FCFM
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

INFERENCIA ESTADISTICA

Tema 5

- ✓ Pruebas de hipótesis

5.1 Hipótesis Estadística.

Sea $(X_1, X_2, \dots, X_n)$ una m.a. de una distribución $f(x;\theta)$.

Definición: Una Hipótesis estadística es una afirmación o conjetura acerca de la distribución de una variable aleatoria.

Ejemplos:

- ✓ Una compañía recibe un gran cargamento de piezas. Solo acepta el envío si no hay mas de un 5% de piezas defectuosas. ¿Como tomar una decisión sin verificar todas las piezas?
- ✓ Se quiere saber si una propuesta de reforma legislativa es acogida de igual forma por hombres y mujeres. ¿Como se puede verificar esa conjetura?

5.2 Hipótesis Simple y Compuesta.

Tipos Generales de Hipótesis:

a) **Simple** si es pesifica por completo a la distribución en estudio y/o aquella que especifica un único valor para el parámetro. Ejemplos: $H: \theta = 0,\ H: \theta = -6,\ H: \theta = 10,\ etc.$

b) **Compuesta** si no especifica por completo a la distribución y/o aquella que especifica un intervalo de valores para el parámetro. Ejemplos: $H: \theta \geq 0,\ H: 1 \leq \theta \leq 6,\ etc.$

$$\text{H}ipótesis\ Unilateral\text{: } H: \theta \leq 6, H: 0 \leq \theta, etc.$$

$$\text{H}ipótesis\ Bilateral\text{: } H: \theta \neq 6 \Leftrightarrow\ H: \theta < 6\ y\ \theta > 6$$

5.2 Hipótesis Simple y Compuesta.

Contrastes de Hipótesis

$$H_0\text{: Hipótesis } Nula\ Vs\ H_1\text{: Hipótesis } Alternativa$$
$$Simple\ Vs\ Simple$$
$$Simple\ Vs\ Compuesta$$
$$Compuesta\ Vs\ Simple$$
$$Compuesta\ Vs\ Compuesta$$

Ejemplo: Sea x una v.a. con distribución $Ber(\theta)$. Donde lo que interesa contrastar:

$$H_0: \theta = {}^{1}/_{2}\ Vs\ H_0: \theta \neq {}^{1}/_{2}$$
$$Simple\ Vs\ Compuesta$$

5.2 Hipótesis Simple y Compuesta.

Proceso del Contraste de Hipótesis

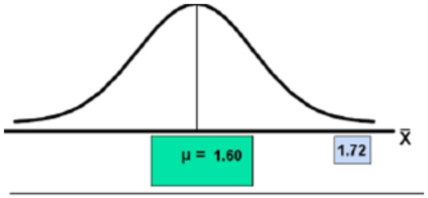
Hipótesis: La Altura media de la población (Colonia, Cuidad, País, etc.) es 1.68 m $(H_0: \mu = 1.68)$

Población

Muestra Aleatoria Simple

La Media Muestral es 1.75 m $(\bar{X} = 1.75)$

¿Es probable que 1.68 m $\bar{X} = 1.75\ si\ \mu = 1.68$



Si no lo es, $\boldsymbol{Rechazaremos\ H_0}$

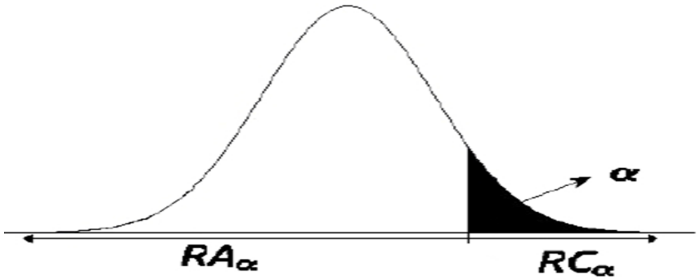
5.3 Región crítica.

Definición 1: Un región de rechazo es un subconjunto de valores de una m.a. para los cuales se rechaza la hipótesis nula.

Definición 2: Un contraste de hipótesis es una regla que determina, a un cierto nivel de significación, α, para que valores de la muestra se rechaza o no se rechaza la hipótesis nula.

Se trata de determinar a un nivel de significación α, una región critica o de rechazo RC_α, y una región de aceptación RA_α.

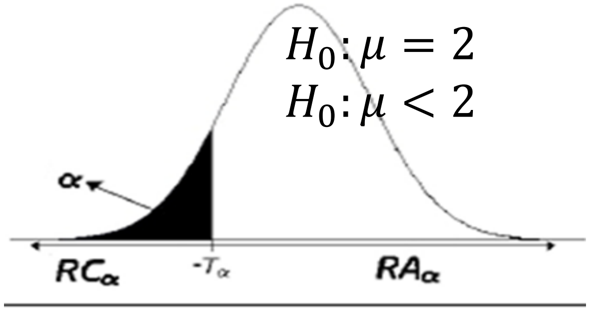
$$\Omega = RC_\alpha \cup RA_\alpha$$
$$RC_\alpha \cup RA_\alpha = \emptyset$$

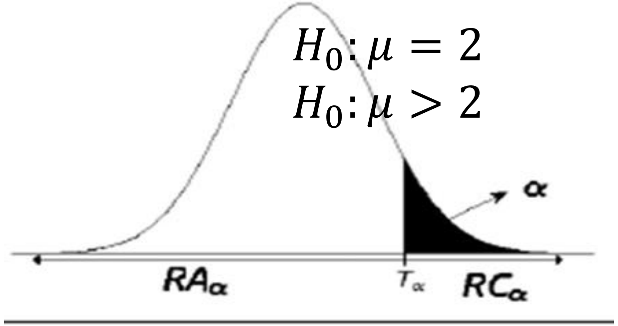


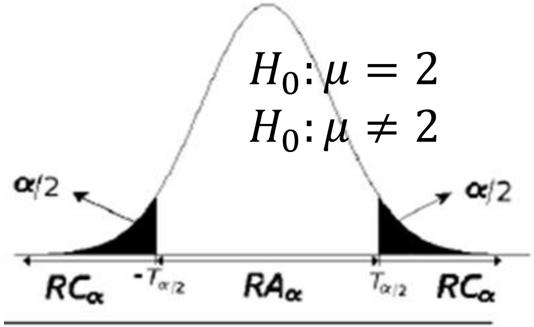
5.3 Región crítica.

