

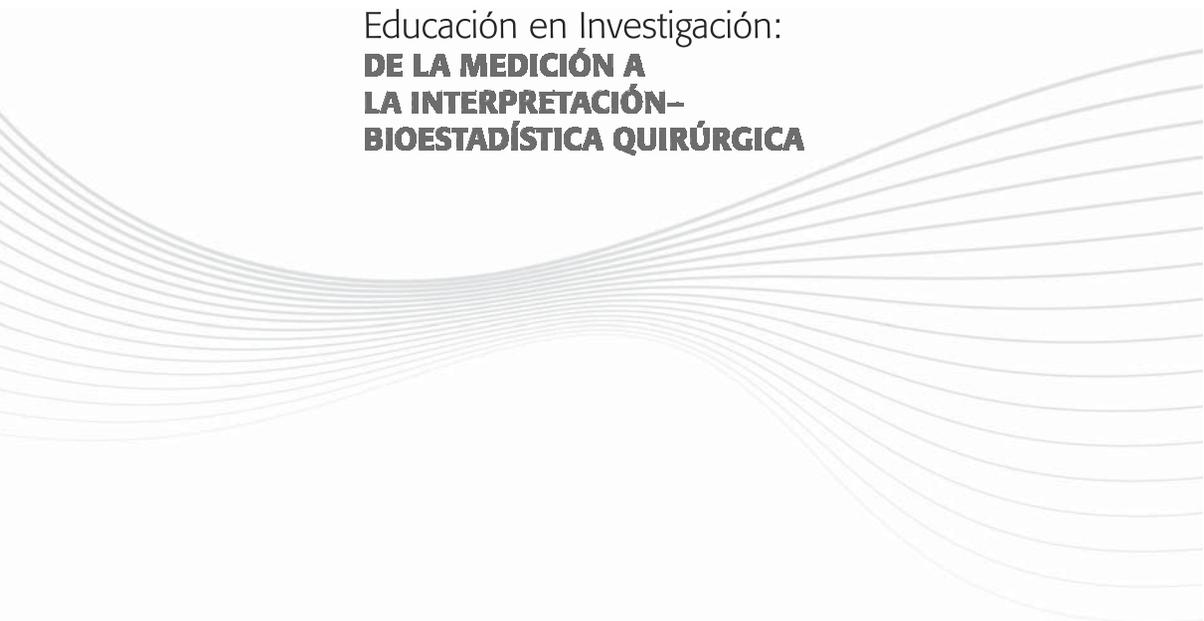
Educación en Investigación: **DE LA MEDICIÓN A LA INTERPRETACIÓN- BIOESTADÍSTICA QUIRÚRGICA**

Organizadores

José María Jiménez Avila
Asdrubal Falavigna



Educación en Investigación:
**DE LA MEDICIÓN A
LA INTERPRETACIÓN-
BIOESTADÍSTICA QUIRÚRGICA**

A decorative graphic consisting of multiple thin, light gray lines that flow from the left side of the page, curve downwards and then upwards, and exit on the right side, creating a sense of movement and depth.

José María Jiménez Avila
Asdrubal Falavigna
Organizadores

Educación en Investigación:
DE LA MEDICIÓN A
LA INTERPRETACIÓN-
BIOESTADÍSTICA QUIRÚRGICA



EDUCS

COMUNIDAD

© de los organizadores

Capa: André Secundino

Corrección: Carla Vinize Casagrande Gianini

Editorial: Traço Diferencial | (54) 3229 7740 | (54) 99901 3978

Datos Internacionales de Catalogación en la Publicación (CIP)
Universidad de Caxias del Sur
UCS - BICE - Procesamiento Técnico

E24 Educación em investigación : de la medición a la interpretación-bioestadística quirúrgica / org. José María Jiménez Avila, Asdrubal Falavigna. – Caxias do Sul, RS : Educs, 2018.
210 p.: il.; 23 cm.
ISBN 978-85-7061-911-2
Apresenta bibliografía
Vários colaboradores

1. Bioestadística. 2. Cirugía – Bioestadística. 3. Investigación médica. I. Falavigna, Asdrubal. II. Jiménez Avila, José María.

CDU 2. ed.: 57.087.1

Índice al catálogo sistemático:

1. Bioestadística	57.087.1
2. Cirugía – Bioestadística	617:57.087.1
3. Investigación médica	61:001.8

Catalogación en la fuente elaborada por la bibliotecaria
Ana Guimarães Pereira – CRB 10/1460

Derechos reservados a:



