

Consideraciones estadísticas en los ensayos de investigación preclínica

Sarah Eliuth Ochoa Hugo¹, Genaro David Santillán Dávila² y Rodolfo Hernández Gutiérrez^{1*}

¹Laboratorio de Investigación Traslacional de Terapias contra el Cáncer, Unidad de Biotecnología Médica y Farmacéutica, Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco. Guadalajara Jalisco.

²Instituto Tecnológico y de Estudios Superiores de Monterrey Campus Guadalajara. Guadalajara Jalisco.

*Autor de correspondencia: rhgutierrez@ciatej.mx

Resumen

Palabras clave:

análisis estadístico, análisis de varianza, evaluación preclínica de medicamentos, Lean Six Sigma.

Durante la última década, el deficiente análisis estadístico de datos obtenidos a partir de ensayos de investigación preclínica ha representado una de las principales causas de la falta de reproducibilidad de resultados en otros laboratorios, tanto de la academia como de la industria durante el proceso de desarrollo de nuevos fármacos. En este trabajo se hace una recapitulación de los conceptos estadísticos relevantes al realizar el análisis de ensayos de investigación preclínica, como lo son: la desviación y error estándar, distribución de probabilidad, distribución normal, t de student, nivel e intervalos de confianza, percentiles y valor de p, prueba de hipótesis, ANOVA y, prueba de Tukey. Así mismo, se describe el alcance de la filosofía Lean Six Sigma y su aplicación en ensayos preclínicos como respuesta a la necesidad de confiabilidad, repetibilidad y reproducibilidad en los resultados.

Introducción

Los ensayos preclínicos desempeñan un papel muy importante en el proceso de desarrollo de nuevos fármacos. Sin embargo, una gran proporción de los resultados de ensayos preclínicos, no se asemejan a los resultados obtenidos en los estudios clínicos. Los artículos científicos publicados suelen presentar baja reproducibilidad, lo cual se atribuye a sesgos que pueden ocurrir en diferentes etapas de la experimentación, así como en el análisis estadístico de los datos obtenidos para un correcto reporte de resultados (Prinz et al., 2011).

Los ensayos de investigación preclínica tienen por objetivo determinar si un posible fármaco tiene posibilidades de conver-

Enfoques Transdisciplinarios:
Ciencia y Sociedad, 3(1), 29-47.
ISSN. 3061-709X. <https://doi.org/10.5281/zenodo.14706445>

Recibido: 09 octubre 2024
Revisado: 01 de noviembre 2024
Aceptado: 1 de diciembre 2024
Publicado: 21 de enero 2025



Este artículo es un artículo de acceso abierto distribuido bajo los términos y condiciones de la licencia CC BY-NC-SA 4.0. Para ver una copia de esta licencia visite <https://creativecommons.org/licenses/by-nc-sa/4.0/>



tirse en un medicamento de uso humano frente a una enfermedad en particular. Durante el proceso de desarrollo de nuevos fármacos, la investigación preclínica llevada a cabo tanto en laboratorios académicos como a nivel industrial hace uso de la estadística para observar patrones, tendencias, variaciones entre mediciones y comparar entre lo observado en la academia y en la industria, en donde se presentan sesgos importantes en los resultados y limitaciones relacionadas con la repetibilidad y reproducibilidad de los estudios (Yenduri & Silver, 2023).

En este trabajo se hace una recapitulación de los conceptos estadísticos relevantes al realizar el análisis de ensayos de investigación preclínica y del alcance de la filosofía Lean Six Sigma como respuesta a la necesidad de confiabilidad, precisión, repetibilidad y reproducibilidad en los resultados. Para esto se realizó una búsqueda bibliográfica en bases de datos como PubMed, EBSCO, Medigraphic, ClinicalKey, Elsevier y GoogleScholar. Se incluyeron guías de The Council for Six Sigma, American Society for Quality y artículos publicados en idioma inglés y español. Los términos de búsqueda utilizados fueron: análisis estadístico, Lean Six Sigma, análisis de varianza, evaluación preclínica de medicamentos.

Desarrollo

Ensayos preclínicos

Proceso de desarrollo de nuevos fármacos

El desarrollo de nuevos fármacos contra alguna enfermedad en específico es un proceso costoso en términos de tiempo, trabajo y recursos económicos (ver figura 1). Este proceso comienza con los ensayos preclínicos en los laboratorios de experimentación, siendo los más comúnmente empleados aquellos tanto *in vitro* como *in vivo*. Los estudios *in vitro* (fuera de un organismo) utiliza tejidos, órganos o cultivo de células; mientras que los estudios *in vivo* (dentro de un organismo) emplean organismos vivos como modelos animales. Existen también otros tipos de ensayos preclínicos como lo son los ensayos *in silico* (modelos computacionales), *ex vivo* (tejido vivo fuera de su origen) e *in situ* (en su entorno natural) (Bowes et al., 2012).

