Concentración en Analítica de Datos y Herramientas de Inteligencia Artificial II

Daniel Otero Fadul

Departamento de Ciencias
Escuela de Ingeniería y Ciencias

Prueba de Hipótesis

En sesiones anteriores se mostró cómo calcular estimaciones puntuales y por intervalo de los parámetros poblacionales. Ahora continuaremos con el estudio de la inferencia estadística mostrando la forma de usar la **prueba de hipótesis** para determinar si una afirmación acerca del valor de un parámetro poblacional debe o no ser rechazada.

Prueba de Hipótesis

Cuando se hace una prueba de hipótesis se empieza por hacer una suposición tentativa acerca del parámetro poblacional. A esta suposición tentativa se le llama **hipótesis nula** y se denota como H_0. Luego se define otra hipótesis, llamada **hipótesis alternativa**, que dice lo contrario de lo que establece la hipótesis nula. La hipótesis alternativa la denotamos como H_a.

Prueba de Hipótesis

La formulación de la hipótesis nula y alternativa no es siempre obvia, pero, en general, la hipótesis alternativa es algo que queremos probar con métodos estadísticos, lo que implica que la hipótesis nula es la negación de eso que queremos demostrar.

Por ejemplo, cierto modelo de automóvil tiene un rendimiento de 24 kilómetros por galón. Un grupo de investigación elabora un nuevo sistema de inyección de combustible diseñado para dar un mejor rendimiento en kilómetros por galón de gasolina. Ya que el grupo de investigación quiere probar que su diseño mejora el rendimiento, es decir, $rendimiento > 24$, esta sería la hipótesis alternativa. Por lo tanto, la prueba de hipótesis se plantea de la siguiente forma:

$$\begin{aligned} H_0: \text{ rendimiento } &\leq 24 \\ H_a: \text{ rendimiento } &> 24 \end{aligned}$$

Prueba de Hipótesis

Hay dos tipos de prueba de hipótesis: las de **una cola** y las de **dos colas**. En el caso de las de una cola, la hipótesis alternativa es *parámetro*$>$ *valor* o *parámetro*$<$ *valor*:

$$\begin{aligned} H_0 : \text{ parámetro } &\leq \quad valor \\ H_a : \text{ parámetro } &> \quad valor \end{aligned}$$

$$\begin{aligned} H_0 : \text{ parámetro } &\geq \quad valor \\ H_a : \text{ parámetro } &< \quad valor \end{aligned}$$

Prueba de Hipótesis

Cuando la prueba de hipótesis es de dos colas su planteamiento es el siguiente:

$$\begin{aligned} H_0 : \text{ parámetro } &= \mathit{valor} \\ H_a : \text{ parámetro } &\neq \mathit{valor} \end{aligned}$$

Nótese que las relaciones $\leq$, $\geq$ y $=$ siempre hacen parte de la hipótesis nula, no de la alternativa.

Prueba de Hipótesis

Las hipótesis nula y alternativa son afirmaciones opuestas acerca de la población. Una de las dos, ya sea la hipótesis nula o la alternativa es verdadera, pero no ambas. Lo ideal es que la prueba de hipótesis lleve a la aceptación de H_0 cuando H_0 sea verdadera y al rechazo de H_0 cuando H_a sea verdadera. Sin embargo, las conclusiones correctas no siempre son posibles. Como la prueba de hipótesis se basa en una información muestral debe tenerse en cuenta que existe la posibilidad de error.

Prueba de Hipótesis

		Situación en la población	
		H_0 verdadera	H_a verdadera
Conclusión	**Se acepta H_0**	Conclusión correcta	Error tipo II
	Se rechaza H_0	Error tipo I	Conclusión correcta

Figura: El aceptar H_0 cuando no es cierta se le conoce como un **error tipo II**; es decir, se acepta H_0 cuando es falsa. El otro tipo de equivocación que puede ocurrir se da cuando se rechaza H_0 si esta es verdadera, lo cual se conoce como **error tipo I**; es decir, se rechaza H_0 cuando es verdadera. Imagen tomada de [1].

Prueba de Hipótesis

La probabilidad de cometer un error tipo I se le conoce como **nivel de significancia**, el cual se denota como α. En el ejemplo del nuevo sistema de inyección, α es igual a la probabilidad de rechazar H_0 cuando el rendimiento del automóvil con que se está comparando es igual a 24 kilómetros por galón. Otra interpretación del nivel de significancia es la probabilidad máxima que se está dispuesto a aceptar de cometer un error de tipo I.

Prueba de Hipótesis

En la práctica la persona responsable de la prueba de hipótesis especifica el nivel de significancia. Al elegir α se controla la probabilidad de cometer un error tipo I. Si el costo de cometer un error tipo I es elevado, los valores pequeños de α son preferibles. Si el costo de cometer un error tipo I no es demasiado elevado, entonces se usan valores mayores para α. En las aplicaciones de la prueba de hipótesis en que solo se controla el error tipo I se les llama **pruebas de significancia**. Muchas aplicaciones de las pruebas de hipótesis son de este tipo.