ASSOCIACION AOSPINE LATIN AMERICA

Av. Silva Jardim 2042 Cj. 1505 | Cep 80250-200 | Curitiba – PR | Brasil

Teléfono: 55 41 3016 4491 | E-mail: aosla@aospine.org

RESUMEN

Nota de los organizadores / 11

Colaboradores / 13

Prólogo / 17

UNIDAD 1 / 19

Capítulo 1 INTRODUCCIÓN / 21

José María Jiménez Avila e Asdrubal Falavigna

Capítulo 2 MATEMÁTICAS: LENGUAJE E INTERPRETACIÓN / 23

Maurício Scussel

Capítulo 3 EL ACTOR PRINCIPAL. LA VARIABLE / 29

José María Jiménez Avila

Capítulo 4 ¿CÓMO PLANEAR LA ELABORACIÓN DE UNA BASE DE DATOS? / 33

Pedro Luis Bazán

Capítulo 5 DETERMINANDO EL TAMAÑO DE LA MUESTRA / 45

Emiliano Vialle

Capítulo 6 ¿CÓMO ELEGIR LAS PRUEBAS ESTADÍSTICAS? / 53

Pedro Luis Bazán

Capítulo 7 BIOESTADÍSTICA DESCRIPTIVA / 63

Alberto Aceves Pérez

Capítulo 8 BIOESTADÍSTICA ANALÍTICA O INFERENCIAL / 69

José María Jiménez Avila

Capítulo 9 MEDIDAS DE TENDENCIA CENTRAL / 75

Alberto Aceves Pérez

Capítulo 10
**MEDIDAS DE DISPERSIÓN (RANGO, VARIANZA, DESVIACIÓN ESTÁNDAR,
COEFICIENTE DE VARIANZA) / 81**

Delio Martins

Capítulo 11
PRESENTACIÓN DE LOS DATOS / 87

José María Jiménez Avila

Capítulo 12
PRUEBA DE HIPÓTESIS (IC) Y VALOR DE “p” / 93

José María Jiménez Avila

Capítulo 13
TEST DE KOLMOGOROV-SMIRNOV / 99

Nelson Astur Neto

Maximiliano Gimenez Gigón

Rodrigo Góes Medéa de Mendonça

Capítulo 14
PRUEBA DE SHAPIRO-WILK / 105

Nelson Astur Neto

Rodrigo Góes Medéa de Mendonça

UNIDAD 2 / 109

Capítulo 15
PRUEBA t DE STUDENT / 111

Barón Zarate

Francisco López Meléndez

Capítulo 16
PRUEBA DE ANOVA / 121

Juan Pablo Guyot

Capítulo 17
CORRELACIÓN DE PEARSON / 125

Alisson R. Teles

Asdrubal Falavigna

UNIDAD 3 / 133

Capítulo 18
CHI² (JI²) / 135

Alberto Aceves Pérez

Capítulo 19
TEST O PRUEBA DE RANGOS DE WILCOXON / 139

Baron Zarate

Capítulo 20**U DE MANN-WHITNEY / 149**

José Manuel Pérez Atanasio
Víctor Fernando Luján Celis
Maritza Belén Sandoval Rincón

Capítulo 21**PRUEBA DE KRUSKAL-WALLIS / 153**

Esteban Araya Ramírez
Fabián Víquez Monge

Capítulo 22**COEFICIENTE DE CORRELACIÓN DE SPEARMAN / 157**

Fernando Alvarado
Miguel Farfán

Capítulo 23**PRUEBA DE MCNEMAR / 163**

Esteban Araya Ramírez
Fabián Víquez Monge

Capítulo 24**REGRESIÓN LINEAL SIMPLE / 169**

Alisson R. Teles
Asdrubal Falavigna

Capítulo 25**DISTRIBUCIÓN DE POISSON / 175**

José Acuña Pinilla
Margarita Gómez Chantraine
Elizabeth Heins

Capítulo 26**ODDS RATIO / 187**

José Manuel Pérez Atanasio
Víctor Fernando Luján Celis
Maritza Belén Sandoval Rincón

UNIDAD 4 / 193**Capítulo 27****LO QUE DA VALOR Y FUERZA A LAS EVIDENCIAS / 195**

Mohsen Sadeghi Naini
Vafa Rahimi Movaghar

AGRADECIMIENTO

Agradecemos a los autores por su valiosa contribución en la elaboración de los capítulos, y los felicitamos por la calidad didáctica de simplificar el complejo mundo de la bioestadística.

Nuestro agradecimiento a AOSpine International Research Commission Members por el apoyo incondicional en la elaboración del libro.

Agradecemos también el soporte de los miembros del Board de AOSpine Latinoamérica y de la oficina por el apoyo logístico para la finalización de este gran proyecto.

Publicar este 2º libro es motivo de gran satisfacción, ya que estamos seguros de su contribución al fortalecimiento diario del conocimiento de la investigación clínica y experimental.

José María Jiménez Avila

Nuevamente, mi reconocimiento a Georgina, Zaira y Daniela por haber dividido sus horas y, a veces, sus días de convivencia con este libro. Siempre han sido mi motivación y mi razón para continuar con este tipo de proyectos.

Un reconocimiento y agradecimiento muy especial al Dr. Asdrubal Falavigna, actual Chairperson de AOSpine Latinoamérica, quien siempre confió en el proyecto y lo hace posible gracias a su visión y liderazgo.

Asdrubal Falavigna

Dedico este libro a mi esposa, Ana Maria, y a mis hijos Vincenzo y Rebecca, por el apoyo diario. Deseo que este libro potencialice el espíritu científico en cada estudiante y profesional, objetivando capacitarlos a innovar y generar conocimiento, con la finalidad de buscar respuestas a nuestras inquietudes diarias.

NOTA DE LOS ORGANIZADORES

Existe una demanda reprimida para la calificación en investigación, sea para el desarrollo de un pensamiento reflexivo o para un análisis crítica de los hechos. Es fundamental la generación de nuevos conocimientos y productos de innovación que potencien los avances de la ciencia y permitan su aplicación al desarrollo de la sociedad en sus distintas dimensiones.