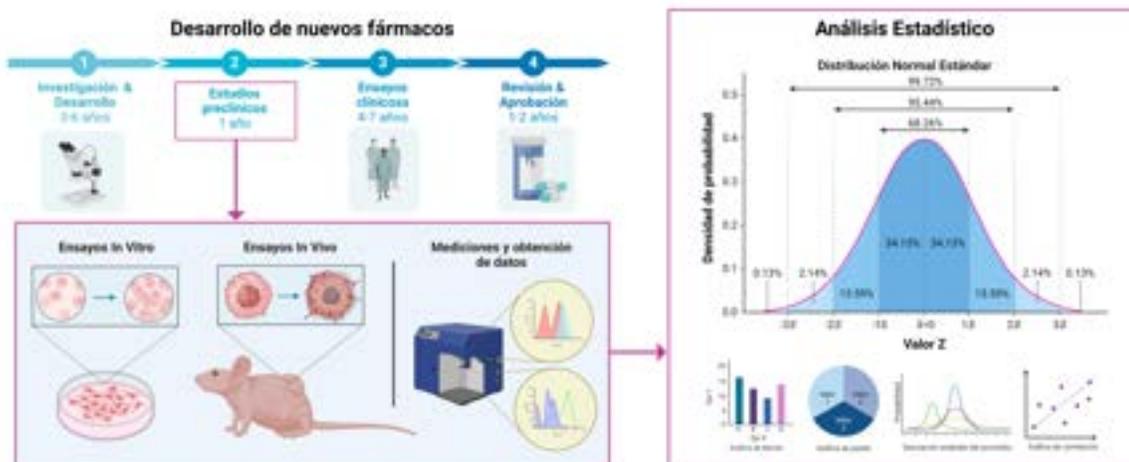


Figura 1. La estadística como parte inherente en la investigación. El desarrollo de nuevos fármacos es un proceso de varias etapas subsecuentes: investigación y desarrollo (I+D), ensayos preclínicos, ensayos clínicos, revisión y aprobación. Los datos obtenidos de las experimentaciones preclínicas, más no limitativas a esta etapa, deben analizarse estadísticamente según su naturaleza para obtener resultados que contribuyan a la toma de decisiones. Estos resultados pueden representarse en diferentes tipos de gráficas dependiendo de lo que se desee analizar. S: desviación estándar. \bar{X} : promedio o media aritmética. Creado en BioRender.

Tras los resultados esperados de los ensayos preclínicos, el desarrollo de nuevos fármacos continúa con los ensayos clínicos. Los investigadores solicitan a las autoridades sanitarias correspondientes la autorización para llevar a cabo el protocolo de investigación del fármaco en seres humanos. En México, la Comisión Federal para la Protección contra Riesgos Sanitarios (COFEPRIS) es la encargada de supervisar esto. Al obtener una respuesta favorable de la autoridad, se comienza con los ensayos clínicos llevados a cabo en cuatro fases: Fase I, con un grupo de entre veinte y ochenta voluntarios no enfermos para determinar seguridad, efectos secundarios, dosis, toxicidad y vía de administración; Fase II, con cientos de voluntarios enfermos para la evaluación de efectividad y seguridad; Fase III, con una población más grande de voluntarios enfermos para el establecimiento de evidencia sobre efectividad y seguridad en miles de personas en comparación con el medicamento estándar; Fase IV de farmacovigilancia del medicamento distribuido de manera comercial (Aban & George, 2015).

Escalamiento de resultados a la industria

Los ensayos preclínicos son una etapa esencial en el desarrollo de nuevos fármacos, ya que proveen los primeros datos sobre su efectividad, no obstante, entre los científicos de la industria y los de la investigación académica, existe un pensamiento común basado en que los artículos publicados en la literatura son difíciles de reproducir. Mientras que en la academia se llevan a cabo ensayos fenotípicos sobre moléculas o dianas específicas, en las grandes empresas farmacéuticas también se



llevan a cabo cribados de alto rendimiento para el análisis a gran escala de fármacos potenciales. En el año 2011, el científico Florian Prinz de la empresa farmacéutica Bayer, realizó un análisis de los resultados de ensayos preclínicos publicados por diferentes investigadores durante los últimos 48 meses en las áreas de oncología, salud femenina y enfermedades cardiovasculares, para ser replicados en su laboratorio. Este análisis reveló que solo el 20-25% de los resultados de estos ensayos publicados tenían relación con los resultados obtenidos por su equipo de trabajo. Algunas de las razones a las que atribuye esta discordancia son un incorrecto o inapropiado análisis estadístico de resultados, tamaño de muestras insuficientes o incluso, la presión a publicar resultados positivos en revistas reconocidas (Prinz et al., 2011).

Cribado de alto rendimiento

El cribado de alto rendimiento (HTS, por sus siglas en inglés) es la estrategia predominante que utilizan las grandes empresas farmacéuticas en el proceso de desarrollo de nuevos fármacos. El HTS a partir de bases de datos robustas consiste en el tamizaje 10,000 – 100,000 compuestos químicos por día, dirigidos a dianas biológicas a través de la aplicación de la automatización, ensayos miniaturizados, análisis de datos a grande escala y otras estrategias como técnicas computacionales y diseño de fármacos basado en fragmentos (Mayr & Bojanic, 2009). El HTS es la opción ideal cuando se tiene poca información de la diana seleccionada sobre su relevancia en la enfermedad estudiada, su eficacia, novedad y espectro del riesgo potencial que posea. Los formatos de los ensayos HTS pueden ser bioquímicos, basados en interacciones y manipulación de células; estos ensayos a su vez deben ser biológicamente relevantes, sensibles, robustos y de bajo costo (Blay et al., 2020).

La robustez de los ensayos de HTS se puede evaluar con métricas de calidad ideales para su optimización y validación. Ejemplo de esto es el factor Z, el cual es un coeficiente adimensional capaz de reflejar la señal del rango dinámico del ensayo y la variación de los datos asociados a las señales de medición (Bajorath, 2002).

Ensayos fenotípicos

El desarrollo y descubrimiento fenotípico de nuevos fármacos (PDD, por sus siglas en inglés) es la estrategia predominante, más no limitativa, en la investigación biotecnológica orientada al área médica/farmacéutica utilizada en la academia. El enfoque del PDD no se basa en el conocimiento de una molécula o diana específica y en la hipótesis de su rol en alguna enfermedad; al contrario, su importancia se basa en su utilidad para resaltar la complejidad de los mecanismos de enfermedades aún sin ser esclarecidos. Los ensayos que abarca el PDD incluyen modelos *in vitro* e *in vivo*, herramientas de edición genómica, organoides y técnicas de imagenología



validadas por el fenotipo molecular para recapitular de manera realista la biología de la enfermedad estudiada, entre otras (Horvath et al., 2016).

El enfoque de investigación PDD, comparado con el HTS, requiere más tiempo por las características propias de los ensayos, por lo que suele ser menos costoso. Sin embargo, los resultados obtenidos suelen ser analizados estadísticamente por investigadores que buscan obtener resultados favorables, lo que puede resultar en problemas posteriores de repetibilidad y reproducibilidad (Moffat et al., 2017).

El análisis estadístico es una parte inherente de la investigación, ya que guía las decisiones del investigador en cada paso del estudio preclínico desde el diseño de los ensayos experimentales hasta el análisis de los datos obtenidos, determinando la potencial escalabilidad de la molécula, fármaco, diana terapéutica o tecnología estudiada.

