra la hipótesis alternativa (o la significancia de la diferencia observada) a un nivel de significancia apropiado

# EJEMPLO

- UN DISEÑADOR DE PRODUCTOS ESTÁ INTERESADO EN REDUCIR EL TIEMPO DE SECADO DE UNA PINTURA TAPA POROS. SE PRUEBAN DOS FÓRMULAS DE PINTURA; LA FÓRMULA 1 TIENE EL CONTENIDO QUÍMICO ESTÁNDAR, Y LA FÓRMULA 2 TIENE UN NUEVO INGREDIENTE SECANTE QUE DEBE REDUCIR EL TIEMPO DE SECADO. DE LA EXPERIENCIA SE SABE QUE LA DESVIACIÓN ESTÁNDAR DEL TIEMPO DE SECADO ES OCHO MINUTOS, Y ESTA VARIABILIDAD INHERENTE NO DEBE VERSE AFECTADA POR LA ADICIÓN DEL NUEVO INGREDIENTE. SE PINTAN DIEZ ESPECÍMENES CON LA FÓRMULA 1, Y OTROS DIEZ CON LA FÓRMULA 2. LOS DOS TIEMPOS PROMEDIO DE SECADO MUESTRALES SON 121 MIN Y 112 MIN RESPECTIVAMENTE. ¿A QUÉ CONCLUSIONES PUEDE LLEGAR EL DISEÑADOR DEL PRODUCTO SOBRE LA EFICACIA DEL NUEVO INGREDIENTE, UTILIZANDO $\alpha = 0{,}05$?

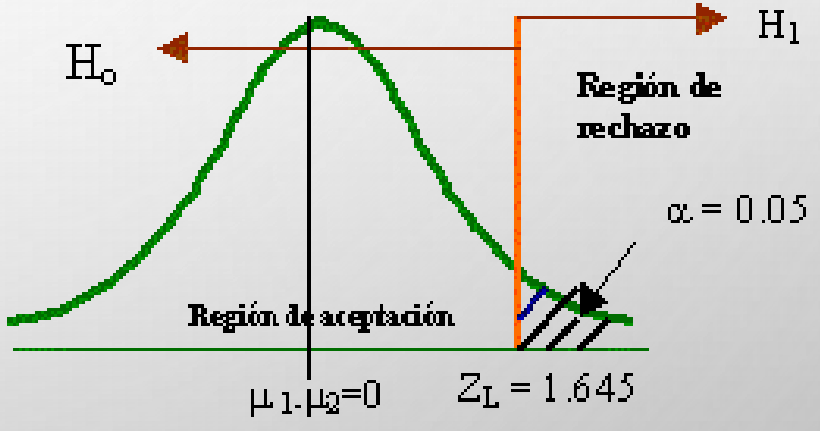
# EJEMPLO: SOLUCIÓN

- SE TRATA DE UNA DISTRIBUCIÓN MUESTRAL DE DIFERENCIA DE MEDIAS CON DESVIACIÓN ESTÁNDAR CONOCIDA
- *DATOS:* $\sigma_1 = \sigma_2 = 8 \;\; \bar{x}_1 = 121min \;\; \bar{x}_2 = 112min \;\; n_1 = n_2 = 10 \;\; \alpha = 0{,}05$
- *ENSAYO DE HIPÓTESIS:* $H_0: \mu_1 - \mu_2 = 0 \;\; H_1: \mu_1 - \mu_2 > 0$

SE DESEA RECHAZAR $H_o$ SI EL NUEVO INGREDIENTE DISMINUYE EL TIEMPO PROMEDIO DE SECADO, POR ESO SE PONE LA DIFERENCIA MAYOR A CERO O SEA POSITIVA PARA PODER PROBAR QUE $\mu_2$ ES MENOR QUE $\mu_1$

- *REGLA DE DECISIÓN:* SI $z_R \leq 1{,}645$ NO SE RECHAZA $H_0$, SI $z_R > 1{,}645$ SE RECHAZA $H_0$
- CÁLCULOS

$$z_R = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(121 - 112) - 0}{\sqrt{\frac{8^2}{10} + \frac{8^2}{10}}} = 2{,}52$$




- *JUSTIFICACIÓN Y DECISIÓN*

PUESTO QUE 2.52>1.645, SE RECHAZA $H_o$, Y SE CONCLUYE CON UN NIVEL DE SIGNIFICANCIA DE 0.05 QUE LA ADICIÓN DEL NUEVO INGREDIENTE A LA PINTURA SI DISMINUYE DE MANERA SIGNIFICATIVA EL TIEMPO PROMEDIO DE SECADO