El estadístico del contraste es un estadístico que se construye a partir de un estimador del parámetro y cuya distribución bajo H_0 es conocida. El nivel de significación es la probabilidad de que, bajo H_0, el estadístico del contraste tome valores en la RC_α.

Ejemplo: Sea $X \sim N(\mu, 5)$. Queremos hacer contrastes sobre la media poblacional μ. Estadístico (común para los tres contrastes): $T = \frac{\bar{X}-2}{5/\sqrt{n}} \sim N(0,1)$







5.4 Errores Tipo I y II.

Definición: El error tipo I se comete cuando **Si se rechaza la hipótesis nula $\boldsymbol{H_0}$** siendo esta verdadera. A la probabilidad de cometer este error se le denota por letra α, es decir, $\alpha = P(\text{Error de tipo I}) = P("Rechazar\ H_0"/"H_0\ es\ verdadera")$

A α se le llama también tamaño de la región critica, tamaño de la región de rechazo, o nivel de significancia de la prueba.

Definición: El error tipo II se comete cuando **NO se rechaza la hipótesis nula $\boldsymbol{H_0}$** siendo esta falsa. A la probabilidad de cometer este error se le denota por letra β, es decir, $\beta = P(\text{Error de tipo II}) = P("No\ Rechazar\ H_0"/"H_0\ es\ falsa")$

5.4 Errores tipo I y II.

	H_0 cierta	H_0 falsa
Rechazar H_0	Error tipo I α	✓
No Rechazar H_0	✓	Error tipo II β

Observaciones:

1. Las probabilidades α y β no son complementarias.
2. Se busca que α y β sean pequeñas.
3. Se fijará un valor de α y buscaremos aquella posible región de rechazo que tenga probabilidad β más pequeña, "Región optima".
4. Si H_0 es simple, y suponiendo que es verdadera, la distribución en estudio queda completamente especificada y α puede calcularse de manera exacta.

5.5 Pruebas de Hipótesis

Prueba de Hipótesis sobre una proporción poblacional p

1. Las observaciones muestrales son una muestra aleatoria simple.
2. Se satisfacen las condiciones para una distribución binomial. (Existe un número fijo de ensayos independientes con probabilidades constantes, y cada ensayo tiene dos categorías de resultados de "éxito" y "fracaso").
3. Se satisfacen las condiciones $np \geq 5$ y $np \leq 5$, por lo tanto, la distribución binomial de proporciones muestrales puede aproximarse con una distribución normal, con $\mu = np$ y $\sigma = \sqrt{npq}$. Observe que p es la proporción supuesta que se utiliza en la aseveración y no la proporción muestral.

5.5 Pruebas de Hipótesis

Prueba de Hipótesis sobre una proporción poblacional p

Notación:

$n = tamaño\ de\ muestra\ o\ número\ de\ ensayos$

$\hat{p} = \frac{x}{n}\ (proporción\ muestral)$

$p = proprción\ de\ la\ población\ (utilizada\ en\ la\ hipótesis\ nula)$

$q = 1 - p$

Estadístico de Prueba:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis sobre una proporción poblacional p

Ejemplo: Se aplico una encuesta a un grupo de empleados, para saber quien obtuvo empleo por medio de redes de contactos, los siguientes resultados de la encuesta, fue de 703 empleados elegidos al azar, el 61% obtuvo trabajo por medio de redes de contactos. Utilice los datos muestrales, con un nivel de significancia de 0.05, para probar la aseveración de que más de la mitad de los empleados consiguen su trabajo por medio de redes de contactos.

$$H_0: \hat{p} = 0.5$$
$$H_1: \hat{p} > 0.5$$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.5 - 0.61}{\sqrt{\frac{(0.61)(0.39)}{703}}} = -5.98$$

$$valor\ de\ tabla\ de\ \alpha$$
$$z_\alpha = z_{.05} = 1.645$$

$$Rechazamos\ H_0\ SI:$$
$$z \leq -z_\alpha \Rightarrow -5.98 \leq -1.645$$
$$\therefore Rechazamos\ H_0$$

$$Otra\ forma\ z = 5.83$$
$$Valor\ de\ tabla = Valor\ P = 1 - .9999 = .0001$$
$$Como\ 0.001\ es\ menor\ a\ 0.05$$
$$\therefore Rechazamos\ H_0$$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis respecto de una media σ conocida

1. La muestra es aleatoria simple
2. Se conoce el valor de la desviación estándar poblacional σ.
3. Se satisface una o ambas de las siguientes condiciones: la población se distribuye normalmente y $n > 30$.