El aprendizaje del conocimiento de bioestadístico normalmente es largo y difícil. Este libro enfoca los principios básicos de la estadística aplicada al especialista del área de la salud bajo la visión de especialistas e investigadores oriundos de diferentes áreas del conocimiento. En esa jornada, el investigador encontrará sugerencias y ejemplos que facilitarán el entendimiento de esta compleja área, entre otros, problemas en el entendimiento del significado de $p < 0.05$, cuando elija un método estadístico sobre otro, etc. Estos son los obstáculos que se deben superar para volverse un profesional crítico y reflexivo capaz de intervenir en la realidad y producir conocimiento específico del área, de modo a atender las demandas de la población.

La lectura de los capítulos permite que el lector adquiera conocimiento de las dificultades comúnmente encontradas en el área de bioestadística. Al final, nada puede recompensar más que ver su trabajo publicado tras varios esfuerzos, disciplina y dedicación.

José María Jiménez Avila
Asdrubal Falavigna

COLABORADORES

Alberto Aceves Pérez

Ortopedista. Centro Médico Nacional de Occidente. Instituto Mexicano del Seguro Social, Guadalajara, Jalisco, México. Profesor de asignatura de la Universidad de Guadalajara. AOSpine Research Officer México.

Alisson R Teles

Neurocirujano de la columna vertebral. McGill Scoliosis & Spine Group. McGill University, Montreal, Canadá.

Asdrubal Falavigna

AOSpine Latin America Chairperson. Director del Centro de Ciencia de la Vida de la Universidad de Caxias do Sul. Coordinador del Science Post-Graduation Program. Profesor de Neurocirugía de la Universidad de Caxias do Sul. Coordinador del Grupo de Investigación Estudios Clínicos y Modelos Básicos de Patologías Raquimedulares y del Laboratorio de Patologías Raquimedulares de la UCS. Caxias do Sul, RS – Brasil.

Barón Zárate Kalfópulos

Cirujano de Columna. Instituto Nacional de Rehabilitación / Medica Sur Ciudad de México. Profesor Alta Especialidad Cirugía de Columna Universidad Nacional Autónoma de México. Director Centro AOSpine Ciudad de México.

Delio Martins

Ortopedista. Hospital Israelita Albert Einstein. Doctor en Ciencias por la Universidad Federal de São Paulo (UNIFESP); Coordinador del Curso de Perfeccionamiento en Cirugía de Columna del Hospital Israelita Albert Einstein; Docente de Ortopedia de la Universidad Anhembi-Morumbi.

Elizabeth Heins

Médico Cirujano Pontificia Universidad Javeriana. Epidemióloga- Johannes Gutenberg Universität Mainz. Senior Manager de Medical Affairs Johnson & Johnson Medical Devices Northern Cluster.

Emiliano Vialle

Jefe del grupo de cirugía de columna, Hospital Universitário Cajuru, Banco de Tejidos Humanos, Pontificia Universidad Católica de Paraná, AOSpine Latin America Fellowship Officer. Brasil.

Esteban Araya

Ortopedista, Universidad de San Carlos de Guatemala. Cirujano de Columna, Universidad Militar Nueva Granada, Colombia. Profesor Departamento de cirugía Universidad de Ciencias Médicas, Costa Rica. Coordinador Clínica de Columna Hospital del Trauma y Coordinador Médico Banco de Tejidos Hospital del Trauma, Ex Jefe de Servicios Quirúrgicos Hospital del Trauma, Costa Rica.

Fabian Vilquez Monge

Máster en Seguros y Gerencia de Riesgos, de la Universidad Pontificia de Salamanca, España. Técnico en Riesgos Asociados a la Operación, Universidad de Costa Rica. Bachiller en Estadística, Universidad de Costa Rica.

Fernando Alvarado Gomez

Cirujano de Columna - Hospital Universitario Fundación Santa Fe de Bogotá. Cirujano de Columna - Instituto de Ortopedia Infantil Roosevelt. Cirugía de Columna, Pelvis y Acetábulo - Universidad Militar Nueva Granada. Ortopedia y Traumatología - Pontificia Universidad Javeriana. Epidemiólogo Clínico – Universidad Nuestra Señora del Rosario. Docente clínico de Ortopedia y Traumatología de la Universidad de los Andes. Docente asociado de Ortopedia y Traumatología de la Universidad Nuestra Señora del Rosario. Docente Ad Honorem del programa de cirugía de columna de la Universidad Militar Nueva Granada. Research AOSpine Colombia. Secretario del Capitulo de Columna SCCOT.

Francisco Manuel López Melendez

Cirujano de Columna. ISSSTE Hospital General Santiago Ramón y Cajal. Durango. México.

José Acuña Pinilla

Instrumentador Quirúrgico Profesional, Universidad de Santander UDES. Coordinador de Medicina Basada en la Evidencia y Asuntos Médicos para la Región Norte Latinoamérica división One Medical Devices de la compañía Johnson & Johnson. Profesor de la Universidad de Santander - UDES, Bogotá.

José Manuel Pérez Atanasio

Médico ortopedista. Maestro en Ciencias. Jefe de la División de Investigación en Salud. UMAE “Dr. Victorio de la Fuente Narváez” IMSS. Ciudad de México.