# UNA MUESTRA: PRUEBA SOBRE UNA SOLA PROPORCIÓN

Consideraremos el problema de probar la hipótesis de que la proporción de éxitos de un experimento binomial es igual a algún valor específico

Probaremos la hipótesis nula $H_0$ que $p=p_0$, donde $p$ es el parámetro de la distribución binomial

La hipótesis alternativa puede ser una de las alternativas unilaterales o bilaterales usuales:

$p < p_0 \qquad p > p_0$

$p \neq p_0$

La variable aleatoria sobre la que basamos nuestro criterio de decisión es la variable aleatoria binomial $X$, también podríamos usar sólo la estadística $\hat{p} = \frac{X}{n}$

Los valores de $X$ que están lejos de la media $\mu = np_0$ conducirá al rechazo de la hipótesis nula

# UNA MUESTRA: PRUEBA SOBRE UNA SOLA PROPORCIÓN

- AL TRATAR CON MUESTRAS PEQUEÑAS, NUESTRA DECISIÓN SE BASA EN VALORES *P*

Para probar la hipótesis $H_o: p = p_o \qquad H_1: p < p_0$

Utilizamos la distribución binomial para calcular el valor *P*

$$P = P(X \leq x \; cuando \; p = p_0)$$

El valor $x$ es el número de éxitos en nuestra muestra de tamaño *n*. Si este valor *P* es menor o igual a $\alpha$, nuestra prueba es significativa en el nivel $\alpha$ y rechazamos $H_0$ a favor de $H_1$

Para probar la hipótesis $H_o: p = p_o \qquad H_1: p > p_0$

En el nivel de significancia $\alpha$, calculamos

$$P = P(X \geq x \; cuando \; p = p_0)$$

y rechazamos $H_0$ a favor de $H_1$ si este valor *P* es menor o igual a $\alpha$

Para probar la hipótesis $H_o: p = p_o \qquad H_1: p \neq p_0$

al nivel de significancia $\alpha$, calculamos

$$P = 2P(X \leq x \; cuando \; p = p_0) \quad si \; x < np_0 \quad o \qquad P = 2P(X \geq x \; cuando \; p = p_0)$$

Si $x > np_0$ y se rechaza $H_0$ a favor de $H_1$ si el valor *P* calculado es menor o igual a $\alpha$

# UNA MUESTRA: PRUEBA SOBRE UNA SOLA PROPORCIÓN

Los pasos para probar una hipótesis nula acerca de una proporción contra varias alternativas mediante el uso de variables binomiales son los siguientes:

$H_0\!: \; p = p_0$

$H_1\!: \; las\ alternativas\ son\ p < p_0 \qquad p > p_0 \qquad p \neq p_0$

$Elegir\ un\ nivel\ de\ significancia\ igual\ a\ \alpha$

$Estadística\ de\ prueba\!: variable\ binomial\ X\ con\ p = p_0$

$Cálculos\!: Encontrar\ x, el\ número\ de\ éxitos, y\ calcular\ el\ valor\ P\ apropiado$

$Decisión\!: Extraer\ las\ conclusiones\ apropiadas\ basadas\ en\ el\ valor\ P$

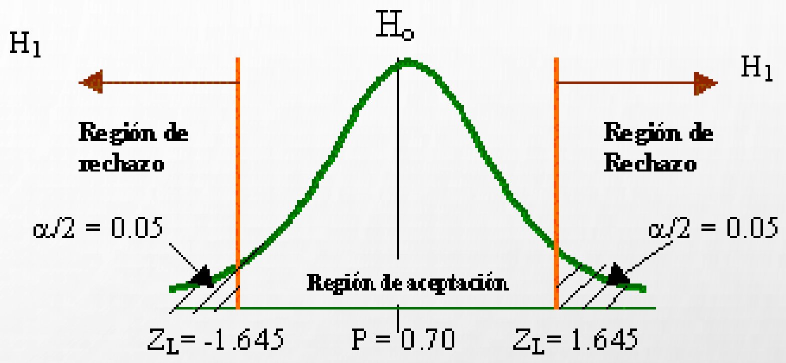
- UN CONSTRUCTOR AFIRMA QUE SE INSTALAN BOMBAS DE CALOR EN 70% DE TODAS LAS CASAS QUE SE CONSTRUYEN HOY EN DÍA EN UNA CIUDAD DE EEUU. ¿ESTARÍA DE ACUERDO CON ESTA AFIRMACIÓN SI UNA INVESTIGACIÓN DE CASAS NUEVAS EN ESA CIUDAD MUESTRA QUE 8 DE 15 TIENEN INSTALADAS BOMBAS DE CALOR?
- UTILICE UN NIVEL DE SIGNIFICANCIA DE 0,10