Estadístico de Prueba:

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma / \sqrt{n}}$$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis respecto de una media σ conocida

Ejemplo: Se aplico conjunto de pesos, de 13 dulces M&M rojos elegidos al azar de una bolsa que contiene 465 dulces. La desviación estándar de los pesos de todos los dulces M&M que están en la bolsa es $\sigma = 0.0565g$. A continuación se presentan los pesos muestrales (en gramos), que tienen una media de $\bar{x} = 0.8635g$. En la bolsa se afirma que el peso neto del contenido es de 396.9*g*, de manera que los dulces M&M deben tener un peso medio de al menos 396.9/465=0.8535g para dar la cantidad anunciada. Utilice los datos muestrales con un nivel de significancia de 0.05 para probar la aseveración de un gerente de producción de que los dulces M&M en realidad tienen una media mayor que 0.8535*g*, por lo que los consumidores están recibiendo una cantidad mayor de la indicada en la etiqueta.

0.751, 0.841, 0.856, 0.799. 0.966, 0.859, 0.857, 0.942, 0.873, 0.809, 0.890, 0.878, 0.905

5.5 Pruebas de Hipótesis

Prueba de Hipótesis respecto de una media σ conocida

$$H_0: \mu = 0.8535$$
$$H_1: \mu > 0.8535$$

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}} = \frac{0.8635 - 0.8535}{\frac{0.0565}{\sqrt{13}}} = 0.64$$

$$valor\ de\ tabla\ de\ \alpha$$
$$z_\alpha = z_{.05} = 1.645$$

$$Rechazamos\ H_0\ SI:$$
$$z \geq z_\alpha \Rightarrow 0.64 \leq 1.645$$
$$\therefore No\ Rechazamos\ H_0$$

$$Otra\ forma\ z = 0.64$$
$$Valor\ de\ tabla = Valor\ P = 1 - 0.7389 = 0.2611$$
$$Como\ 0.2611\ es\ mayor\ a\ 0.05$$
$$\therefore No\ Rechazamos\ H_0$$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis respecto de una media *s* desconocida

1. La muestra es aleatoria simple
2. Se desconoce el valor de la desviación estándar poblacional σ.
3. Se satisface una o ambas de las siguientes condiciones: la población se distribuye normalmente y $n < 30$.

Estadístico de Prueba:

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma / \sqrt{n}}$$

Se utiliza $gl = n - 1$ para el numero de grados de libertad.

5.5 Pruebas de Hipótesis

Prueba de Hipótesis respecto de una media *s* desconocida

Ejemplo: Control de calidad de los dulces M&M El conjunto de datos 13 del apéndice B incluye los pesos de 13 dulces M&M rojos, elegidos al azar de una bolsa que contiene 465 M&M. A continuación se presentan los pesos (en gramos), los cuales tienen una media de $\bar{x} = 0.8635g$ y una desviación estándar de $s = 0.0576g$. En el empaque se afirma que el peso neto del contenido es 396.9 g, de manera que los M&M deben tener un peso medio de al menos 396.9/465 = 0.8535 g para dar la cantidad anunciada. Utilice los datos muestrales con un nivel de significancia de 0.05, para probar la aseveración que hizo un gerente de producción de que los M&M tienen en realidad una media mayor que 0.8535 g, de manera que los consumidores están recibiendo más que la cantidad indicada en la etiqueta.

0.751, 0.841, 0.856, 0.799. 0.966, 0.859, 0.857, 0.942, 0.873, 0.809, 0.890, 0.878, 0.905

FCFM
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

5.5 Pruebas de Hipótesis

Prueba de Hipótesis respecto de una media *s* desconocida

$$H_0: \mu = 0.8535$$
$$H_1: \mu > 0.8535$$

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}} = \frac{0.8635 - 0.8535}{\frac{0.0576}{\sqrt{13}}} = 0.626$$

$$valor\ de\ tabla\ de\ \alpha$$
$$t_{\alpha,n-1} = t_{.05,12-1} = 1.7823$$

$$Rechazamos\ H_0\ SI:$$
$$t \geq t_{\alpha,n-1} \Rightarrow 0.626 \leq 1.782$$
$$\therefore\ No\ Rechazamos\ H_0$$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis respecto de una desviación estándar o de una varianza

1. La muestra es aleatoria simple
2. La población tiene una distribución normal.

Estadístico de Prueba:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

Se utiliza $gl = n - 1$ para el numero de grados de libertad.

5.5 Pruebas de Hipótesis

Prueba de Hipótesis respecto de una desviación estándar o de una varianza

Ejemplo: Los ingenieros de control de calidad desean asegurarse de que un producto tenga una media aceptable, pero también quieren producir artículos con una calidad consistente, eliminando los defectos. La Compañía Pepsi ha fabricado latas de bebidas de cola con cantidades que tienen una desviación estándar de 0.051 onzas. Se prueba una nueva máquina embotelladora, y una muestra aleatoria simple de 24 latas produce las cantidades (en onzas) que se listan a continuación. (Las 24 cantidades tienen una desviación estándar de $s = 0.039\ oz$). Utilice un nivel de significancia de 0.05 para probar la aseveración de que las latas de bebidas de cola de la nueva máquina tienen cantidades con una desviación estándar menor que 0.051 oz.