José María Jiménez Avila

Ortopedista, Instituto Tecnológico y de Estudios Superiores de Monterrey (TEC), campus Guadalajara, Profesor Departamento de Ciencias Básicas y Coordinador de Clínica de Ortopedia y Traumatología, Instituto Mexicano del Seguro Social, Clínica de Columna, Regional Research Officer de AOSpine Latinoamérica. México.

Juan Pablo Guyot

Cirugía de Columna, Ortopedia y Traumatología. Hospital Universitario Fundación Favalaro. Argentina.

Maritza Belén Sandoval Rincón

Médico psiquiatra. Especialista en trastornos del sueño. Clínica de Trastornos del Movimiento y Sueño. Hospital General Ajusco Medio. Ciudad de México.

Maurício Scussel

Ingeniero de Control y Automación por la Pontificia Universidad Católica de Paraná (PUCPR), Doctorado en Neurología/ Neurociencias por la Universidad Federal de São Paulo y Centro Nacional de Microelectrónica de Barcelona (UNIFESP/IMB-CNM), Post Doctorado en el Instituto de Fisiología y Fisiopatología en la Universidad de Heidelberg – Alemania, Director de Simplifica - Investigación & Desarrollo para Life Sciences. Brasil.

Margarita Gómez Chantraine

Médico Cirujano - Universidad Nacional de Colombia. Especialista en Obstetricia y Ginecología - Universidad Nacional de Colombia. Magister (C) en Epidemiología Clínica e Investigación - Pontificia Universidad Javeriana. Grouper Manager de Medical Affairs Johnson & Johnson Medical Devices Latinoamérica. Colombia.

Maximiliano Gimenez Gigón

Ortopedía. Cirujano de Columna no Centro de Patología Raquidea - Tucumán, Argentina. AOSpine Fellow en Santa Casa de la Misericórdia - São Paulo, Brasil.

Miguel Farfán

Médico cirujano graduado de la Universidad Nacional de Colombia, Especialista en Epidemiología Clínica de la Universidad el Bosque, Ortopedista y traumatólogo egresado de la Fundación Santa Fé de Bogotá, Universidad del Rosario. Gerente General y Fundador Media For Health SAS, Gerente General y Fundador Unidad Ortopédica de Colombia UNOC.co. Experiencia en innovación en salud, optimización de procesos mediante el uso de las TIC (Apps) y mejoramiento de la experiencia del usuario y satisfacción de pacientes y personal de la salud. Experiencia en investigación clínica, creación de registros clínicos y revisión sistemática de la literatura. Desarrollo de proyectos de innovación educativa en salud para médicos y pacientes.

Mohsen Sadegi Naini

Neurocirujano. Spina Trauma and Sugery Research Center, Tehran University of Medical sciences.

Nelson Astur Neto

Médico Asociado del Grupo de Columna del Departamento de Ortopedia y Traumatología de la Santa Casa de São Paulo - Pavilhão “Fernandinho Simonsen” Investigador de AOSpine Brasil. Cirujano ortopédico de columna en el Hospital Albert Einstein - São Paulo, Brasil. Miembro activo de la Sociedad Brasileña de Columna (SBC), Sociedad Brasileña de Ortopedia y Traumatología (SBOT), North American Spine Society (NASS), American Academy of Orthopedic Surgeons (AAOS), AOSpine. Brasil.

Pedro Luis Bazán

Especialista en Ortopedia y Traumatología, certificado en Cirugía de Columna Vertebral (SAPCV). Jefe de Unidad de Columna HIGA San Martín y Hospital Italiano La Plata. Director de Curso de Capacitación Post Básica en Patología de la Columna Vertebral del Adulto (Departamento de Post Grado Facultad de Ciencias Médicas de la Universidad Nacional de La Plata). Ayudante Diplomado Rentado de Ortopedia y Traumatología (FCM-UNLP). Argentina.

Rodrigo Góes Medéa de Mendonça

Cirujano Ortopédico de Columna. Médico Asociado del Grupo de Columna de la Santa Casa de São Paulo. Médico en el Hospital Albert Einstein - São Paulo. Miembro activo de la Sociedad Brasileña de Ortopedia y Traumatología (SBOT), Sociedad Brasileña de Columna (SBC), Sociedad Brasileña de Ortopedia Pediátrica (SBOP), North American Spine Society (NASS), AOSpine. Brasil.

Vafa Rhimi Movarghar

Profesor de Neurocirugía. Spine Trauma and Surgery Research Center, Tehran University of Medical Sciences.

Víctor Fernando Luján Celis

Médico pasante de servicio social en investigación. División de Investigación en Salud. UMAE “Dr. Victorio de la Fuente Narváez” IMSS. Ciudad de México.

PRÓLOGO

El actual enfoque de AOSpine está centrado en ir más allá de una diseminadora de conocimiento de clase mundial, una Creadora de Conocimiento. Este libro refleja con precisión el objetivo de la *International Research Commission de AOSpine* de convertir a cada uno de sus miembros en 'Cirujanos Científicos'. La Cirugía de Columna se encuentra en rápida evolución y cada día nos enfrentamos a nuevas técnicas, implantes modernos y tecnología de alto costo. Se ha vuelto imperativo que cada cirujano cuente con conocimiento profundo de investigación clínica, lo que le permitirá no solamente realizar su trabajo individual de investigación, sino evaluar la calidad de los trabajos de investigación publicados.