Prueba de Hipótesis

Aunque en la mayor parte de las aplicaciones de las pruebas de hipótesis se controla la probabilidad de cometer un error tipo I, no siempre sucede lo mismo con un error tipo II. Por tanto, si se decide aceptar H_0 no es posible depositar toda nuestra confianza en esa decisión. Debido a la incertidumbre de cometer un error tipo II al realizar una prueba de significancia se suele recomendar que se diga "no se rechaza H_0" en lugar de "se acepta H_0". Siempre que no se determine y controle la probabilidad de cometer un error tipo II no se dirá "se acepta H_0". Dicho esto, las conclusiones que se sugieren son **no se rechaza $\mathbf{H_0}$** o **se rechaza $\mathbf{H_0}$**.

Media Poblacional: σ conocida

En toda prueba de hipótesis se trabaja con un **estadístico de prueba** y la distribución asociada a dicho estadístico. En el caso de que el parámetro que nos interesa es la media poblacional, asumiendo que conocemos la desviación estándar de la población σ, el estadístico que se emplea es el siguiente:

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

donde μ_0 es el umbral de la hipótesis nula, $\bar{x}$ la media muestral, y n el tamaño de la muestra.

Media Poblacional: σ conocida

Al realizar una prueba de hipótesis siempre asumimos que la hipótesis nula H_0 es cierta cuando el parámetro de interés es igual a el umbral, el cual sería μ_0 en este caso. Si el estadístico de prueba arroja un valor que es poco probable si H_0 es cierta, entonces se rechaza H_0. La forma en que se rechaza H_0 depende del tipo de prueba que se realice y el método que se use para este fin.

Media Poblacional: σ conocida

La "Federal Trade Commission", FTC, realiza periódicamente estudios estadísticos con objeto de comprobar las afirmaciones de los fabricantes acerca de sus productos. Por ejemplo, en la etiqueta de una lata grande de "Hilltop Coffee" dice que la lata contiene 3 libras de café. La FTC sabe que el proceso de producción de Hilltop no permite llenar las latas con 3 libras exactas de café por lata, incluso si la media poblacional del peso de llenado de todas las latas es de 3 libras por lata. Sin embargo, mientras la media poblacional del peso de llenado sea por lo menos 3 libras por lata los derechos del consumidor estarán protegidos. Por tanto, la FTC interpreta que la información de la etiqueta de una lata grande de café Hilltop tiene una media poblacional del peso de llenado de por lo menos 3 libras por lata.

Media Poblacional: σ conocida

Dicho lo anterior, la FTC plantea la siguiente prueba de hipótesis de una cola para verificar que Hilltop está cumpliendo con lo que dice en las etiquetas de su producto:

$$\begin{aligned} H_0 : \mu &\geq 3 \\ H_a : \mu &< 3 \end{aligned}$$

Media Poblacional: σ conocida

La FTC toma una muestra de 36 elementos y obtiene una media muestral igual a $\bar{x} = 2.92$ libras. Supongamos que $\sigma = 0.18$. Nótese que en este caso $\mu_0 = 3$. Con estos datos, nuestro estadístico de prueba es igual a

$$\begin{aligned} Z &= \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \\ &= \frac{2.92 - 3}{0.18/\sqrt{36}} \\ &= -2.67. \end{aligned}$$

Si el nivel de significancia de la prueba es $\alpha = 0.01$, ¿cómo podemos saber si rechazamos o no H_0? Hay dos formas de hacer esto.

Método del Valor Crítico

En el **método del valor crítico** primero se determina un valor para el estadístico de prueba llamado **valor crítico**. En una prueba de la cola inferior, el valor crítico sirve como punto de referencia para determinar si el valor del estadístico de prueba es lo suficientemente pequeño para rechazar la hipótesis nula. El valor crítico es el valor del estadístico de prueba que corresponde a un área α (nivel de significancia) en la cola inferior de la distribución muestral del estadístico de prueba. En otras palabras, el valor crítico es el mayor valor del estadístico de prueba que hará que se rechace la hipótesis nula.

Método del Valor Crítico

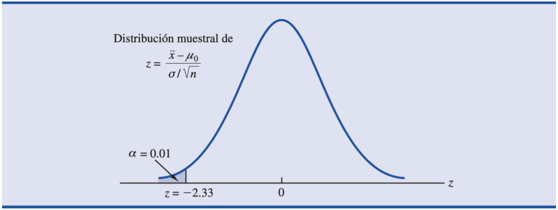

Figura: Con el método del valor crítico H_0 se rechaza si el estadístico de prueba es menor o igual al valor crítico, el cual se denota como $-z_\alpha$ en este ejemplo. Para un nivel de significancia $\alpha = 0.01$, el valor crítico es igual a $-z_\alpha = -2.33$. Ya que el estadístico z tiene un valor de -2.67, se tiene que $z \leq -z_\alpha$, así que se rechaza H_0. Imagen tomada de [1].

Método del Valor p

En el **método del valor-p** se usa el valor del estadístico de prueba Z para calcular una probabilidad llamada **valor-p**. Este valor es una probabilidad que aporta una medida de una evidencia suministrada por la muestra contra la hipótesis nula. Valores-p pequeños indican una evidencia mayor contra la hipótesis nula.