11.98, 11.98, 11.99, 11.98, 11.90, 12.02, 11.99, 11.93, 12.02, 12.02, 12.02, 11.98, 12.01, 12.00, 11.99, 11.95, 11.95, 11.96, 11.96, 12.02, 11.99, 12.07, 11.93, 12.05

5.5 Pruebas de Hipótesis

Prueba de Hipótesis respecto de una desviación estándar o de una varianza

$H_0: \sigma = 0.051$
$H_1: \sigma < 0.051$

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(24-1)(0.039)^2}{0.051^2} = 13.450$$

$$valor\ de\ tabla\ de\ \alpha$$
$$\chi^2_{1-\alpha,n-1} = \chi^2_{0.95,23} = 13.09$$

$$Rechazamos\ H_0\ SI:$$
$$\chi^2 \leq \chi^2_{1-\alpha,n-1} \Rightarrow 13.450 \geq 13.09$$
$$\therefore No\ Rechazamos\ H_0$$

Regiones de Rechazo

$\chi^2 \geq \chi^2_{\alpha,\,n-1}$
$\chi^2 \leq \chi^2_{1-\alpha,\,n-1}$
$\chi^2 \leq \chi^2_{1-\alpha/2,\,n-1}$ ó $\chi^2 >= \chi^2_{\alpha/2,\,n-1}$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis sobre dos proporciones

1. Tenemos proporciones de dos muestras aleatorias simples que son independientes. (Las muestras son independientes si los valores muestrales seleccionados de una población no están relacionados ni apareados de alguna forma con los valores muestrales seleccionados de la otra población).
2. Para ambas muestras, el número de éxitos es de al menos 5 y el número de fracasos es de al menos 5.

Notación:

$p_1 = proporción\ poblacional, n_1 = tamaño\ muestral\ y\ x_1$
$= numero\ de\ exitos\ en\ la\ muestra$

$\hat{p}_1 = \frac{x_1}{n_1}\ (proporción\ muestral)$

$\hat{q} = 1 - \hat{p}_1$

Se adjuntan los significados correspondientes a $p_2, n_2, x_2, \hat{p}_2\ y\ \hat{q}_2$, que provienen de la población 2.

5.5 Pruebas de Hipótesis

Prueba de Hipótesis sobre dos proporciones
Proporción muestral agrupada:

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}, donde\ \bar{q} = 1 - \bar{p}$$

Estadístico de Prueba para dos proporciones (con $\boldsymbol{H_0: p_1 = p_2}$):

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis sobre dos proporciones

Ejemplo: ¿La cirugía es mejor que el entablillado? El problema incluye los resultados de una prueba clínica en la que se dio tratamiento a pacientes con síndrome de túnel carpiano, y los resultados se resumen en la tabla. Utilice los datos muestrales de la tabla, con un nivel de significancia de 0.05 para probar la aseveración de que la tasa de éxito de la cirugía es mejor que la tasa de éxito del entablillado.

	Cirugía	Entablillado
Éxito un año Después	67	60
Total de personas	73	83
Porcentaje de Éxitos	**92%**	**72%**

5.5 Pruebas de Hipótesis

Prueba de Hipótesis sobre dos proporciones

$$H_0: p_1 = p_2$$
$$H_1: p_1 > p_2$$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{67 + 60}{73 + 83} = 0.811410256$$
$$\bar{q} = 1 - \bar{p} = 1 - 0.811410256 = 0.18589744$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$$

$$= \frac{\left(\frac{67}{73} - \frac{60}{83}\right) - 0}{\sqrt{\frac{(0.811410256)(0.18589744)}{73} + \frac{(0.811410256)(0.18589744)}{n_2}}} = 3.12$$

$$valor\ de\ tabla\ de\ \alpha$$
$$z_\alpha = z_{.05} = 1.645$$

$$Rechazamos\ H_0\ SI:$$
$$z \geq z_\alpha \Rightarrow 3.12 \geq 1.645$$
$$\therefore Rechazamos\ H_0$$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis de muestras independientes con s_1 y s_2 desconocidas y sin suposición de igualdad

1. $\boldsymbol{s_1}$ **y** $\boldsymbol{s_2}$ se desconocen y no se hace una suposición sobre la igualdad de $\boldsymbol{s_1}$ **y** $\boldsymbol{s_2}$.
2. Las dos muestras son independientes.
3. Ambas muestras son aleatorias simples.
4. Cualquiera o ambas de estas condiciones se satisfacen: los dos tamaños muestrales son grandes (con $n_1 < 30$ y $n_2 < 30$) o ambas muestras provienen de poblaciones que tienen distribuciones normales. (En muestras pequeñas, el requisito de normalidad es menos estricto, en el sentido de que los procedimientos se comportan bien siempre y cuando no existan valores extremos ni grandes sesgos).

5.5 Pruebas de Hipótesis

Prueba de Hipótesis de muestras independientes con s_1 y s_2 desconocidas y sin suposición de igualdad

Estadístico de Prueba:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Grados de Libertad:

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

5.5 Pruebas de Hipótesis

Prueba de Hipótesis de muestras independientes con s_1 y s_2 desconocidas y sin suposición de igualdad

Ejemplo: Discriminación por edad, La UNAM realizo un concurso de promoción. A continuación se muestran las edades de los solicitantes que tuvieron éxito y de los que no tuvieron. Algunos de los solicitantes que no tuvieron éxito para obtener la promoción se quejaron de que hubo discriminación por edad en la competencia. Maneje los datos como muestras de poblaciones más grandes y utilice un nivel de significancia de 0.05 para poner a prueba la aseveración de que los solicitantes sin éxito provienen de una población con una edad media mayor que la de los solicitantes exitosos. Con base en el resultado, ¿parece haber discriminación por la edad?