Estadística y variabilidad en las mediciones

Estadística descriptiva e inferencial

Una investigación se enfoca en una colección definida de datos u objetos que constituyen una población de interés y con características que pueden ser categóricas o numéricas. Cualquier característica cuyo valor puede cambiar de un dato a otro en la población se denomina variable; por lo general, las letras minúsculas son utilizadas para denotar estas variables. Los datos multivariantes surgen cuando se realizan observaciones en más de una variable y de la misma manera pueden ser numéricos o categóricos (Montgomery, 2007).

La estadística es el conjunto de técnicas orientadas a recolectar, organizar, resumir, presentar, analizar datos y generar conclusiones válidas sobre las características de la fuente de donde se obtuvieron los datos, considerando su variación. Las técnicas estadísticas se pueden dividir en estadística descriptiva y en estadística inferencial.

En la estadística descriptiva se recopilan, organizan, resumen, presentan y analizan datos. Al recopilar datos, es posible que el analista desee resumir y describir las características importantes de los mismos de manera gráfica en histogramas, diagramas de cajas, gráficas de puntos, gráfica de dispersión, etcétera. Otros métodos descriptivos implican el cálculo de medidas numéricas como medias, desviaciones estándar, coeficiente de correlación, entre otros. Los softwares computacionales funcionan como herramientas que ayudan al analista a poder diseñar estas tareas permitiendo tener más tiempo para estudiar los datos y extraer mensajes importantes (Green et al., 2022).

La estadística inferencial es la rama dedicada a generar conclusiones sobre las características de la fuente de donde se obtuvieron los datos, en la que se congregan



las técnicas para generalizar desde una muestra a la población. Después de haber obtenido una muestra de la población, el analista podrá utilizar la información muestral para sacar alguna conclusión o hacer inferencia de alguna clase (Marshall & Jonker, 2011).

En general, la investigación preclínica puede clasificarse en dos categorías (ver figura 2), según el objetivo del ensayo: 1) Estudios en los que se formulan hipótesis o estudios exploratorios (estadística descriptiva), en donde se hace un enfoque en cuestiones poco estudiadas por lo que se evalúa la relación entre una o dos variables en un contexto en particular y, 2) Estudios que retan las hipótesis o estudios confirmatorios (estadística inferencial), en los cuales los ensayos parten de hipótesis definidas para posteriormente obtener datos e interpretarlos con los conceptos de análisis estadísticos que se describen a continuación (Aban & George, 2015).

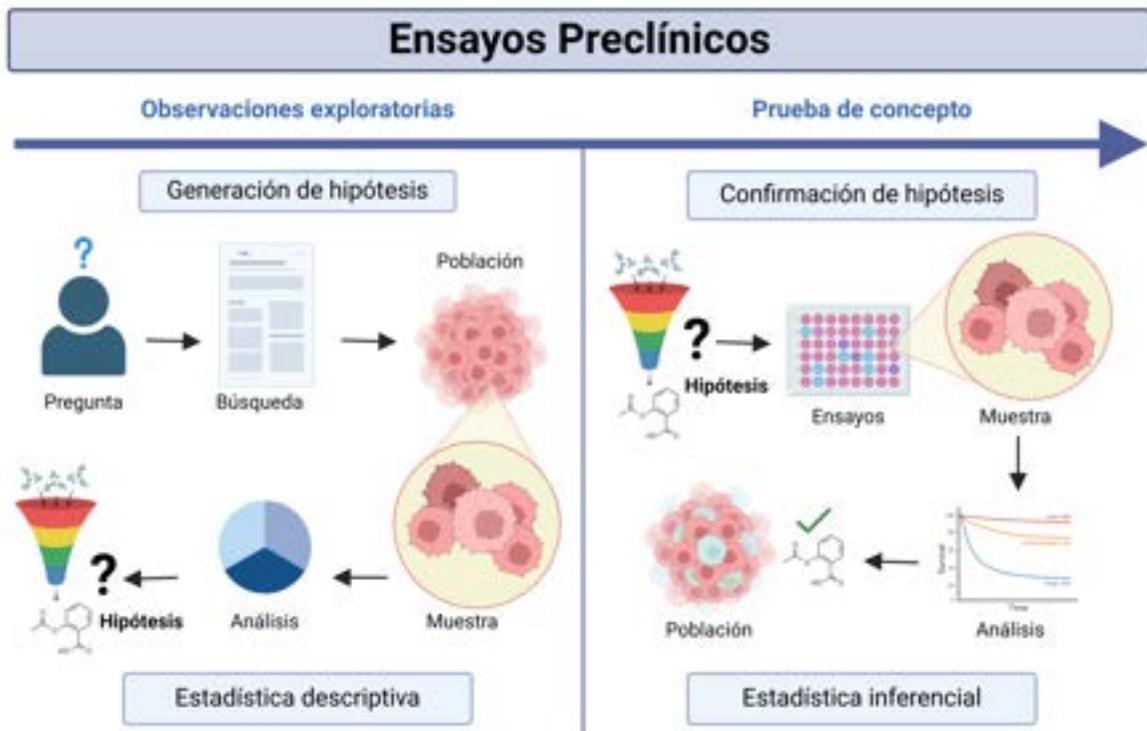


Figura 2. Los ensayos preclínicos se dividen en estudios exploratorios u observacionales y en estudios confirmatorios o de prueba de conceptos. En el primero, lado izquierdo de la figura, los estudios exploratorios toman una muestra de la población para generar una hipótesis mediante el empleo de estadística descriptiva como cálculos numéricos, gráficas y tablas. El segundo, lado derecho de la figura, muestra cómo este tipo de estudios de prueba de concepto parten de una hipótesis ya formulada y se apoya del análisis estadístico de las muestras para inferir que los resultados pueden ser generalizados a la población. Creado en BioRender.



Desviación estándar

La desviación estándar es un parámetro ampliamente utilizado en la estadística inferencial, ya que permite observar la variación de un conjunto de datos y así identificar si existe estabilidad en un ensayo. La desviación estándar es una medida de dispersión e indica qué tan alejados están los datos con respecto a la media (1). Para ello se calcula la distancia que hay entre cada uno de los datos de manera individual x_p , versus la media del conjunto de datos y se denota con la letra S para referirse a la desviación estándar muestral (Devore, 2008).