Para la prueba de hipótesis de este ejemplo, el valor-p es igual a

$$p = P\left(Z \leq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right).$$

La hipótesis nula se rechaza si $p \leq \alpha$.

Método del Valor p

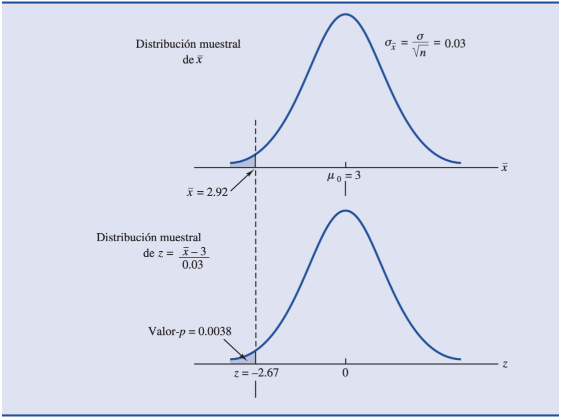

Figura: Con el método del valor-p H_0 es rechazada si esta cantidad es menor o igual al nivel de significancia α. En este caso tenemos que $p = 0.0038$. Ya que $p \leq \alpha$ se rechaza H_0. Imagen tomada de [1].

Prueba de dos Colas

En las pruebas de hipótesis, la forma general de una prueba de dos colas para la media poblacional es la siguiente:

$$\begin{aligned} H_0 : \mu &= \mu_0 \\ H_a : \mu &\neq \mu_0 \end{aligned}$$

A continuación se ilustrará cómo se realiza este tipo de prueba para la media poblacional cuando la desviación estándar σ de la población es conocida.

Prueba de dos Colas

La "U.S. Golf Association", USGA, establece reglas que deben satisfacer los fabricantes de equipos de golf si quieren que sus productos se acepten en los eventos de USGA. "MaxFlight" emplea procesos de fabricación de alta tecnología para producir pelotas de golf que tienen una distancia media de recorrido de 295 yardas. Sin embargo, algunas veces el proceso se desajusta y se producen pelotas de golf que tienen una distancia media de recorrido diferente a 295 yardas. Cuando la distancia media es menor que 295 yardas, a la empresa le preocupa perder clientes porque las pelotas de golf no proporcionen la distancia anunciada. Cuando la distancia es mayor que 295 yardas, las pelotas de MaxFlight pueden ser rechazadas por la USGA por exceder los estándares respecto de distancia de vuelo y carrera.

Prueba de dos Colas

Dicho lo anterior, la prueba de hipótesis planteada por MaxFlight es la siguiente:

$$\begin{aligned} H_0 : \mu &= 295 \\ H_a : \mu &\neq 295 \end{aligned}$$

Sean $\sigma = 12$ y $\alpha = 0.05$. Si se toma una muestra de $n = 50$ pelotas con media muestral igual a $\bar{x} = 297.6$, ¿será $\bar{x}$ lo suficientemente grande para rechazar H_0?

Prueba de dos Colas

En este tipo de prueba se consideran las dos "colas" de la distribución. Si utilizamos el método del valor crítico, para un nivel de significancia igual a α, debemos encontrar un valor crítico $z_{\alpha/2}$ tal que

$$P(Z \geq z_{\alpha/2}) = \frac{\alpha}{2},$$

donde Z es el estadístico de prueba:

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

De este forma se obtiene que

$$P(Z \leq -z_{\alpha/2}) + P(Z \geq z_{\alpha/2}) = \alpha.$$

Prueba de dos Colas

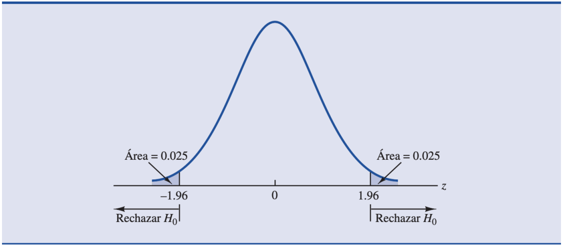

Figura: En esta gráfica se pueden observar las regiones de rechazo de H_0 cuando la distribución del estadístico de prueba es una normal y el nivel de significancia es igual a $\alpha = 0.05$. Imagen tomada de [1].

Prueba de dos Colas

Ya que $\alpha = 0.05$, tenemos que $z_{\alpha/2} = 1.96$, por lo tanto, H_0 se rechaza si el estadístico de prueba es mayor o igual que $z_{\alpha/2} = 1.96$ o menor o igual que $-z_{\alpha/2} = -1.96$. En este caso, se tiene que el estadístico es igual a

$$\begin{aligned} Z &= \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \\ &= \frac{297.6 - 295}{12/\sqrt{50}} \\ &= 1.53. \end{aligned}$$

Así que no se rechaza H_0. En otras palabras, no hay suficiente evidencia para pensar que las pelotas de golf no cubren una distancia distinta a 295 yardas.

Prueba de dos Colas

En una prueba de dos colas, el valor-p es la probabilidad de obtener un valor para el estadístico de prueba tan improbable o más improbable que el obtenido con la muestra:

$$p = P\left(Z \leq -\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right) + P\left(Z \geq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right).$$

La hipótesis nula se rechaza si $p \leq \alpha$.