Edades de solicitantes sin éxito									
34	37	37	38	41	42	43	44	44	45
45	45	46	48	49	53	53	54	54	55
56	57	60							

Edades de solicitantes con éxito									
27	33	36	37	38	38	39	42	42	43
43	44	44	44	45	45	45	45	46	46
47	47	48	48	49	49	51	51	52	54

5.5 Pruebas de Hipótesis

Prueba de Hipótesis de muestras independientes con s_1 y s_2 desconocidas y sin suposición de igualdad

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 > \mu_2$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(47.0 - 43.9) - 0}{\sqrt{\frac{7.2^2}{23} + \frac{5.9^2}{30}}} = 1.678$$

$$gl = v = \frac{\left({s_1^2}/{n_1} + {s_2^2}/{n_2}\right)^2}{\frac{\left({s_1^2}/{n_1}\right)^2}{n_1 - 1} + \frac{\left({s_2^2}/{n_2}\right)^2}{n_2 - 1}} = \frac{\left(\frac{7.2^2}{23} + \frac{5.9^2}{30}\right)^2}{\frac{\left(\frac{7.2^2}{23}\right)^2}{23 - 1} + \frac{\left(\frac{5.9^2}{30}\right)^2}{30 - 1}} = 41.868 \approx 42$$

$$valor\ de\ tabla\ de\ \alpha$$
$$t_{\alpha,v} = t_{.05,42} = 1.682$$

$$Rechazamos\ H_0\ SI:$$
$$t \geq t_{\alpha,v} \Rightarrow 1.678 \leq 1.682$$
$$\therefore No\ Rechazamos\ H_0$$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis de muestras que supone $\boldsymbol{\sigma_1 = \sigma_2}$ y se agrupan las varianzas muestrales

1. Se desconocen las dos desviaciones estándar poblacionales, pero se supone que son iguales. Es decir, $\sigma_1 = \sigma_2$.
2. Las dos muestras son independientes.
3. Ambas muestras son aleatorias simples.
4. Cualquiera de estas condiciones (o ambas) se satisfacen: los dos tamaños muestrales son grandes con $n_1 < 30$ y $n_2 < 30$) o las dos muestras provienen de poblaciones que tienen distribuciones normales. (Para muestras pequeñas, el requisito de normalidad es menos estricto en el sentido de que los procedimientos funcionan bien siempre y cuando no existan valores extremos ni desviaciones de la normalidad demasiado pronunciadas).

5.5 Pruebas de Hipótesis

Prueba de Hipótesis de muestras que supone $\boldsymbol{\sigma_1 = \sigma_2}$ y se agrupan las varianzas muestrales

Estadístico de Prueba:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$donde\ s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}\ y\ gl = n_1 + n_2 - 2$$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis de muestras que supone $\boldsymbol{\sigma_1 = \sigma_2}$ y se agrupan las varianzas muestrales

Ejemplo: Se tienen las mediciones del nivel de hierro en la sangre de dos muestras de niños: un grupo de niños sanos y el otro padece fibrosis quística. Los datos obtenidos se dan en la siguiente tabla:

sanos	$n_1 = 9$	$\bar{x}_1 = 18.9$	$s_1^2 = 5.9^2$
enfermos	$n_2 = 13$	$\bar{x}_2 = 11.9$	$s_2^2 = 6.3^2$

Podemos asumir que las muestras provienen de poblaciones normales independientes con iguales varianzas. Es de interés saber si las dos medias del nivel de hierro en sangre son iguales o distintas. Utilizar $\alpha = 0.05$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis de muestras que supone $\boldsymbol{\sigma_1 = \sigma_2}$ y se agrupan las varianzas muestrales

Estadístico de Prueba:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(18.9 - 11.9) - 0}{6.14\sqrt{\frac{1}{9} + \frac{1}{13}}} = 2.63$$

$$s_p = \sqrt{\frac{(9-1)5.9^2 + (13-1)6.3^2}{(9-1) + (13-1)}} = 6.14$$

$$valor\ de\ tabla\ de\ \alpha$$
$$t_{\alpha/2, n_1+n_2-2} = t_{.025,20} = 2.085$$

$$Rechazamos\ H_0\ SI:$$
$$t \geq t_{\frac{\alpha}{2}, n_1+n_2-2} \Rightarrow 2.63 > 2.086$$
$$\therefore Rechazamos\ H_0$$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis de muestras independientes con $\boldsymbol{\sigma_1}$ y $\boldsymbol{\sigma_2}$ conocidas

1. Se conocen las dos desviaciones estándar poblacionales.
2. Las dos muestras son independientes.
3. Ambas muestras son aleatorias simples.
4. Cualquiera de estas condiciones (o ambas) se satisfacen: los dos tamaños muestrales son grandes con $n_1 > 30$ y $n_2 > 30$) o las dos muestras provienen de poblaciones que tienen distribuciones normales. (Para muestras pequeñas, el requisito de normalidad es menos estricto en el sentido de que los procedimientos funcionan bien siempre y cuando no existan valores extremos ni desviaciones de la normalidad demasiado pronunciadas).

5.5 Pruebas de Hipótesis

Prueba de Hipótesis de muestras independientes con σ_1 y σ_2 conocidas

Estadístico de Prueba:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis de muestras que supone $\boldsymbol{\sigma_1 = \sigma_2}$ y se agrupan las varianzas muestrales

Ejemplo: Un diseñador de productos está interesado en reducir el tiempo de secado de una pintura tapaporos. 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 8 minutos, y esta variabilidad no debe verse afectada por la adición del nuevo ingrediente. Se pintan 10 especímenes con la fórmula 1 y otros 10 con la fórmula 2. los tiempos promedio de secado muestrales fueron $\bar{x}_1 = 121\, minutos$ y $\bar{x}_2 = 112\, minutos$ respectivamente.

¿A qué conclusiones debe llegar el diseñador del producto sobre la eficacia del nuevo ingrediente utilizando α = 0.05?

FCFM
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

5.5 Pruebas de Hipótesis

Prueba de Hipótesis de muestras independientes con $\boldsymbol{\sigma_1}$ y $\boldsymbol{\sigma_2}$ conocidas

Estadístico de Prueba:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 > \mu_2$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(121 - 112) - 0}{\sqrt{\frac{8^2}{10} + \frac{8^2}{10}}} = 2.52$$

$$valor\ de\ tabla\ de\ \alpha$$
$$z_\alpha = z_{.05} = 1.645$$

$$Rechazamos\ H_0\ SI:$$
$$z \geq z_{.05} \Rightarrow 2.52 > 1.645$$
$$\therefore Rechazamos\ H_0$$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis de dos medias para datos aperados.