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (1)$$

En donde: x_i es el valor. \bar{x} , es la media de la muestra. n , es el tamaño de la muestra. Σ , es la suma.

Cuando la diferencia de cada valor individual versus la media es muy pequeña o cercana a cero, entonces la desviación estándar S , tendrá un valor cercano a cero y se dice que hay muy poca variación en el ensayo. Cuando la diferencia es significativamente grande entonces la desviación estándar será mayor a 1 y se dice que hay variación. Para lo anterior, la cantidad de datos n desempeña un papel importante, ya que cuanto mayor es el tamaño de muestra se puede describir con mayor precisión para la toma de decisiones.

Una vez conocida la desviación estándar se pueden realizar más análisis como el cálculo de la varianza o el cuadrado de la desviación estándar S^2 también usado como medida de dispersión, así como, el cálculo del error estándar, intervalos de confianza o pruebas de hipótesis, ya sea para una o más muestras (Yenduri & Silver, 2023).

Error estándar

Cada vez que se realiza una medición bajo estudio, esta va acompañada de cierto tipo de error, conocido como el error estándar el cual necesita ser calculado pues ayuda a determinar qué tan precisas son las estimaciones que se realizan (2). Si el error es cercano al cero, la medición observada es muy cercana a la realidad. Si el error estándar es muy alto, entonces la medición dará un resultado muy diferente a lo que se considera real (Rodrigues et al., 2017):

$$\text{Error estándar} = (S/\sqrt{n}) \quad (2)$$

En donde: S es la desviación estándar de la muestra; n , el tamaño de la muestra.

Una manera de conocer la precisión de una medición es a través de un intervalo de confianza que indica un rango de dónde podría ubicarse el parámetro, es decir, la probabilidad de que este se encuentre entre estos dos valores.



Distribución de Probabilidad

Existen muchas fuentes de variación entre los procesos de estudio, por las diferencias significativas ya que ninguno es igual al otro, como el tiempo requerido para realizar el experimento y los pasos para llevarlo a cabo. Mientras cada elemento se mide, estos pueden ser diferentes entre ellos, pero en grupo forman un patrón que puede describirse como distribución. Las distribuciones se caracterizan de la siguiente manera: 1) Localización, referente al valor central; 2) Dispersión, desde el valor menor al valor más grande y, 3) Forma, patrón de la variación si es simétrica o no (Daimler Chrysler Corporation et al., 2005).

Una distribución de probabilidad es un modelo que explica cómo se relaciona el valor de una variable aleatoria con la probabilidad que tiene esa misma de ocurrir en el ensayo (Gutiérrez Pulido & de la Vara Salazar, 2013). Existen dos tipos de distribución:

1. Distribuciones continuas: cuando una variable que se está midiendo se expresa de manera continua y puede contener cualquier valor dentro de los reales $[-\infty, \infty]$.
2. Distribuciones discretas: cuando la variable que se está midiendo solo puede asumir ciertos valores, como enteros 0, 1, 2.

Regularmente en los ensayos preclínicos, la distribución que más se utiliza es la de tipo continua, ya que los datos con los que se estará trabajando son variables cuantitativas.

Distribución normal

Una representación gráfica de los datos es la distribución de probabilidad normal (ver figura 1), la cual asemeja la forma de una campana. La distribución normal es de las más utilizadas en los análisis estadísticos ya que se espera que cualquier ensayo con datos cuantitativos siga un comportamiento similar. La curva de normalidad describe el comportamiento de los datos alrededor de la media, y de forma visual se puede identificar la dispersión de estos. Si la gráfica de distribución normal es ancha de la base, denota que los datos están dispersos por la gran variación que existe entre ellos lo cual infiere que la desviación estándar también es alta. Por el contrario, si la base de la distribución es estrecha, denota que los datos tienen poca dispersión y que hay muy poca variación por lo que la desviación estándar es baja (The Council for Six Sigma, 2018).



Distribución normal estándar (Z)

La distribución normal estándar es útil para realizar inferencias estadísticas respecto a la media y proporciones con muestras grandes. Una forma de apreciar la desviación estándar como medida de dispersión entorno a la media es la desigualdad de Chebyshev a través de la relación entre la media y la desviación estándar (ver figura 1). En esta desigualdad se afirma que el porcentaje de la medición de los datos, que surgen del análisis en los ensayos, caen en los siguientes intervalos:

- a. En el intervalo desde $\bar{X}-1S$ hasta $\bar{X}+1S$, se encuentra el 68.26% de los datos.
- b. En el intervalo desde $\bar{X}-2S$ hasta $\bar{X}+2S$, se encuentra el 95.44% de los datos.
- c. En el intervalo desde $\bar{X}-3S$ hasta $\bar{X}+3S$, se encuentra el 99.72% de los datos.

Esta regla aplica para la mayoría de los casos que tienen un comportamiento similar al de una distribución de probabilidad normal, que busca relacionar cómo la desviación estándar mide la variabilidad entorno a la media (Salas-Parra et al., 2023).

Distribución t de Student

Cuando se realizan ensayos experimentales es necesario que la cantidad de datos seleccionados en el estudio tengan robustez, para lo cual se selecciona un tamaño de muestra grande ($n>30$). Cuanto mayor sea el número de datos analizados estos seguirán una distribución normal. De forma contraria, si por el tiempo, los recursos u otros factores no se pueden realizar ensayos grandes, se selecciona una muestra pequeña; siempre que el tamaño sea menor a treinta ($n<30$) se utilizará la distribución t de Student (3). Una distribución t se rige por un solo parámetro, llamado grados de libertad de la distribución, abreviado como: gl ($n-1$). Este parámetro se denota con la letra griega ν con valores enteros positivos y cada valor del parámetro ν corresponde a una distribución t diferente. La distribución t es útil para determinar intervalos de confianza y pruebas de hipótesis para la media con muestras pequeñas y σ desconocidas. Para lo que se deberá utilizar el valor del estadístico de prueba T en la distribución normal estándar (Gutiérrez Pulido & de la Vara Salazar, 2013).

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \quad (3)$$

En donde: T es el estadístico de prueba. \bar{x} es la media de la muestra. μ es la media poblacional. S es la desviación estándar de la muestra. n es el tamaño de la muestra.