Sin embargo, la transformación de un cirujano talentoso en un cirujano científico requiere conocimiento claro de los protocolos de investigación, lo que este libro brinda con absoluta adecuación. Los capítulos están diseñados para guiar al lector paso a paso por la planificación de su estudio, pasando por la recolección de muestras, selección de pruebas estadísticas apropiadas, análisis de datos y presentación de resultados. Todos los autores son expertos en el área y cada capítulo está escrito con claridad de manera que el cirujano pueda elegir fácilmente el mejor protocolo.

Me encuentro particularmente feliz por los cirujanos de columna en formación, fellows y residentes activamente involucrados en investigación, ya que este libro ofrece una clara comprensión de la metodología de investigación.

Felicito a los autores, quienes pueden sentirse muy orgullosos de su trabajo, una compilación impresionante de información valiosa para todo aquel que se desempeña en la investigación clínica

Prof. S. Rajasekaran, PhD.
*Chairman, Dept of Orthopaedics,
Ganga Hospital, Coimbatore, India
Chair, AOSpine International Research Commission*

Unidad 1

CAPÍTULO 1

Introducción

Asdrubal Falavigna e José María Jiménez Avila

CAPÍTULO 2

Matemáticas: lenguaje e interpretación

Maurício Scussel

CAPÍTULO 3

El actor principal. La variable

José María Jiménez Avila

CAPÍTULO 4

¿Cómo planear la elaboración de una base de datos?

Pedro Luis Bazán

CAPÍTULO 5

Determinando el tamaño de la muestra

Emiliano Vialle

CAPÍTULO 6

¿Cómo elegir las pruebas estadísticas?

Pedro Luis Bazán

CAPÍTULO 7

Bioestadística descriptiva

Alberto Aceves Perez

CAPÍTULO 8

Bioestadística analítica ou inferencial

José María Jiménez Avila

CAPÍTULO 9

Medidas de tendencia central

Alberto Aceves Pérez

CAPÍTULO 10

Medidas de dispersión (Rango, Varianza, Desviación Estándar, Coeficiente de Varianza)

Delio Martins

CAPÍTULO 11

Presentación de los datos*José María Jiménez Avila*

CAPÍTULO 12

Prueba de hipótesis (IC) y valor de “p”*José María Jiménez Avila*

CAPÍTULO 13

Prueba de Kolmogorov-Smirnov*Nelson Astur Neto**Maximiliano Gimenez Gigón**Rodrigo Góes Medéa de Mendonça*

CAPÍTULO 14

Prueba de Shapiro-Wilk*Nelson Astur Neto**Rodrigo Góes Medéa de Mendonça*

CAPÍTULO 1

INTRODUCCIÓN

*Asdrubal Falavigna
José María Jiménez Avila*

El proceso de la investigación, es clave en la generación, paliación y renovación de herramientas diagnósticas, preventivas y de tratamiento oportuno para los pacientes.

Un conflicto paradójico, es que la investigación en estos campos se genera en países desarrollados que distan mucho de tener los problemas de salud del tercer mundo, como también la vivencia de una realidad económica y de desigualdades en salud muy diferentes a las que se tienen a diario.

Esta es la razón de la necesidad de generar investigación en un nivel práctico secundario, no orientada tanto a las ciencias básicas, sino tendientes a los campos de la epidemiología clínica, las ciencias sociales, la economía en salud y la satisfacción del usuario.

Como parte del proceso de la investigación, se encuentra el análisis de los datos y algo más importante, la interpretación de la información que genera la suma de todos los detalles que arroja, al recabar la información de las encuestas.

Para ello debemos de conocer fundamentos que estan asociados a la estadística, la cual es una disciplina de las matemáticas aplicadas que se dedica al manejo de los datos numéricos esto es, es el método para recolectar, elaborar, analizar e interpretar datos numéricos.

A la estadística no se le considera como una ciencia, sino un método que enseña los procedimientos lógicos de la observación y del análisis, para lo cual es necesario tener en cuenta, para aprovechar al máximo las experiencias de otras áreas del conocimiento.

Este libro se presenta, pensando en todo cirujano de columna, estudiante de medicina o médico en general, que tiene amplia experiencia clínica quirúrgica y que cuenta con basta información en su computadora, pero que no ha llegado a dar ese gran paso de

interpretar los datos, no únicamente desde el punto de vista descriptivo, sino desde el análisis o la “inferencia”, lo que permitirá obtener información que será de gran utilidad para la comunidad científica.

El libro consta de 28 capítulos, divididos en 4 unidades, que van desde cómo se estructuran las bases de datos, como se interpretan en la cirugía de columna las pruebas paramétricas y no paramétricas y algo muy importante que es como utilizar los datos obtenidos, no únicamente para mostrarlos en un artículo original, sino saber utilizar la información para gestionar y producir un cambio en la atención médica quirúrgica.

Este libro está pensado en aquellas personas que desean seguir realizando su actividad asistencial y quirúrgica, que les gusta la investigación y tienen la necesidad de saber interpretar las pruebas estadísticas que aparecen en el apartado de material y métodos y resultados, en los artículos originales y que muchas veces no se analizan, tomando los datos como una verdad absoluta, sin darnos cuenta de que, en muchos casos en sus conclusiones, estas carecen de “Congruencia Metodológica”.