Prueba de dos Colas

Para el ejemplo de las pelotas de golf tenemos que el valor-p es igual a

$$\begin{aligned} p &= P\left(Z \leq -\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right) + P\left(Z \geq \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right) \\ &= P(Z \leq -1.53) + P(Z \geq 1.53) \\ &= 0.063 + 0.063 \\ &= 0.126. \end{aligned}$$

Ya que $p > \alpha$, no se rechaza H_0.

Prueba de dos Colas

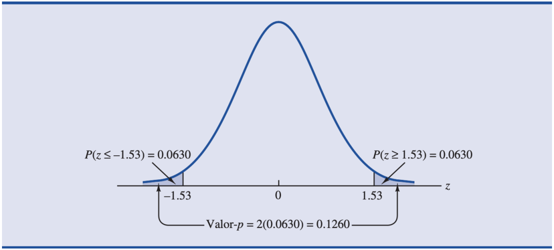

Figura: En la figura se ilustra el cálculo del valor-p para el ejemplo de las pelotas de golf. Se puede ver que p es mayor que el nivel de significancia de la prueba, así que se rechaza H_0. Imagen tomada de [1].

Prueba de dos Colas

Vale la pena mencionar que existe una relación entre la prueba de hipótesis de dos colas y los intervalos de confianza: si el nivel de significancia de una prueba es α, entonces la hipótesis H_0 se rechaza si el valor μ_0 no está contenido en el intervalo de confianza de μ con un nivel de confianza del $(1 - \alpha) \times 100\%$.

Prueba de dos Colas

Consideremos nuevamente los datos del caso de las pelotas de golf fabricadas por la empresa "MaxFlight": $\sigma = 12$, $n = 50$, $\bar{x} = 297.6$ y $\alpha = 0.05$. Si construimos un intervalo de confianza para la media poblacional μ de la distancia cubierta por las pelotas tendríamos lo siguiente:

$$\begin{aligned}\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] &= \left[297.6 - (1.96)\frac{12}{\sqrt{50}}, 297.6 + (1.96)\frac{12}{\sqrt{50}}\right] \\ &= [294.27, 300.92].\end{aligned}$$

En este caso $\mu_0 = 295 \in [294.27, 300.92]$. Ya que μ_0 está contenido en el intervalo de confianza, no se rechaza H_0.

Resumen de Pruebas de Hipótesis

	Prueba de la cola inferior	Prueba de la cola superior	Prueba de dos colas
Hipótesis	$H_0: \mu \geq \mu_0$ $H_a: \mu < \mu_0$	$H_0: \mu \leq \mu_0$ $H_a: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$
Estadístico de prueba	$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
Regla de rechazo: método del valor-*p*	Rechazar H_0 si valor-$p \leq \alpha$	Rechazar H_0 si valor-$p \leq \alpha$	Rechazar H_0 si valor-$p \leq \alpha$
Regla de rechazo: método del valor crítico	Rechazar H_0 si $z \leq -z_\alpha$	Rechazar H_0 si $z \geq z_\alpha$	Rechazar H_0 si $z \leq -z_{\alpha/2}$ o si $z \geq z_{\alpha/2}$

Figura: En esta tabla se compilan los tres tipos de pruebas de hipótesis que podemos realizar para la media poblacional cuando la desviación estándar de la población se conoce. Nótese que independientemente del tipo de prueba que se haga, el criterio para rechazar H_0 siempre es el mismo cuando se utiliza el método del valor-p. Imagen tomada de [1].

Media Poblacional: σ desconocida

Los procedimientos que se realizan para hace una prueba de hipótesis cuando la desviación estándar poblacional σ no se conoce son los mismos que vimos para el caso en que σ es conocida, sin embargo debemos estimar este parámetro σ con algún estimador, por ejemplo s, y utilizar la distribución t-Student ya que esta es la distribución del estadístico de prueba en este caso:

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

donde μ_0 es el umbral de la hipótesis nula, $\bar{x}$ la media muestral, y n el tamaño de la muestra.

Media Poblacional: σ desconocida

	Prueba de la cola inferior	Prueba de la cola superior	Prueba de dos colas
Hipótesis	$H_0: \mu \geq \mu_0$ $H_a: \mu < \mu_0$	$H_0: \mu \leq \mu_0$ $H_a: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$
Estadístico de prueba	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
Regla de rechazo: *método del valor-p*	Rechazar H_0 si valor-$p \leq \alpha$	Rechazar H_0 si valor-$p \leq \alpha$	Rechazar H_0 si valor-$p \leq \alpha$
Regla de rechazo: método del valor crítico	Rechazar H_0 si $t \leq -t_\alpha$	Rechazar H_0 si $t \geq t_\alpha$	Rechazar H_0 si $t \leq -t_{\alpha/2}$ o si $t \geq t_{\alpha/2}$

Figura: En esta tabla se compilan los tres tipos de pruebas de hipótesis que podemos realizar para la media poblacional cuando la desviación estándar de la población no se conoce. Imagen tomada de [1].