Las propiedades que determinan a la distribución t son las siguientes:



1. La variable de interés t , para enfatizar que los datos no siguen una distribución estándar cuando el tamaño de muestra n es pequeña.
2. Cada curva t_v tiene forma de campana y con su centro en 0.
3. Cada curva t_v está más parecida a la campana de la distribución normal estándar, curva (z) .
4. Conforme v incrementa la dispersión de t_v correspondiente disminuye.
5. A medida que v incrementa $\rightarrow \infty$ la secuencia de la curva t_v tiende aproximarse más a la distribución normal estándar, por lo que un tamaño de muestra grande arrastra menor % de error en la estimación puntual.

Se denomina $t_{a,v}$ al valor crítico t y es el número sobre el eje de medición con el área bajo la curva t . Para calcular el valor de $t_{a,v}$ se utilizan las tablas de distribución t de Student o paquetes de software computacionales con v grados de libertad a la derecha de $t_{a,v}$ y a es el nivel de confianza deseado (Devore, 2008).

Nivel de confianza

Una estrategia para determinar el tamaño de muestra necesario al hacer una estimación puntual en el ensayo es primero especificar el nivel de confianza deseado. El nivel de confianza α podría ser del 90, 95 o 99%, pero una estimación por intervalo altamente confiable no siempre puede ser precisa, ya que los puntos extremos del intervalo pueden estar lejos, mientras que un intervalo preciso puede acarrear una confiabilidad relativamente baja debido a que el nivel de confianza del intervalo está relacionado inversamente con su precisión; por lo que, se suele utilizar un nivel de confianza del 95% (NIH, 2024).

Intervalo de confianza

La inferencia y las conclusiones estadísticas sobre algún parámetro, se puede deducir con un intervalo de confianza (IC), reportando la probabilidad de que un parámetro esté contenido en el rango (θ_1, θ_2) . Para ello se toman en cuenta las siguientes consideraciones:

- a. IC para la media y muestras grandes $n > 30$ se utiliza la distribución normal estándar
- b. IC para proporciones, también se utiliza la distribución normal estándar.
- c. IC para la media y muestras pequeñas $n < 30$ y varianza desconocida, se utiliza la distribución t de Student.
- d. IC para la varianza, se utiliza la distribución Ji cuadrada.



Para realizar un intervalo de confianza se debe establecer el nivel de significancia (6), el más utilizado es el 95%, así se espera que el parámetro esté dentro del rango con un 95% de probabilidad (PennState, 2024).

Percentiles y valor de p

El área debajo de la curva entre dos puntos es la probabilidad de que una variable distribuida normalmente asuma un valor entre ellos. Una medida de localización son los llamados cuantiles que dividen al conjunto de datos ordenados en grupos que contienen la misma cantidad de datos. Cuando el área debajo de la curva de la distribución Z se divide en 100 partes iguales se obtienen los percentiles (p). Con cualquier valor p entre 0 y 1, se puede hacer uso de las tablas Z o de un software estadístico para obtener el percentil 100° (ver tabla 1). El 95° percentil de la distribución Z es el valor del eje horizontal, tal que el área debajo de la curva es el 0.9500. Es la pregunta contraria al tratar de encontrar la probabilidad al P ($Z=95\%$), por lo que la tabla se utiliza de manera inversa, encontrar el valor que mejor se acerca al 0.9500; $z= 1.645$. Se denotará al $Z_{\alpha} = 100(1 - \alpha)^\circ$ percentil de la distribución normal estándar. En un intervalo de confianza, el nivel de significancia esperado es de $1-\alpha$, valor del área que queda a la izquierda de la distribución (Egermann et al., 2015).

Tabla 1. Percentiles normales estándar y valores críticos

Percentil (p)	Nivel de confianza deseado		
	90	95	99
Área de la curva (α)	0.1	0.05	0.01
$Z_{\alpha} = 100(1-\alpha)$	1.28	1.645	2.33

Fuente: (Gutiérrez Pulido & de la Vara Salazar, 2013)

Si el nivel de confianza dentro de un IC es del 95% entonces se utilizará el área de $(1 - \alpha)$, es decir el 0.05 el área debajo de la curva de distribución normal estándar, por lo que su percentil $Z=1.645$. Esta tabla solo aplica para la distribución normal estándar. Si la distribución es distinta entonces se utilizará la tabla correspondiente. Los percentiles aplican para cualquier distribución continua y es útil para calcular niveles de confianza, así como también para pruebas de hipótesis.

Prueba de Hipótesis

La prueba de hipótesis es una metodología utilizada en la toma de decisiones al realizar ensayos preclínicos. El objetivo de una prueba de hipótesis es el de comprobar si cierta observación es verdadera o falsa, por lo que se generan dos hipótesis: la hipótesis nula (H_0) y la hipótesis alterna (H_A). La H_0 es la que se va a probar y siempre incluye a la igualdad, quiere decir que no tiene ningún efecto en lo que se



está midiendo; por otro lado, la H_A se refiere a que el experimento influye significativamente en la variable de respuesta (Wisniewski et al., 2018).

En una prueba de hipótesis el valor p , que puede ser calculado con percentiles de la distribución, reporta la probabilidad de observar un efecto tan grande como aquel que se está midiendo si la H_0 fuera verdadera. Cuanto más pequeño sea el valor p , se puede concluir que la H_0 no se puede aceptar. Cuando la H_0 no se pueda aceptar o comprobar que es verdadera con la evidencia estadística, entonces se define que la variable de interés de salida tiene un efecto en las mediciones de la prueba (Huang et al., 2020).

El nivel de significancia se define al inicio de la prueba. Cuanto más pequeño sea el nivel de confianza ($1 - \alpha$) el efecto para poder detectar si hay relevancia o no también será mínimo. Este valor debe estar predefinido de manera que pudiera detectar que hay una relevancia biológica en el experimento. El nivel de significancia deberá calcularse junto con el tamaño de la muestra y con esto confirmar que el experimento tiene el potencial suficiente de detectar efectos relevantes (The Council for Six Sigma, 2018).