Ya se vio el caso, cuando se hablo de intervalos de confianza para una diferencia de medias, de datos dados de a pares, es decir $(X_{11}, X_{21}); (X_{12}, X_{22}); \dots; (X_{1n}, X_{2n})$. Las v.a. X_1 y X_2 tienen medias μ_1 y μ_2 respectivamente.

Considerando $D_j = X_{1j} - X_{2j}$ con $j = 1,2, \dots, n$.

Entonces

$$E(D_j) = E(X_{1j} - X_{2j}) = E(X_{1j}) - E(X_{2j}) = \mu_1 - \mu_2$$

$$V(D_j) = V(X_{1j} - X_{2j}) = V(X_{1j}) - V(X_{2j}) - 2Cov(X_{1j}, X_{2j}) = \sigma_1^2 + \sigma_2^2 - 2Cov(X_1, X_2)$$

$$Estimamos\ E(D_j) = \mu_1 - \mu_2\ con\ \bar{D} = \frac{1}{n} D_j = \frac{1}{n}\sum_{j=1}^{n}(X_{1j} - X_{2j}) = \bar{X}_1 - \bar{X}_2$$

$$Estimamos\ la\ \ V(D_j)\ con\ S_D^2 = \frac{1}{n-1}\sum_{j=1}^{n}(D_j - \bar{D})^2$$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis de dos medias para datos aperados.

Anotamos $\mu_D = \mu_1 - \mu_2 \; y \; \sigma_D^2 = V(D_j)$ respectivamente.

Asumimos que $D_j \sim N(\mu_D, \sigma_D^2) \; con \; j = 1,2, \dots, n$

Las v.a.en pares diferentes son independientes, no lo son dentro de un mismo par. Para construir una regla de decisión nuevamente, consideremos el estadístico

Estadístico de Prueba:

$$t = \frac{\bar{D} - \Delta_0}{S_D/\sqrt{n}}$$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis de dos medias para datos aperados.

Ejemplo: Se comparan dos microprocesadores en una muestra de 6 códigos de puntos de referencia para determinar si hay una diferencia en la rapidez. Los tiempos en segundos utilizados para cada procesador en cada código están dados en la siguiente tabla:

	Código					
	1	**2**	**3**	**4**	**5**	**6**
Procesador A	27.2	18.1	27.2	19.7	24.5	22.1
Procesador B	24.1	19.3	26.8	20.1	27.6	29.8

¿Puede concluir que las medias de la rapidez de ambos procesadores son diferentes con nivel de significancia 0.05?

5.5 Pruebas de Hipótesis

Prueba de Hipótesis de dos medias para datos aperados.
Estadístico de Prueba:

$$t = \frac{\overline{D} - \Delta_0}{S_D/\sqrt{n}} = \frac{-1.483333 - 0}{3.66246/\sqrt{6}} = -0.99206$$

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

$$Calculamos:$$
$$\overline{D} = -1.483333 \ y \ S_D = 3.66246$$

$$Rechazamos\ H_0\ SI:$$
$$t \leq -t_{\frac{\alpha}{2}, n_1-1} \Rightarrow -0.99206 \leq -2.571$$
$$\therefore NO\ Rechazamos\ H_0$$

$$valor\ de\ tabla\ de\ \alpha$$
$$t_{\alpha/2, n_1-1} = t_{.025,5} = 2.571$$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis sobre la igualdad de dos varianzas.

Supongamos que tenemos interés en dos poblaciones normales independientes, donde las medias y las varianzas de la población son desconocidas. Se desea probar la hipótesis sobre la igualdad de las dos varianzas y normalmente distribuidas:

Estadístico de Prueba:

$$F = \frac{s_1^2}{s_2^2}$$

$$Region\ de\ Rechazo \begin{cases} F \geq F_{\alpha, v_1, v_2} \\ F \leq F_{1-\alpha, v_1, v_2} \\ F \geq F_{\frac{\alpha}{2}, v_1, v_2} \text{ ó } F \leq F_{1-\frac{\alpha}{2}, v_1, v_2} \end{cases}$$

$$Donde\colon\ v_1 = n_1 - 1\ y\ v_2 = n_2 - 1$$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis sobre la igualdad de dos varianzas.

Ejemplo: En una serie de experimentos para determinar la tasa de absorción de ciertos pesticidas en la piel se aplicaron cantidades medidas de dos pesticidas a algunos especímenes de piel. Después de un tiempo se midieron las cantidades absorbidas (en μg). Para el pesticida A la varianza de las cantidades absorbidas en 6 muestras fue de 2.3; mientras que para el B la varianza de las cantidades absorbidas en 10 especímenes fue de 0.6. Suponga que para cada pesticida las cantidades absorbidas constituyen una muestra aleatoria de una población normal. ¿Se puede concluir que la varianza en la cantidad absorbida es mayor para el pesticida A que para el B? Utilizar $\alpha = 0.05$

5.5 Pruebas de Hipótesis

Prueba de Hipótesis sobre la igualdad de dos varianzas.
Estadístico de Prueba:

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 > \sigma_2^2$$

$$F = \frac{s_1^2}{s_2^2} = \frac{2.3}{0.6} = 3.83$$

$$valor\ de\ tabla\ de\ \alpha$$
$$F_{\alpha,v_1,v_2} = F_{.05,5,9} = 3.48$$

$$Rechazamos\ H_0\ SI:$$
$$F \geq F_{\alpha,v_1,v_2} \Rightarrow 3.83 > 3.48$$
$$\therefore\ Rechazamos\ H_0$$

5.5 Pruebas de Hipótesis

P-Valor.