Media Poblacional: σ desconocida

Consideremos la siguiente prueba de hipótesis:

$$\begin{aligned} H_0 : \mu &\leq 12 \\ H_a : \mu &> 12 \end{aligned}$$

Si el tamaño de la muestra es $n = 25$, la media muestral es $\bar{x} = 14$ y la desviación estándar muestral es igual a $s = 4.32$, para un nivel de significancia $\alpha = 0.05$, ¿es posible rechazar H_0? ¿A qué sería igual el valor crítico? ¿A qué es igual el valor-p?

Prueba de Hipótesis para $\mu_1 - \mu_2$

Cuando queremos establecer si las medias de dos poblaciones son significativamente diferentes podemos plantear una prueba de hipótesis de la diferencia de las medias μ_1 y μ_2. Asumiendo que las desviaciones estándar de las poblaciones son distintas y conocidas, el estadístico de prueba se define como

$$Z = \frac{\bar{x}_1 - \bar{x}_2 - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}},$$

donde $\bar{x}_1$ y $\bar{x}_2$ son las medias muestrales, σ_1 y σ_2 las desviaciones estándar poblacionales, n_1 y n_2 el tamaño de las muestras, y d_0 la diferencia hipotética de las medias. Un valor típico para d_0 es cero.

Para este tipo de prueba la distribución del estadístico es la normal estándar.

Prueba de Hipótesis para $\mu_1 - \mu_2$

Cuando las desviaciones estándar de las poblaciones no se conocen, el estadístico de prueba sigue una distribución t-Student y las desviaciones σ_1 y σ_2 son reemplazadas por sus estimaciones s_1 y s_2:

$$T = \frac{\bar{x}_1 - \bar{x}_2 - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

El número de grados de libertad k de la distribución es igual a la parte entera de la siguiente fórmula:

$$\hat{k} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}.$$

En otras palabras, $k = \lfloor \hat{k} \rfloor$.

Prueba de Hipótesis para $\mu_1 - \mu_2$

Supongamos que existe un nuevo "software" que ayuda a los analistas de sistemas a reducir el tiempo requerido para diseñar, elaborar y poner en marcha un sistema de información. Para evaluar las ventajas del nuevo "software" se toma una muestra de 24 analistas de sistemas. A cada analista se le da información sobre un sistema de información hipotético. A 12 de ellos se les pide que elaboren el sistema de información usando la tecnología existente y a los otros 12 analistas se les capacita para usar el nuevo software y se les pide que lo empleen para elaborar el sistema de información.

Prueba de Hipótesis para $\mu_1 - \mu_2$

Sea μ_1 la media poblacional del tiempo promedio que les toma a los analistas completar el sistema de información con la tecnología ya existente. Sea μ_2 la media poblacional del tiempo promedio que les toma a los analistas completar el sistema de información con el nuevo "software".

Si el nuevo "software" reduce el tiempo que les lleva a los analistas llevar a cabo su trabajo, deberíamos tener $\mu_1 > \mu_2$, o en otras palabras, $\mu_1 - \mu_2 > 0$. Por lo tanto, planteamos la siguiente prueba de hipótesis:

$$\begin{aligned} H_0 : \mu_1 - \mu_2 &\leq 0 \\ H_a : \mu_1 - \mu_2 &> 0 \end{aligned}$$

Nótese que $d_0 = 0$.

Prueba de Hipótesis para $\mu_1 - \mu_2$

	Tecnología existente	Software nuevo
	300	274
	280	220
	344	308
	385	336
	372	198
	360	300
	288	315
	321	258
	376	318
	290	310
	301	332
	283	263
Resumen estadístico		
Tamaño de la muestra	$n_1 = 12$	$n_2 = 12$
Media muestral	$\bar{x}_1 = 325$ horas	$\bar{x}_2 = 286$ horas
Desviación estándar muestral	$s_1 = 40$	$s_2 = 44$

Figura: Los datos del ejemplo de los analistas que desarrollan un sistema de información con tecnología ya existente versus usando un nuevo "software". Imagen tomada de [1].

Prueba de Hipótesis para $\mu_1 - \mu_2$

Teniendo en cuenta los datos de la diapositiva anterior, si el nivel de significancia es igual a $\alpha = 0.05$, ¿a qué es igual el valor crítico? ¿Se rechaza H_0?

Prueba de Independencia

Otra aplicación importante de la distribución chi-cuadrada es el empleo de datos muestrales para probar la independencia de dos variables. Para ilustrar la prueba de independencia se considerará la prueba de independencia realizada por la Alber's Brewery de Tucson, Arizona. Alber's produce y distribuye tres tipos de cerveza: ligera, clara y oscura. Al analizar los segmentos de mercado de las tres cervezas, el grupo de investigación de mercado de la empresa se pregunta si las preferencias de los consumidores por estos tipos de cerveza difieren entre hombres y mujeres. En caso de que las preferencias fueran independientes del sexo biológico del consumidor, iniciarían una campaña publicitaria para todas las cervezas de Alber's, pero si las preferencias por los distintos tipos de cerveza dependen del sexo biológico del consumidor, la empresa ajustaría sus promociones a los mercados.