Antes de hacer conclusiones en el análisis, primero se debe probar la hipótesis de que los datos siguen una distribución normal y de ser significativa, no rechazar la hipótesis y tomar la decisión correcta. Si no se realiza la prueba de hipótesis de la normalidad de los datos, no se deberá continuar con el análisis, pues las decisiones tomadas podrían ser erróneas. Si el valor p generado de la prueba es menor que el nivel α ($p < 0.05$) seleccionado, entonces la H_0 se rechaza y la H_A se acepta. Si el valor p generado de la prueba es mayor que el nivel α ($p > 0.05$) se acepta la H_0 por lo que la H_A se rechaza (The Council for Six Sigma, 2018).

Análisis de varianza

El análisis de la varianza (ANOVA, por sus siglas en inglés), es un procedimiento utilizado para determinar si los datos de tres o más muestras, provenientes de diferentes poblaciones, apuntan diferencias entre las medias de estas poblaciones. Durante la prueba de hipótesis del análisis de la varianza, la H_0 indica que todas las medias de los tratamientos son iguales entre sí, y la H_A , indica que al menos un par de medias es diferente. Los cálculos deben cumplir la suposición de que los datos en ANOVA son independientes y siguen una distribución normal (realizar prueba de hipótesis de normalidad de los datos) y con varianzas iguales (4). La variabilidad total de los datos se mide como la variación de los tratamientos más la variación debido al error (Kubiak & Benbow, 2009).

$$\sigma_{Total}^2 = \sigma_{Tratamientos}^2 + \sigma_{Error}^2 \quad (4)$$



En donde: la variación de los tratamientos $\sigma_{Tratamientos}^2$, se calcula como la suma de los cuadrados de las diferencias del promedio de cada dato y el gran promedio total; la variación debido al error σ_{Error}^2 , se calcula como la suma de los cuadrados de las diferencias con los tratamientos y el promedio del tratamiento.

Las partes que componen a la variación son las que se muestran en la tabla resultado del estudio ANOVA, a través de software de computadora (ver tabla 2). En donde k son los números de muestras (tratamientos), n es el número de lecturas por tratamiento y, N es la cantidad total de datos. La media de los cuadrados del error es la variación total del estudio y se espera que sea menor a la variación entre los tratamientos. El estadístico F compara la variación entre tratamientos y la variación debido al error.

Tabla 2. Análisis de la varianza y fuentes de variación

Fuentes de variación	Suma de los cuadrados (SC)	Grados de libertad	Media de los cuadrados (MC)	Estadístico F	Valor-p
Variación debido a los tratamientos	$SC_{Tratamientos}$	k-1	$MC = \frac{SCt}{k-1}$	$F_0 = \frac{MCt}{MCE}$	p < 0.05
Variación debido al error	SC_{error}	N-k	$MCE = \frac{SCe}{N-k}$		
Variación Total	SC_{Total}	N-1			

Fuente: (Kubiak & Benbow, 2009)

Si la MCE es muy grande comparado con MC, entonces el valor F_0 será muy pequeño y no se puede considerar el estudio ANOVA como aceptable. Si MC es más grande, se tendrá que hacer la prueba de hipótesis con el valor α seleccionado previamente. Cuando el valor p sea menor que α (0.05) entonces la H_0 se rechaza y se concluye que al menos un par de medias es diferente. Si el valor p es mayor que α (0.05) entonces la H_0 se acepta y se infiere que la media de los tratamientos es igual entre sí.

Prueba de Tukey

La prueba de ANOVA puede detectar si todas las medias de los tratamientos son iguales o al menos algún par es diferente, pero no puede determinar cuál es el distinto. La prueba de Tukey es utilizada para comparar cada una de las medias en pares e indicar si esos pares son iguales o distintos. El estadístico de prueba de Tukey se calcula con la raíz cuadrada del MCE contra el tamaño de muestra de cada una de ellas a un nivel de significancia α . Después, se compara la diferencia de medias formando pares: la media de la muestra uno contra la media de la muestra dos, media de la muestra uno contra la media de la muestra tres y así sucesivamente hasta terminar de formar todos los pares posibles. Cualquier par de tratamientos en donde



la diferencia de medias sea mayor que el estadístico de Tukey, implica que este par de muestras son diferentes entre sí (Montgomery, 2013).

Lean Six Sigma

A nivel industrial existen estrategias útiles para la gestión de la calidad, las cuales pueden ser aplicadas en el análisis de los procesos y de los pasos que se llevan a cabo durante los ensayos de investigación tanto preclínica como clínica. A partir del año 2000, en Estados Unidos de América se comenzó a implementar la filosofía Lean Six Sigma en proyectos para la mejora de la calidad en los servicios de salud a nivel hospitalario, dando resultados positivos. Sin embargo, cuando se habla de la estrategia Lean Six Sigma, regularmente se piensa en un solo concepto, pues ambos, Lean Manufacturing y Six Sigma, se enfocan en la calidad de los procesos, pero en realidad se trata de dos filosofías (Schweikhart & Dembe, 2009).

Six Sigma

La filosofía Six Sigma es utilizada para mejorar la calidad y la eficiencia de los procesos operacionales, esta se enfoca en hacer procesos más uniformes y precisos a través de la aplicación de herramientas estadísticas. El objetivo principal de Six Sigma es el de reducir la variación para mejorar la calidad a un nivel de competitividad mundial, lo que implica tener un desempeño de proceso de 3.4 defectos por millón de oportunidades (y su estabilidad en el tiempo, considerando que la media se puede desplazar 1.5 veces la desviación estándar). La perspectiva filosófica de Six Sigma es ver todas las actividades como un proceso que necesita ser definido, medido, analizado, mejorado y controlado (DMAIC, por sus siglas en inglés). Six Sigma como conjunto de herramientas incluye técnicas cualitativas (mapeo de procesos) y cuantitativas (análisis estadístico), que pueden ser conducidas por un experto para llevar a cabo el proceso de mejora (Munro et al., 2015).

Lean Manufacturing

Por otra parte, Lean Manufacturing involucra una serie de principios, métodos y prácticas que ayudan a diseñar, mejorar y administrar una organización. El principal objetivo de Lean es mejorar la eficiencia eliminando los desperdicios que pueden llegar a absorber tiempo y recursos, pero sin agregar valor: sobre producción, sobre inventarios, área, transporte, movimientos, tiempo de espera o mala calidad. Una de las principales herramientas es el mapeo del flujo de valor (VSM, por sus siglas en inglés) que ayudan a identificar de manera visual los pasos requeridos para completar una tarea con la finalidad de observar los recursos que están involucrados como personal clave, flujo de información y actividades, de manera que se puedan iden-



tificar las oportunidades e integrar los pasos del proceso en términos de eficiencia (The Council for Six Sigma, 2018).