Hasta ahora se dieron los resultados de una prueba de hipótesis estableciendo si la hipótesis nula fue o no rechazada con un valor especificado de α o nivel de significancia.

A menudo este planteamiento resulta inadecuado, ya que no proporciona ninguna idea sobre si el valor calculado del estadístico está apenas en la región de rechazo o bien ubicado dentro de ella. Además, esta forma de establecer los resultados impone a otros usuarios el nivel de significancia predeterminado.

Para evitar estas dificultades, se adopta el enfoque del **p-valor**. El **valor p o p-valor** es la probabilidad de que el estadístico de prueba tome un valor que sea al menos tan extremo como el valor observado del estadístico de prueba cuando la hipótesis nula es verdadera. Es así como el p-valor da mucha información sobre el peso de la evidencia contra H_0, de modo que el investigador pueda llegar a una conclusión para cualquier nivel de significancia especificado.

5.5 Pruebas de Hipótesis

La definición formal del p-valor es la siguiente: El **valor p** es el nivel de significancia más pequeño que conduce al rechazo de la hipótesis nula H_0.

Valor P para una prueba Z	
Valor P = $1 - P(Z < z_{calculada})$	para una prueba de cola superior
Valor P = $P(Z < -z_{calculada})$	para una prueba de cola inferior
Valor P = $2*[1 - P(Z < \|z_{calculada}\|)]$	para una prueba de dos colas
Si, Valor P $\leq \alpha$ Rechazar H_0 al nivel α Valor P $> \alpha$ No rechazar H_0 al nivel α	

5.5 Pruebas de Hipótesis

Ejemplo: En el ejemplo anteúltimo referido al porcentaje deseado de SiO2 en cierto tipo de cemento aluminoso las hipótesis eran: H_0: $\mu = 5.5$ contra H_1: $\mu \neq 5.5$, y el estadístico de prueba es $z = 0.333333$.

$$\text{Calculamos el } Valor\ P = 2 * (1 - \mathrm{P}(\mathrm{Z} < |Z_{calculada}|)$$
$$= 2\big(1 - (\mathrm{Z} < 0.333333)\big) = 2(1 - 0.6293) = 0.719$$

Si,

$$Valor\ P \leq \alpha \ \ Rechazamos\ H_0\ al\ nivel\ \alpha$$
$$Valor\ P > \alpha \ \ No\ Rechazamos\ H_0\ al\ nivel\ \alpha$$

$$Valor\ P = 0.719 > \alpha = .05 \ \ No\ Rechazamos\ H_0\ al\ nivel\ \alpha$$

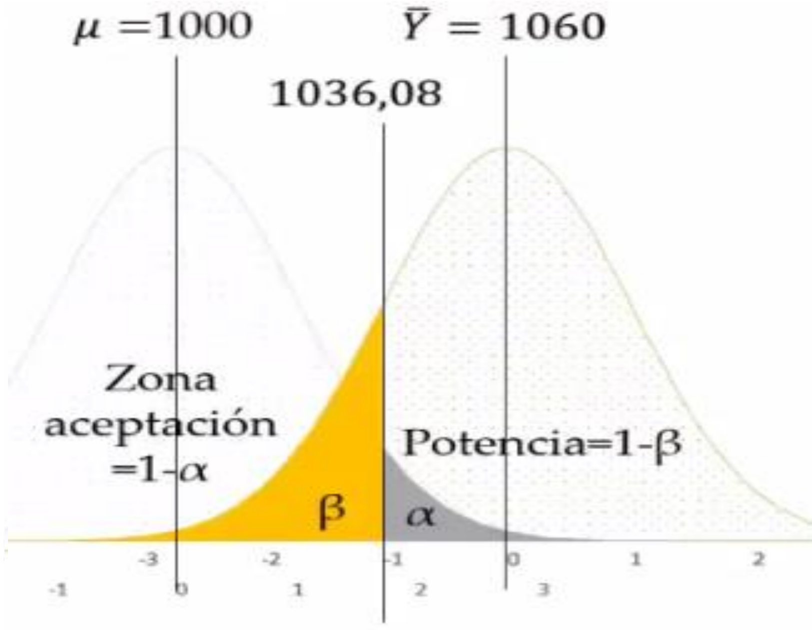
5.6 Función potencia.

Definición: La potencia de una prueba de hipótesis es la probabilidad $1-\beta$ de rechazar una hipótesis nula falsa; se calcula utilizando un nivel de significancia a particular y un valor específico del parámetro de población que representa una alternativa al valor considerado como verdadero en la hipótesis nula. Es decir, la potencia de una prueba de hipótesis es la probabilidad de sustentar una hipótesis alternativa que es verdadera.