Prueba de Independencia

La prueba de hipótesis que se plantearía en este caso para saber si las variables son independientes sería la siguiente:

H_0 : La preferencia por un tipo de cerveza y el sexo biológico son independientes

H_a : La preferencia por un tipo de cerveza y el sexo biológico no son independientes

Prueba de Independencia

	Ligera	Clara	Oscura	Total
Hombre	20	40	20	80
Mujer	30	30	10	70
Total	50	70	30	150

Figura: Para probar si dos variables son independientes, se toma una muestra y se usa una tabulación cruzada para resumir los datos de las dos variables simultáneamente. En esta tabla se pueden ver las **frecuencias observadas** para cada posible categoría. Por ejemplo, el número de mujeres que prefieren la cerveza "clara" son 30. Imagen tomada de [1].

Prueba de Independencia

Como en toda prueba de hipótesis, asumimos primero que la hipótesis nula es cierta. Lo que procede a continuación es calcular las **frecuencias esperadas** asumiendo que H_0 es verdadera.

Prueba de Independencia

Definamos dos eventos: $S = \{\text{Hombre}\}$ y $C = \{\text{Oscura}\}$. Si estos eventos son independientes deberíamos tener que

$$\begin{aligned} P(S \cap C) &= P(S)P(C) \\ &= \left(\frac{80}{150}\right)\left(\frac{30}{150}\right) \\ &= \frac{8}{75}. \end{aligned}$$

Así que la frecuencia esperada para la intersección de estos eventos será

$$\begin{aligned} \text{frecuencia esperada} &= P(S)P(C)(\text{total de consumidores}) \\ &= \left(\frac{8}{75}\right)(150) \\ &= 16. \end{aligned}$$

Prueba de Independencia

Las frecuencias esperadas también se organizan en una tabla. La fórmula de la frecuencia esperada de la posición ij, la cual se denota como e_{ij}, está dada por

$$e_{ij} = \frac{(\text{total del renglón } i)(\text{total de la columna } j)}{\text{tamaño de la muestra}}.$$

Por ejemplo, la frecuencia calculada en la diapositiva anterior correspondería a la frecuencia esperada e_{13}:

$$\begin{aligned} e_{13} &= \frac{(\text{total del renglón 1})(\text{total de la columna 3})}{\text{tamaño de la muestra}} \\ &= \frac{(80)(30)}{150} \\ &= 16. \end{aligned}$$

Prueba de Independencia

	Ligera	Clara	Oscura	Total
Hombre	26.67	37.33	16.00	80
Mujer	23.33	32.67	14.00	70
Total	50.00	70.00	30.00	150

Figura: En esta tabla se pueden ver las **frecuencias esperadas** para cada posible escenario. Las frecuencias esperadas se calculan con la fórmula que se encuentra en la diapositiva anterior. Imagen tomada de [1].

Prueba de Independencia

Si las frecuencias se organizan en una tabla de n filas y m columnas, el estadístico de prueba se define como sigue:

$$\chi^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{(f_{ij} - e_{ij})^2}{e_{ij}},$$

donde f_{ij} y e_{ij} son la frecuencia observada y esperada de la fila i y la columna j, respectivamente. En esta prueba el estadístico sigue una distribución χ-cuadrado de $(n-1)(m-1)$ grados de libertad.

Prueba de Independencia

Género	Cerveza preferida	Frecuencia observada (f_{ij})	Frecuencia esperada (e_{ij})	Diferencia ($f_{ij} - e_{ij}$)	Cuadrado de la diferencia ($f_{ij} - e_{ij})^2$	Cuadrado de la diferencia dividido entre frecuencia esperada $(f_{ij} - e_{ij})^2/e_{ij}$
Hombre	Ligera	20	26.67	−6.67	44.44	1.67
Hombre	Clara	40	37.33	2.67	7.11	0.19
Hombre	Oscura	20	16.00	4.00	16.00	1.00
Mujer	Ligera	30	23.33	6.67	44.44	1.90
Mujer	Clara	30	32.67	−2.67	7.11	0.22
Mujer	Oscura	10	14.00	−4.00	16.00	1.14
	Total	150				$\chi^2 = 6.12$

Figura: Cálculo del estadístico de la prueba de independencia. En este caso se obtiene que $\chi^2 = 6.12$. Para este ejemplo, el nivel de significancia de la prueba es $\alpha = 0.05$. El número de grados de libertad de la distribución es $(n-1)(m-1) = (2-1)(3-1) = 2$. Bajo estas circunstancias se obtiene que el valor-p es igual a $p = 0.0469$. Ya que $p \leq \alpha$, la hipótesis de que la preferencia del tipo de cerveza y el sexo biológico son variables independientes se rechaza. Imagen tomada de [1].

Análisis de Varianza

El **análisis de varianza**, conocido como **ANOVA** por sus siglas en inglés, es una técnica estadística utilizada para analizar la variabilidad en un conjunto de datos y determinar si existen diferencias significativas entre las medias de tres o más poblaciones. Esta técnica es una herramienta poderosa para comparar múltiples grupos y determinar si las diferencias observadas son el resultado de factores reales o si podrían deberse al azar.