Este libro fortalece el Análisis Crítico de la Literatura Médica, para integrar el conocimiento apegado a los lineamientos científicos.

Recuerde: “Lo que no se mide, no se evalúa y por lo tanto no se mejora”.

CAPÍTULO 2

MATEMÁTICAS: LENGUAJE E INTERPRETACIÓN*Maurício Scussel*

Las Matemáticas es la ciencia del razonamiento lógico y abstracto, que investiga geometrías, proporciones, variaciones, entre otros temas. Es una ciencia comprensiva e inherente a las otras ciencias que varias veces, impresiona por aparentes niveles elevados de complejidad.

Muchos ven en esa supuesta complejidad, un obstáculo que dificulta la comprensión de los fenómenos cuya representación se da en lenguaje matemático. Sin embargo, interrogamos: ¿Estaría la dificultad de interpretación de un determinado problema numérico en el lenguaje simbólico de las matemáticas aplicadas, o sería esta derivada del no entendimiento de la razón básica por detrás de los símbolos y notaciones matemáticas?

Muchas veces, la evaluación estadística de un estudio se restringe a los estadísticos responsables del análisis de resultados. No obstante, se puede estimular la reflexión sobre el lenguaje y la interpretación de las matemáticas en aquel investigador que pretende describir la metodología o el resultado de un estudio. De esta forma, la intención de este libro no es transformar al lector en un estadístico, sino despertar su interés por la razón que constituye las bases de este conocimiento.

Desde la Antigüedad, la matemática siempre estuvo presente en el pensamiento de filósofos como Platón, Pitágoras, Aristóteles, entre otros. Esto es comprensible, una vez que la matemática es una herramienta que hace posible demostrar, a través de la lógica, la razón de la mecánica de algún conocimiento y, al mismo tiempo, vincular lo real a lo abstracto. Es paradójico que una ciencia tan exacta y lógica, que define que dos más dos son cuatro, también vincule lo real a lo imaginario, donde $i^2 = -1$.

¿Cómo es posible comprender y dominar este lenguaje con toda esa aparente complejidad?

La matemática está compuesta por un simbolismo extenso, para describir diferentes funciones, ecuaciones, variaciones... Sin embargo, podemos decir resumidamente que esos símbolos, en esencia, sirven para transmitir algún conocimiento que puede estar, al mismo tiempo, en un nivel objetivo o subjetivo. Y, para ilustrar esos aspectos objetivo y subjetivo de los símbolos, reflexionemos sobre los símbolos a continuación:



En un análisis simple y objetivo, nos damos cuenta de que cada elemento está relacionado a algo concreto. La “cruz”, al cristianismo. Lo “correcto”, a una marca de material deportivo. Los “anillos”, a las olimpiadas. Pero, en cada elemento presentado existe un conocimiento de orden subjetivo, enmascarado, que se asocia a la figura. Por ejemplo, para la gran mayoría de los lectores, probablemente no será necesario mencionar el nombre de la marca deportiva que representa la figura, ni mencionar el tamaño de la influencia que esta marca ejerce en el mundo deportivo, ya que eso está almacenado en su inconsciente.

Para crear una imagen sobre lo objetivo y lo subjetivo, podríamos decir que un símbolo es una especie de tarjeta de memoria. Su diseño sería como la unidad física, objetiva (objeto), y su contenido es lo subjetivo (“poner sub”), lo que está en otro estrato.

Pero, ¿por qué lo subjetivo ayuda a entender las matemáticas?

Lograr acercarse más al aspecto subjetivo de los símbolos utilizados en las matemáticas facilita entender la razón por la cual un elemento se aplica a una simple cuenta de suma o de resta. Así, recorreremos algunas operaciones aritméticas, con la mente abierta, para tratar de observar esta propiedad de los símbolos desde otro ángulo.

La aritmética es el ramo de las matemáticas dedicada a los números y a las operaciones posibles entre ellos. Segmento más antiguo y más elemental que comprende, entre varios elementos, las cuatro operaciones: suma, resta, división y multiplicación. Para buscar una aproximación de la relación de lo “objetivo x subjetivo”, analizaremos cada símbolo de las cuatro operaciones, buscando extraer una “razón” por la cual el símbolo presenta determinada forma:

- **Resta:** operación representada por una línea continua. Según historiadores, este símbolo surge en la literatura occidental a partir del siglo XIV y XV, con la metamorfosis de la palabra latina *minus* a *m̄us* hasta llegar al $-$.

Analizando la función de una resta, por las matemáticas o por la lingüística, llegamos a la idea de que restar representa el hecho de sacarle algo a algo. Así, teniendo en mente la necesidad de representar una separación, una ruptura, ¿cuál sería un elemento simple que representaría esta acción?

Imagínese las siguientes situaciones: Dos vecinos que necesiten delimitar sus propiedades: ¿cuál el elemento geométrico que serviría para representar esa separación de áreas? Cuando es necesario separar la piel para una incisión quirúrgica, ¿cuál el diseño que se forma en la piel? Una línea, ¿verdad?



Claro que esta aproximación abstracta de la razón se debe encarar de una forma lúdica, sin confrontar la explicación lógica de la razón que le ha dado origen al símbolo. Sin embargo, una vez que el conocimiento se incorpora, no importa si su origen fue estrictamente lógico o lúdico, lo importante es entender el concepto del elemento para poder aplicarlo mejor.