5.6 Función potencia.

Ejemplo: Determinar la potencia de la prueba utilizando un N.C. del 95%

$\mu = 1{,}000 \quad n = 100 \quad \bar{Y} = 1{,}060 \quad \sigma = 220$



$$H_0: \mu = 1{,}000$$
$$H_1: \mu = 1{,}060$$

$$P(Z \leq z) = 0.05$$
$$P(Z \leq 1.64) = 0.05$$

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

Sustituimos

$$1.64 = \frac{\bar{Y} - 1{,}000}{220 / \sqrt{100}}$$

Despegamos

$$(22)(1.64) = \bar{Y} - 1{,}000$$
$$\bar{Y} = (22)(1.64) + 1{,}000$$
$$\bar{Y} = 1{,}036.08$$

Calculamos Z

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

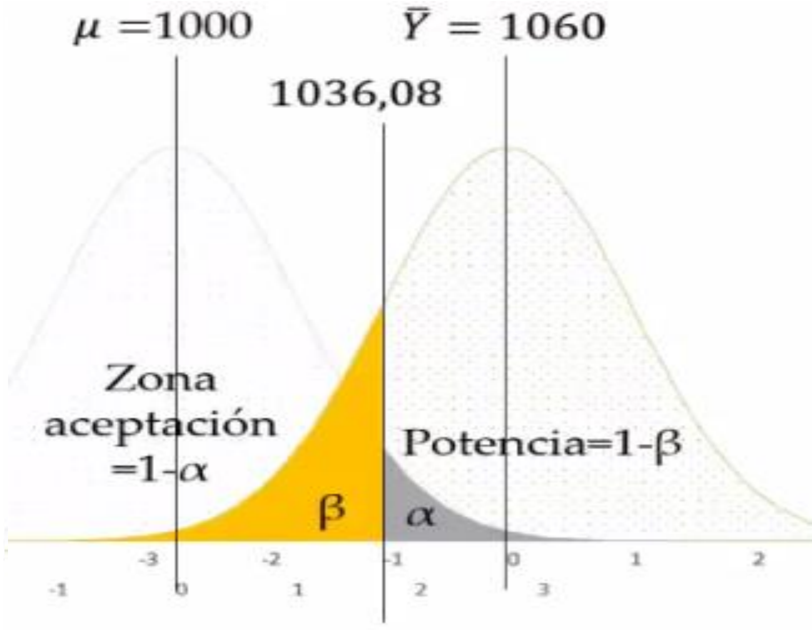
$$Z = \frac{1{,}036.08 - 1{,}000}{220 / \sqrt{100}}$$

$$Z = -1.09$$

5.6 Función potencia.

Ejemplo: Determinar la potencia de la prueba utilizando un N.C. del 95%

$\mu = 1{,}000 \;\; n = 100 \;\; \bar{Y} = 1{,}060 \;\; \sigma = 220$



$$P(Z \leq -1.09) = 0.1379$$
$$\approx 0.14 \; (\beta \; Error \; tipo \; II)$$
$$Potencia = 1 - Error \; tipo \; II =$$
$$1 - 0.14 = 0.86$$

5.7 Lema de Neyman-Pearson.

Si $X_1, X_2, \dots, X_n$ una m.a. de una distrubión $f(x; \theta)$.

Proposición: (Lema de Neyman-Person)
Sea $\alpha \in (0,1)$ y sea θ_0 y θ_1 dos valores distintos del parámetro desconocido θ. La región de rechazo más potente de tamaño α para el contraste de hipótesis

$$H_0: \theta \in \Theta_0 Vs H_0: \theta \in \Theta_1 \Rightarrow H_0: \theta = \theta_0 \; Vs \; H_1: \theta = \theta_1, en\ donde\ \theta_0 \neq \theta_1$$

está dada por $RC_\alpha = \zeta = \left\{ (x_1, \dots, x_n): \dfrac{L(x_1, \dots, x_n; \theta_1)}{L(x_1, \dots, x_n; \theta_0)} > c \right\}$,

en donde $L(x_1, \dots, x_n; \theta_1)$ es la función de verosimilitud de una muestra aleatoria y *"c"* es una constante que hace que hace que esta región de rechazo sea de tamaño α.

5.7 Lema de Neyman-Pearson.

Ejemplo: Dada una función de densidad exponencial con parámetro $\theta > 0$

$$f(x) = \begin{cases} \theta e^{-\theta x} & si \;\; x \geq 0 \\ 0 & si \;\; x < 0 \end{cases}$$

Se desea contrastar la hipótesis $H_0\!: \theta = 2\, Vs\, H_1\!: \theta > 2$ con un nivel de significancia del 7% mediante una muestra de tamaño uno. La mejor región critica según el lema de Neyman-Pearson es $x_1 \leq K$ y

$$\begin{aligned} 0.07 &= \mathbb{P}(\text{aceptar}\, H_1/es\ cierta\ H_0) \\ &= \mathbb{P}(x_1 \leq K/\theta = 2) \\ &= \int_0^K 2e^{-2x_1}\, dx_1 \\ &= 1 - e^{-2K} \end{aligned}$$

5.7 Lema de Neyman-Pearson.

Por lo tanto la función potencia del contraste es

$$
\begin{aligned}
Pot(\theta) &= \mathbb{P}(X \in C/H_1) \\
&= \mathbb{P}(x_1 \leq 0.03625/H_1) \\
&= \int_0^{0.03625} \theta e^{\theta x_1}\, dx_1 \\
&= 1 - e^{-0.03625\theta}
\end{aligned}
$$

Bibliografía.

- ✓ Canavos, G. C. Probabilidad y Estadística: Aplicaciones y Métodos. México. McGraw-Hill. 1987.
- ✓ Casella, G. and Berger, R. L. Statistical Inference. California. Wadsworth. 1990.
- ✓ Ipiña, S. L. (2008). Inferencia Estadística y Análisis de Datos. Madrid: PEARSON Prentice-Hall.
- ✓ Larsen, R. J. and Marx, M. L. An Introduction to Mathematical Statistics and its Applications. USA. Englewood Cliffs-Prentice-Hall. 1986.
- ✓ Lindgren, B. W. Statistical Theory. New York. Macmillan Publishing. 1976.
- ✓ Mood, A. M. et al. Introduction to the Theory of Statistics. New York. McGraw-Hill. 1974.
- ✓ Vazquez, J. (2017). Estimación Puntual. México: Proyecto PAIME UNAM.