# EJEMPLO

***SOLUCIÓN***

- SE TRATA DE UNA DISTRIBUCIÓN MUESTRAL DE PROPORCIONES
- *DATOS:*

$P = 0{,}70 \quad p = \frac{8}{15} = 0{,}5333 \quad n = 15 \quad \alpha = 0{,}10$




- *ENSAYO DE HIPÓTESIS*

$H_0 : P = 0{,}70 \quad H_1 : P \neq 0{,}70$

- *REGLA DE DECISIÓN*

SI $-1{,}645 \leq z_R \leq 1{,}645$ NO SE RECHAZA $H_0$, SI $z_R < -1{,}645$ Ó SI $z_R > 1{,}645$ SE RECHAZA $H_0$

- *CÁLCULOS*

$$z_R = \frac{p - P}{\sqrt{\frac{Pq}{n}}} = \frac{0{,}533 - 0{,}70}{\sqrt{\frac{0{,}70 \cdot 0{,}30}{15}}} = -1{,}41$$

- *JUSTIFICACIÓN Y DECISIÓN*

COMO $-1{,}645 \leq \mathbf{-1{,}41} \leq 1{,}645$ NO SE RECHAZA $H_0$ Y SE CONCLUYE CON UN NIVEL DE SIGNIFICANCIA DE 0.10 QUE LA AFIRMACIÓN DEL CONSTRUCTOR ES CIERTA

# UNA MUESTRA: PRUEBA SOBRE UNA SOLA PROPORCIÓN

Para *n* grande, se requieren procedimientos de aproximación

Cuando el valor hipotético $p_0$ es muy cercano a 0 o a 1, se puede utilizar la distribución de Poisson, con parámetro $\mu = np_0$

La aproximación de la curva normal, con parámetros $\mu = np_0$ y $\sigma^2 = np_0q_0$, se prefiere para *n* grande y es muy precisa mientras que $p_0$ no esté extremadamente cerca de 0 o de 1

Si utilizamos la aproximación normal, el ***valor z para probar p=p₀*** está dado por

$$z = \frac{x - np_0}{\sqrt{np_0q_0}}$$

Es un valor de la variable normal estándar *Z*

Para una prueba de dos colas al nivel de significancia $\alpha$, la región crítica es:

$$z < -z_{\alpha/2} \quad \text{y} \quad z > z_{\alpha/2}$$

Para la alternativa unilateral $p < p_0$, la región crítica es $z < -z_\alpha$, y para la alternativa $p > p_0$ , la región crítica es $z > z_\alpha$

# PRUEBAS PARA DIFERENCIAS MUESTRALES: DIFERENCIAS ENTRE PROPORCIONES

Sean $P_1$ y $P_2$ las proporciones muestrales de muestras grandes de tamaños $N_1$ y $N_2$ obtenidas de poblaciones cuyas proporciones son $p_1$ y $p_2$

Considérese que la hipótesis nula de *no hay diferencia* entre estos parámetros poblacionales (es decir $p_1 = p_2$)

La distribución muestral de las diferencias entre las proporciones es aproximadamente normal con media y desviación estándar dadas por:

$$\mu_{P1-P2} = 0 \quad y \quad \sigma_{P1-P2} = \sqrt{pq\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}$$

Donde $p = \frac{N_1 P_1 + N_2 P_2}{N_1 + N_2}$ se usa como estimación de la proporción poblacional y donde $q = 1 - p$

Empleando la variable estandarizada:

$$z = \frac{P_1 - P_2 - 0}{\sigma_{P1-P2}} = \frac{P_1 - P_2}{\sigma_{P1-P2}}$$

Se puede probar la diferencia observada a nivel de significancia apropiado y con esto probar la hipótesis nula