- **División:** es la operación que actualmente se representa con una línea continua separando dos puntos (\div). Según historiadores, este símbolo es más moderno y surge por primera vez en el siglo XVII.

De forma análoga a la resta, la división también se refiere a una ruptura entre dos partes. Sin embargo, en la división existe un elemento divisor que delimita en cuantas porciones **iguales** se fraccionará el dividendo.



De esta forma, para representar simbólicamente una ruptura de algo en partes iguales, ¿no sería necesario que el símbolo que representa esta operación contuviese la fragmentación con pesos iguales entre los dos lados?

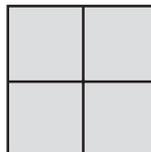
- **Suma:** operación representada por una línea continua superpuesta por otra línea continua. El surgimiento de este símbolo es concomitante al de la resta, apareciendo en las mismas publicaciones. La explicación histórica es que este elemento surge de la palabra “*et*” (que corresponde a la conjunción y en el español: Hansel y Gretel; dos y dos son cuatro, etc.), que con el pasar del tiempo comenzó a ser abreviada por “*t*”, hasta llegar al símbolo que se utiliza actualmente (+).



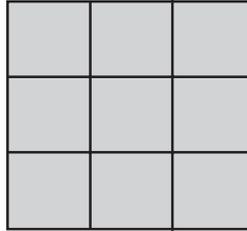
Para alcanzar una imagen lúdica de este elemento, podemos imaginarnos el resultado de alguna actividad que necesite ligar dos puntos separados por una línea o espacio. Como ejemplo, podemos imaginarnos el cordón de un tenis que junta dos partes, o el punto quirúrgico que une dos porciones de piel separadas por un corte.

- **Multiplicación:** para demostrar una razón por la cual el símbolo de la multiplicación presenta la forma (x), se necesita el auxilio de la geometría. Este recurso ayuda a entender que cuando multiplicamos un número por él mismo, lo estamos llevando al cuadrado (potenciación).

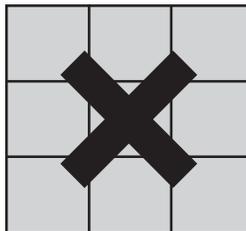
Imaginemos la multiplicación del número 2×2 . También podemos decir que estamos elevando el número 2 al cuadrado, pues $2 \times 2 = 4$, así como $2^2 = 4$, por lo tanto $2 \times 2 = 2^2$. Pero, ¿por qué el uso de la geometría? Observe la imagen abajo:



Podemos decir que el cuadrado dibujado arriba representa la ecuación 2^2 , pues tenemos 4 elementos dentro del cuadrado. O, podemos decir también que representa la multiplicación de los elementos 2 y 2, pues su resultado será 4. En el caso del 3×3 , o, 3^2 , tenemos una situación análoga:



Representamos el 3^2 a través de la multiplicación del 3 por sí mismo, resultando en el número 9, que corresponde a la cantidad de elementos dentro del cuadrado. Observando desde la geometría, vemos que en la multiplicación, los elementos no solamente crecen en largo, sino también en ancho para atender el requisito de la multiplicación y, por consiguiente, de la potenciación. Así, el símbolo que posiblemente representa la expansión hacia las dos dimensiones (largo y ancho) podría ser la X.



Claro que para presentar resultados científicos, la subjetividad del simbolismo de los elementos matemáticos no queda evidente, ya que es necesario fijar conceptos y bases para asegurar el avance del conocimiento, y el objetivismo propicia un caminar más sólido. Por lo tanto, lo que se propone es solamente despertar la curiosidad de la razón de ser de cada elemento matemático. Esa forma lúdica de encarar las matemáticas tal vez no aclare un problema, en primera instancia, pero el mirar hacia cualquier problema desde otro ángulo, ampliamos su comprensión.

Si durante la lectura de este libro enfrenta dificultades para analizar o entender alguna fórmula matemática, recuerde que entender una ecuación es análogo a entender el caso clínico de un paciente. Cada señal, cada símbolo equivale a un síntoma, a un análisis; y cada síntoma posee un historial, una razón. Cuando usted se enfrente a algo complejo, recuerde los estudios de casos clínicos. Comience analizando punto por punto y, a través de la razón de cada elemento la ecuación comenzará a hacerse más comprensible.

¡Buena Lectura!

LECTURA COMPLEMENTARIA

1. Logic, Wesley C. Salmon. Prentice-Hall, 1963.
2. SUPPES, Patric. Introduction to Logic, Princeton, N.J., D Van Nostrand Company, Inc., 1957.

CAPÍTULO 3

EL ACTOR PRINCIPAL. LA VARIABLE

José María Jiménez Avila

En una investigación y desde sus inicios, todo gira alrededor de lo que se ha denominado como “variable” y se puede definir como la cualidad, propiedad o característica de las personas o cosas en estudio, que puede ser enumerado o medido cuantitativamente o cualitativamente y que varía de un sujeto a otro.

Debe de tomarse siempre en cuenta, desde que se elabora la pregunta de investigación (PICOT), como durante todo el proceso metodológico.