Análisis de Varianza

Hay varios tipos de ANOVA, pero estos se pueden clasificar en dos grandes grupos: los de un factor y los de dos o más factores.

- **ANOVA de un factor:** Se utiliza para comparar las medias de tres o más poblaciones cuando solo se tiene en cuenta un factor o variable independiente. Por ejemplo, se podría utilizar para determinar si hay diferencias significativas en las calificaciones de estudiantes de diferentes escuelas.
- **ANOVA de dos o más factores:** Esta prueba, también conocida como **experimento factorial**, nos permite considerar dos o más factores para analizar cómo influyen conjuntamente en la variable dependiente. Por ejemplo, se podría utilizar para estudiar cómo tanto el género como la edad influyen en el rendimiento académico.

En nuestro caso hablaremos solamente del primer tipo de ANOVA.

Análisis de Varianza

El ANOVA de un solo factor se utiliza para probar la igualdad de k medias poblacionales en un diseño completamente aleatorio. La forma general de esta prueba de hipótesis es la siguiente:

$$
\begin{aligned}
H_0 &: \quad \mu_1 = \mu_2 = \cdots = \mu_k \\
H_a &: \quad \text{No todas las medias poblacionales son iguales}
\end{aligned}
$$

Análisis de Varianza

Un ejemplo de un estudio estadístico experimental es el siguiente: la empresa "Chemitech, Inc." elaboró un sistema de filtración para los suministros de aguas municipales. Los componentes del sistema de filtración se comprarían a varios proveedores y Chemitech armaría el sistema de filtración en su fábrica en Columbia, Carolina del Sur. El grupo de ingenieros industriales es el encargado de determinar el mejor método para armar el sistema de filtración. Después de considerar varios métodos, quedaron solo tres alternativas: el método A, el método B y el método C. La diferencia entre estos métodos era el orden en los pasos para armar el sistema. Chemitech quiere saber con cuál método se pueden producir más sistemas de filtración en una semana.

Análisis de Varianza

En el experimento de Chemitech, el método para armar el sistema es la **variable independiente** o **factor**. Como a este factor le corresponden tres métodos para armar el sistema, se dice que en este experimento hay tres **tratamientos**; cada tratamiento corresponde a uno de los tres métodos para armar el sistema. El problema de Chemitech es un ejemplo de un experimento de un solo factor; interviene solo un factor cualitativo (el método para armar el sistema). En experimentos más complejos caben múltiples factores; los factores pueden ser cualitativos o cuantitativos.

Análisis de Varianza

Los tres tratamientos o métodos para armar el sistema constituyen las **tres poblaciones de interés** del experimento de Chemitech. Una población está formada por todos los trabajadores que emplean el método A, otra población es la de todos los trabajadores que emplean el método B, y otra población es la de todos los trabajadores que emplean el método C. Nótese que en cada población la **variable dependiente**, o **variable de respuesta**, es el número de sistemas de filtración que se arman por semana. El objetivo estadístico del experimento es determinar si el número medio producido por semana es el mismo en las tres poblaciones (con los tres métodos).

Análisis de Varianza

Supongamos que tenemos una muestra aleatoria de tres trabajadores de la empresa Chemitech. En el lenguaje del diseño de experimentos, estos tres trabajadores son las **unidades experimentales**. El diseño de experimentos que se usará para el problema de Chemitech se le llama **diseño completamente aleatorio**. En este tipo de diseño se requiere que cada uno de los tratamientos o métodos para armar el sistema se asigne de manera aleatoria a cada una de las unidades experimentales o trabajadores.

Análisis de Varianza

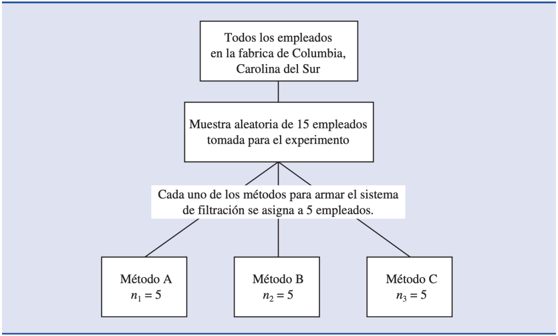

Figura: Diseño experimental completamente aleatorio para evaluar el método de armado del sistema de filtración de la empresa Chemitech. Imagen tomada de [1].

Análisis de Varianza

	Método		
	A	**B**	**C**
	58	58	48
	64	69	57
	55	71	59
	66	64	47
	67	68	49
Media muestral	62	66	52
Varianza muestral	27.5	26.5	31.0
Desviación estándar muestral	5.244	5.148	5.568

Figura: Datos del número de unidades producidas por los trabajadores utilizando los distintos métodos. Imagen tomada de [1].

Análisis de Varianza

Para que el ANOVA sea válido se deben cumplir los siguientes supuestos:

- Para cada población, la variable de respuesta tiene una distribución normal. En el caso de Chemitech asumimos que el número de unidades producida por semana (variable de respuesta) con cada uno de los métodos debe tener una distribución normal.
- La varianza de la variable de respuesta σ^2 es la misma en todas las poblaciones. En el experimento de Chemitech, para los tres métodos, la varianza en el número de unidades producidas por semana debe ser la misma.
- Las observaciones deben ser independientes. Para el experimento de Chemitech la cantidad de unidades producida por semana por un empleado debe ser independiente del número de unidades producidas por semana por cualquier otro empleado.