Lean Six Sigma en ensayos preclínicos

Las metodologías Lean Six Sigma fueron desarrolladas en el complejo ambiente de manufactura y sistemas ingenieriles en donde la calidad, la precisión y el consumidor son puntos tan críticos para el éxito de los proyectos como lo son en la investigación biomédica, durante el proceso de desarrollo de nuevos fármacos tanto en ensayos preclínicos como clínicos.

Son pocos los reportes en la literatura sobre la aplicación de Lean Six Sigma en los ensayos preclínicos en comparación con los ensayos clínicos y servicios de salud. Sin embargo, Lean Six Sigma se ha recomendado para evitar contaminación cruzada en cultivos celulares (Lindgren, 2008) y para reducir errores de laboratorio en diferentes disciplinas incluyendo la biología molecular, hematología, citología, microbiología y patología (Hollensead et al., 2004). De igual manera, esta filosofía se ha direccionado al aseguramiento de la calidad en laboratorios farmacéuticos e industrias relacionadas; el ejemplo más reciente es la empresa internacional de biotecnología Novozymes, en donde se aplicó un enfoque Lean Six Sigma para analizar la cultura de la empresa, el flujo de los procesos, y el trabajo de los analistas responsables de llevar a cabo ensayos preclínicos, dando como resultado la implementación de la metodología 5S mediante cinco pasos (clasificar, ordenar, limpiar, estandarizar, mantener) logrando así una mejor utilización de los espacios de trabajo de cada analista en el laboratorio (Shanley, 2020).

Lean Six Sigma en ensayos clínicos

Las metodologías Lean Six Sigma son aplicables en la parte preclínica y clínica del proceso de desarrollo de nuevos fármacos debido a la evidente necesidad de precisión estadística, calidad en las mediciones durante las experimentaciones y reproducibilidad de los ensayos en el laboratorio. Se han reportado en la literatura casos de éxito en la aplicación de Lean Six Sigma en la práctica clínica. Por ejemplo, en laboratorios clínicos donde se ven reducidos errores en los sistemas de captura de la información de pacientes (Riebling N, 2005), se manejan los volúmenes necesarios para análisis sanguíneos (Elder, 2008), se asegura la repetibilidad y reproducibilidad de ensayos anticoagulantes en diferentes laboratorios de una comunidad (Hurley et al., 2008) y se establece un flujo de trabajo eficiente y continuo dentro de un laboratorio histológico hospitalario (Schweikhart & Dembe, 2009).

En la actualidad, son numerosas las organizaciones que han implementado la estrategia Lean Six Sigma en la mejora del desempeño ya que es un conjunto de he-



ramientas sumamente útiles para ayudar a incrementar la eficiencia y la efectividad de los procesos reduciendo al mínimo los errores.

Conclusiones

Los ensayos preclínicos son esenciales en el proceso de desarrollo de nuevos fármacos al proveer los primeros datos sobre la efectividad de una nueva terapia. El análisis estadístico de estos datos debe estar alineado a las características propias del ensayo, es decir, para obtener resultados confiables, repetibles y reproducibles se deben aplicar los conceptos estadísticos que correspondan a la naturaleza de este.

Se propone la implementación de la filosofía Lean Six Sigma en las actividades diarias realizadas por todo analista dentro de un laboratorio de investigación mediante la estandarización de metodologías, el análisis y solución de problemas desde la causa raíz reduciendo así las fuentes de variación de manera sistémica, aumentando la eficacia de los procesos y el éxito de los ensayos preclínicos.

Conflicto de intereses

Los autores declaran que la investigación se realizó en ausencia de cualquier relación comercial o financiera que pudiera interpretarse como un potencial conflicto de interés.

Agradecimientos

Sarah Eliuth Ochoa Hugo con CVU 891162 agradece a CONAHCYT por la beca otorgada para su formación en el programa de Doctorado en Ciencias en Innovación Biotecnológica de CIATEJ. Fondo CONAHCYT, FORDECYT-PRONACES/568483/2020.

Referencias

- Aban, I. B., & George, B. (2015). Statistical Considerations for Preclinical Studies. *Experimental Neurology*, 270, 82. <https://doi.org/10.1016/J.EXPNEUROL.2015.02.024>
- Bajorath, J. (2002). Integration of virtual and high-throughput screening. *Nature Reviews Drug Discovery*, 1(11), 882-894. <https://doi.org/10.1038/nrd941>
- Blay, V., Tolani, B., Ho, S. P., & Arkin, M. R. (2020). High-Throughput Screening: today's biochemical and cell-based approaches. *Drug Discovery Today*, 25(10), 1807–1821. <https://doi.org/10.1016/j.drudis.2020.07.024>
- Bowes, J., Brown, A. J., Hamon, J., Jarolimek, W., Sridhar, A., Waldron, G., & Whitebread, S. (2012). Reducing safety-related drug attrition: The use of *in vitro* pharmacological profiling. *Nature Reviews Drug Discovery*, 11(12), 909-922. <https://doi.org/10.1038/nrd3845>