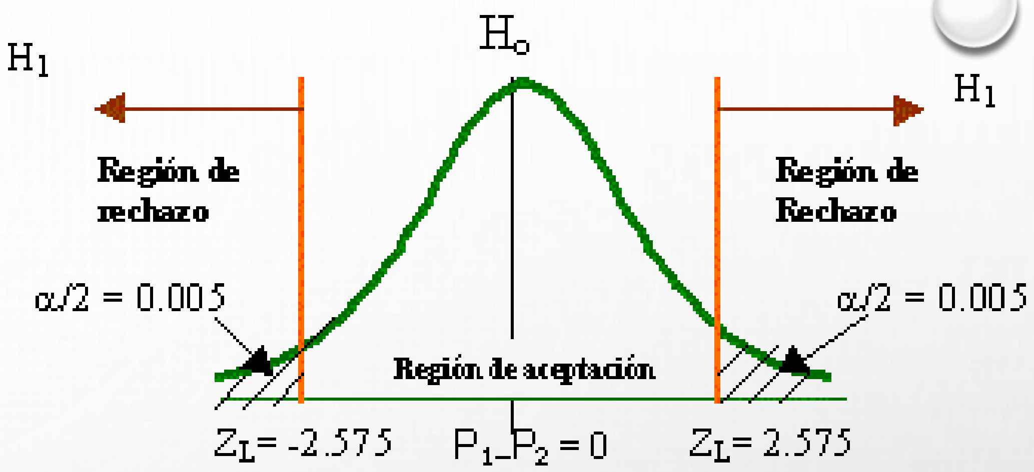
# EJEMPLO

- SE EVALÚAN DOS TIPOS DIFERENTES DE SOLUCIONES PARA PULIR, PARA SU POSIBLE USO EN UNA OPERACIÓN DE PULIDO EN LA FABRICACIÓN DE LENTES INTRAOCULARES UTILIZADOS EN EL OJO HUMANO DESPUÉS DE UNA CIRUGÍA DE CATARATAS. SE PULEN 300 LENTES CON LA PRIMERA SOLUCIÓN Y, DE ÉSTOS, 253 NO PRESENTARON DEFECTOS INDUCIDOS POR EL PULIDO. DESPUÉS SE PULEN OTROS 300 LENTES CON LA SEGUNDA SOLUCIÓN, DE LOS CUALES 196 RESULTAN SATISFACTORIOS. ¿EXISTE ALGUNA RAZÓN PARA CREER QUE LAS DOS SOLUCIONES PARA PULIR SON DIFERENTES? UTILICE $\alpha = 0{,}01$

# EJEMPLO: SOLUCIÓN

- SE TRATA DE UNA DISTRIBUCIÓN MUESTRAL DE DIFERENCIA DE PROPORCIONES
- *DATOS*: $p_1 = \frac{253}{300} = 0{,}8433 \;\; p_2 = \frac{196}{300} = 0{,}6533 \;\; n_1 = n_2 = 300 \;\; \alpha = 0{,}01$
- *ENSAYO DE HIPÓTESIS*: $H_0: P_1 - P_2 = 0 \;\; H_1: P_1 - P_2 \neq 0$
- *REGLA DE DECISIÓN*: SI $-2{,}575 \leq z_R \leq 2{,}575$ NO SE RECHAZA $H_0$,

  SI $z_R < -2{,}575$ Ó SI $z_R > 2{,}575$ SE RECHAZA $H_0$
- *CÁLCULOS*: $z_R = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 \cdot q_1}{n_1} + \frac{P_2 \cdot q_2}{n_2}}}$




EN ESTA FÓRMULA SE PUEDE OBSERVAR QUE EN EL DENOMINADOR SE TIENEN A LAS PROPORCIONES POBLACIONALES O SEA LOS PARÁMETROS, LOS CUALES NO SE CONOCEN, POR LO QUE EN EL ENSAYO DE HIPÓTESIS LA FÓRMULA PARA PODER CALCULAR LA $Z_R$ CAMBIA, ESTIMANDO A EL PARÁMETRO COMÚN P DE LA SIGUIENTE FORMA: $p = \frac{x_1 + x_2}{n_1 + n_2}$, ENTONCES LA FÓRMULA DE $Z_R$ QUEDARÍA DE LA SIGUIENTE MANERA: $z_R = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{P \cdot q \cdot \left(\frac{1}{n_1} + \frac{1}{n_1}\right)}}$

SE CALCULARÁ EL VALOR DE P:

$$p = \frac{x_1 + x_2}{n_1 + n_2} = \frac{253 + 196}{300 + 300} = 0{,}7483$$

$$z_R = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{P \cdot q \cdot \left(\frac{1}{n_1} + \frac{1}{n_1}\right)}} = \frac{(0{,}8433 - 0{,}6533) - 0}{\sqrt{0{,}7483 \cdot 0{,}2517 \cdot \left(\frac{1}{300} + \frac{1}{300}\right)}} = 5{,}36$$

- *JUSTIFICACIÓN Y DECISIÓN*: PUESTO QUE 5.36>2.575, SE RECHAZA LA HIPÓTESIS NULA Y SE CONCLUYE CON UN NIVEL DE SIGNIFICANCIA DE 0.01 QUE LOS DOS FLUIDOS PARA PULIR SON DIFERENTES