Dado lo anterior, la prueba de hipótesis que se plantea es la siguiente:

$$
\begin{aligned}
H_0 &: \quad \mu_1 = \mu_2 = \mu_3 \\
H_a &: \quad \text{No todas las medias poblacionales son iguales}
\end{aligned}
$$

Las medias poblacionales de los métodos A, B y C son μ_1, μ_2 y μ_3, respectivamente.

Análisis de Varianza

Si las medias de las tres poblaciones son iguales, se esperaría que las tres medias muestrales fueran muy parecidas. En efecto, entre más parecidas sean las medias muestrales, mayor será la evidencia para concluir que las medias poblacionales son iguales. En otras palabras, si la variabilidad entre las medias muestrales es "pequeña", esto favorece a H_0; si la variabilidad entre las medias muestrales es "grande", esto favorece a H_a.

Análisis de Varianza

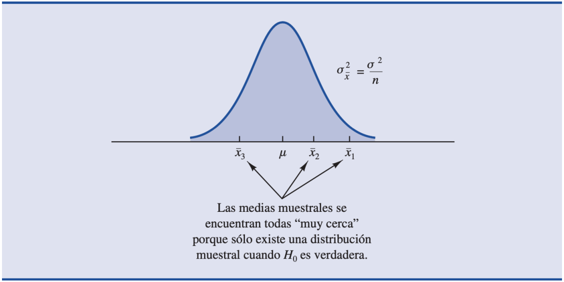

Figura: Visualización de lo que ocurre si la hipótesis de que todas las medias poblacionales son iguales es cierta. Nótese que la varianza de la distribución muestral es la varianza del estimador de la media muestral $\bar{x}$. Imagen tomada de [1].

Análisis de Varianza

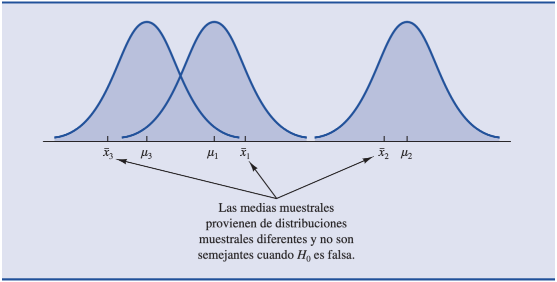

Figura: Visualización de lo que ocurre si la hipótesis de que todas las medias poblacionales son iguales es falsa. Nótese que la variabilidad de las medias muestrales aumenta. Imagen tomada de [1].

Análisis de Varianza

Dado lo anterior, para verificar si las medias poblacionales son iguales, en el análisis de varianza se estima la varianza poblacional σ^2 midiendo la variabilidad **dentro** de los tratamientos y **entre** los tratamientos. Si H_0 es cierta, estas dos estimaciones deberían coincidir, pero si esto no es así, la estimación de la varianza entre los tratamientos debería sobreestimar σ^2.

Análisis de Varianza

El **estimador de la varianza poblacional entre tratamientos**, que denotamos como **CMTR**, se define como

$$\text{CMTR} = \frac{1}{k-1}\sum_{j=1}^{k} n_j(\bar{x}_j - \bar{x})^2,$$

donde k es el número de tratamientos, n_j es el número de unidades experimentales del tratamiento j, $\bar{x}_j$ es la media muestral del tratamiento j, y $\bar{x}$ es la media muestral de la media poblacional μ asumiendo que H_0 es verdadera. Este es un estimador insesgado de la varianza poblacional σ^2.

Análisis de Varianza

El **estimador de la varianza poblacional dentro de los tratamientos**, que denotamos como **CME**, se define como

$$\text{CME} = \frac{1}{n-k}\sum_{j=1}^{k}(n_j - 1)s_j^2,$$

donde n_j y s_j^2 son el número de unidades experimentales y la varianza muestral del tratamiento j, respectivamente. CME también es un un estimador insesgado de la varianza poblacional σ^2.

El estadístico de prueba F de la prueba de la igualdad de k medias poblacionales se define como

$$F = \frac{\text{CMTR}}{\text{CME}},$$

el cual sifue una distribución F con $k-1$ grados de libertad en el numerador y $n-k$ grados de libertad en el denominador.

Como es de esperarse, para un nivel de significancia igual a α, H_0 se rechaza si el estadístico F es mayor que el valor crítico f_α, o si el valor-p es menor que α.

Análisis de Varianza

Con base en los datos del experimento de Chemitech, ¿serán las medias poblacionales del número de unidades producidas por semana co los tres métodos de armado de los sistemas de filtración iguales?

BIBLIOGRAFÍA

1 Anderson, D. R., Sweeney, D. J., Williams, T. A., *"Estadística para Administración y Economía"*, Décima Edición, CENGAGE Learning, 2008.

2 Walpole R. E., Myers R. H., Myers S. L., Keying Y., *"Probability and Statistics for Engineers and Scientists"*, 9th Edition, Prentice Hall, 2012.