- Daimler Chrysler Corporation, General Motors Corporation, & Ford Motor Company. (2005). *Statistical Process Control (Segunda Edición)*. Automotive Industry Action Group.
- Devore, J. L. (2008). *Probabilidad y Estadística para Ingeniería y Ciencias (Séptima Edición)*. Cengage Learning.
- Egermann, H., Fernando, N., Chuen, L., & McAdams, S. (2015). P-values as percentiles. Commentary on: “Null hypothesis significance tests. A mix-up of two different theories: the basis for widespread confusion and numerous misinterpretations”. *Frontiers in Psychology*, 6(JAN), 411–432. <https://doi.org/10.3389/FPSYG.2015.00341>
- Elder, B. L. (2008). Six Sigma in the microbiology laboratory. *Clinical Microbiology*, 30(19), 143-147. <https://doi.org/10.1016/j.clinmicnews.2008.09.001>
- Green, J. L., Manski, S. E., Hansen, T. A., & Broatch, J. E. (2022). Descriptive statistics. En R. J. Tierney, F. Rizvi & K. Ercikan (eds.), *International Encyclopedia of Education: Fourth Edition* (pp. 723–733). Elsevier Science <https://doi.org/10.1016/B978-0-12-818630-5.10083-1>
- Gutiérrez Pulido, H., & de la Vara Salazar, R. (2013). Herramientas básicas para Seis Sigma. En *Control Estadístico de La Calidad y Seis Sigma* (pp. 134–164). McGrawHill.
- Hollensead, S. C., Lockwood, W. B., & Elin, R. J. (2004). Errors in pathology and laboratory medicine: consequences and prevention. *Journal of Surgical Oncology*, 88(3), 161–181. <https://doi.org/10.1002/JSO.20125>
- Horvath, P., Aulner, N., Bickle, M., Davies, A. M., Nery, E. Del, Ebner, D., Montoya, M. C., Östling, P., Pietiäinen, V., Price, L. S., Shorte, S. L., Turcatti, G., Von Schantz, C., & Carragher, N. O. (2016). Screening out irrelevant cell-based models of disease. *Nature Reviews Drug Discovery*, 15(11),751–769). <https://doi.org/10.1038/nrd.2016.175>
- Huang, W., Percie du Sert, N., Vollert, J., & Rice, A. S. C. (2020). Good Research Practice in Non-Clinical Pharmacology and Biomedicine. *Handbook of Experimental Pharmacology*, 257, 55. https://doi.org/10.1007/164_2019_277
- Hurley, B., Taylor, T., Levett, J., Huber, C., & Hahn, E. (2008). Implementation of six sigma and lean methodology into the anticoagulation management process. *Journal of Thrombosis and Thrombolysis*, 25(1), 106–106. <https://doi.org/10.1007/S11239-007-0148-Z>
- Kubiak, T. M., & Benbow, D. W. (2009). *The Certified Six Sigma Black Belt Handbook*. ASQ Quality Press - Scientific Research Publishing. <https://www.scirp.org/reference/referencespapers?referenceid=1790270>



- Lindgren, V. (2008). To err is human; to follow the SOP divine. *International Journal of Cancer*, 123(4), 979–980. <https://doi.org/10.1002/IJC.23567>
- Marshall, G., & Jonker, L. (2011). An introduction to inferential statistics: A review and practical guide. *Radiography*, 17(1), e1-e6. <https://doi.org/10.1016/j.radi.2009.12.006>
- Mayr, L. M., & Bojanic, D. (2009). Novel trends in high-throughput screening. *Current Opinion in Pharmacology*, 9(5), 580–588). <https://doi.org/10.1016/j.coph.2009.08.004>
- Moffat, J. G., Vincent, F., Lee, J. A., Eder, J., & Prunotto, M. (2017). Opportunities and challenges in phenotypic drug discovery: An industry perspective. *Nature Reviews Drug Discovery*, 16(8), 531–543. <https://doi.org/10.1038/nrd.2017.111>
- Montgomery, D. (2007). *Control estadístico de la calidad (Tercera Edición)*. Limusa Wiley.
- Montgomery, D. (2013). *Design and analysis of experiments (Octava Edición)*. John Wiley & Sons, Ltd.
- Munro, R. A., Ramu, G., & Zrymiak, D. J. (2015). *The Certified Green Belt Handbook (M. T. Meinholz, Ed.; Segunda Edición)*. American Society for Quality.
- National Library of Medicine. (s.f.). *Common Terms and Equations*. Recuperado el 11 de diciembre de 2024 de <https://www.nlm.nih.gov/oet/ed/stats/02-950.html>
- PennState. (2024). *Introduction to Confidence Intervals | STAT 200*. The Pennsylvania State University. <https://online.stat.psu.edu/stat200/lesson/4/4.2>
- Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, 10(9), 712–713. <https://doi.org/10.1038/NRD3439-C1>
- Riebling N, T. L. (2005). Six sigma project reduces analytical errors in an automated lab. *MLO: medical laboratory observer*, 37(6), 20–23. . <https://pubmed.ncbi.nlm.nih.gov/16028478/>
- Rodrigues, C. F. de S., Lima, F. J. C. de, & Barbosa, F. T. (2017). Importance of using basic statistics adequately in clinical research. *Brazilian Journal of Anesthesiology (English Edition)*, 67(6), 619–625. <https://doi.org/10.1016/j.bjane.2017.01.011>
- Salas-Parra, R. D., Castro-Ochoa, K. J., & Machado-Aranda, D. A. (2023). Basic science statistics. En A. E.M. Eltorai, J. A. Bakal, P. C. Newell & A. J. Osband (eds.), *Translational Surgery* (pp. 195–199). Academic Press <https://doi.org/10.1016/B978-0-323-90300-4.00022-7>
- Schweikhart, S. A., & Dembe, A. E. (2009). The Applicability of Lean and Six Sigma Techniques to Clinical and Translational Research. *Journal of Investigative Medicine : The Official Publication of the American Federation for Clinical Research*, 57(7), 748. <https://doi.org/10.231/JIM.0B013E3181B91B3A>



- Shanley, A. (2020). *Using Six Sigma and QbD to Improve Preclinical R&D*. <https://www.pharmtech.com/view/bringing-six-sigma-to-preclinical-pharma-r-d>
- The Council for Six Sigma. (2018). *Six Sigma: a complete step-by-step guide*. The Council for Six Sigma.
- Wisniewski, C. S., Jones, E. P., Weeda, E. R., Pilch, N. A., & Picone, M. F. (2018). Medical literature evaluation and biostatistics. En D. Thomas (ed.), *Clinical Pharmacy Education, Practice and Research: Clinical Pharmacy, Drug Information, Pharmacovigilance, Pharmacoeconomics and Clinical Research* (pp. 143–162). Elsevier. <https://doi.org/10.1016/B978-0-12-814276-9.00010-6>
- Yenduri, R., & Silver, R. E. (2023). Basic statistical principles. En A. E.M. Eltorai, J. A. Bakal, P. C. Newell & A. J. Osband (eds.), *Translational Surgery* (pp. 139–143). Academic Press. <https://doi.org/10.1016/B978-0-323-90300-4.00049-5>