La función de la variable, es la de descomponer la hipótesis planteada en sus elementos más simples y es toda aquella característica o evento que de las personas o los grupos de estudio, cambian de una situación a otra o de un tiempo a otro y que por lo tanto pueden tomar diversos valores.

Encontramos en la literatura varios conceptos en relación a la definición de las variables, por lo que se debe de analizar cada uno de los términos.

Las variables se clasifican en independientes o variables explicativas, exposición (**causa**) y las dependientes o respuesta (**efecto o enfermedad**) y existen otras que estarán modificando algún evento, las cuales se denominan como intervinientes.

En el momento que se identifica un problema y se aventura a dar una explicación hipotética de la variable, es necesario someterla a prueba, esto se puede lograr descomponiéndola en variables pequeñas, las cuales deben ser susceptibles de ser evaluadas empíricamente.

Si los procedimientos empíricos no contrarían la hipótesis planteada, ésta se acepta como probablemente verdadera.

En la mayoría de los casos es necesario medir las variables durante la contrastación empírica de la hipótesis y la medición resulta un procedimiento indispensable en la práctica científica.

Una variable debe de tener 2 características indispensables: “Medible y Observable”.

Medición

La medición consiste en asignar un número o una calificación a alguna propiedad específica de un individuo, una población o un evento, usando ciertas reglas, no se mide al individuo, sino cierta característica suya; no obstante la medición es un proceso de abstracción.

Para realizar la medición, es necesario seguir un proceso, el cual consiste en el paso de una entidad teórica a una escala conceptual y posteriormente a una escala operativa.

Los pasos que se siguen son:

- a) Se delimita la parte del evento que se medirá;
- b) Se selecciona la escala con la que se medirá;
- c) Se compara el atributo medido con la escala y;
- d) Finalmente, se emite un juicio de valor acerca de los resultados de la comparación.

Principales escalas de medición (tipos de variables)

Las escalas se clasifican en cualitativas (nominal y ordinal) y cuantitativas (discreta y continua), un requisito indispensable de todas las escalas, es que las categorías deben ser exhaustivas y mutuamente excluyentes.

• Escala nominal

Es aquella cuya característica se define por un nombre y al ser definida por uno, NO implica ser mas o menos que la característica definida por otro nombre diferente, ejemplo: Sexo: masculino, femenino, ocupación, tipo de parto, tabaquismo, etc.

De manera práctica en ocasiones se opta por darles un número a manera de codificación, esto para facilitar la captura, en estos casos el número carece de significado, unicamente diferenciar un grupo del otro, sin tener mayor o menor valor.

- **Escala ordinal**

Es aquella cuya característica puede recibir algún orden subjetivo, se puede asumir que una característica es mas o menos que las otras, pero se desconoce que tanto más o menos. Ejemplo: bien o mal.

El orden es subjetivo o bien a elección del investigador, escolaridad: (primaria, secundaria etc.).

- **Escala Discreta**

Aquella cuyos valores están separados entre si por una cantidad determinada y la unidad no puede ser fraccionada.

Al momento de dividirse la unidad, esta desaparece. Ejemplo: edad (años cumplidos), No. de hijos.

- **Escala Continua**

Aquella en que la escala de medición, se puede dividir en una cantidad infinita de valores entre dos puntos cualquiera.

El metro puede dividirse en un infinito número de partes.

Todos los derivados con la longitud o del tiempo por ejemplo: volumen, área, peso, talla, temperatura.

Proceso de Operacionalización de las Variables

Las variables que se investigan en un estudio quedan identificadas desde el momento en que se define el problema, en el marco teórico se identifican las variables secundarias y se conceptualizan las mismas, este nivel de definición es abstracto y complejo y no permite la observación o medición.

El proceso de llevar una variable de un nivel abstracto a un plano operacional, se denomina Operacionalización y la función básica de dicho proceso, es precisar o concretar al máximo el significado o alcance que se otorga a una variable en estudio.

Los conceptos representan el sistema teórico de cualquier ciencia y son símbolos de los fenómenos que se estudian.

Un concepto es general y abstracto, no siendo posible la observación de los fenómenos involucrados, por ello es necesario llegar a la operacionalización de las variables.

La operacionalización se logra a través de un proceso que transforma una variable en otras que tengan el mismo significado y

que sean susceptibles de medición empírica para lograrlo, las variables principales se descomponen en otras más específicas llamadas dimensiones, a su vez, es necesario traducir estas dimensiones a indicadores para permitir la observación directa.

Ejemplo:

Variable: Evolución de la fractura.

Indicador: Estado de consolidación.

Dimensión: Fases de la consolidación.

Escalas: 1. Callo blando.

2. Callo duro.

3. Corticalización.

4. Remodelación.

Una vez que este proceso termina, podemos contar que las variables obtenidas de manera conceptualizadas, pueden ser medidas y pueden ser observadas.

Al mismo tiempo, esto permitirá establecer alguna asociación que permita identificar factores causales que pueden ser predictores en la identificación de factores de riesgo, mismos que pueden desencadenar algún tipo de enfermedad o de resultado.

Las variables deben de definirse de manera clara con el objeto de evitar confusiones, esto permitirá facilitar la búsqueda y el análisis posterior de los datos, lo cual garantizará la comparabilidad de los resultados con los que fueron obtenidos con algunos estudios hechos de manera previa.

LECTURA COMPLEMENTARIA

1. Celis de la Rosa A. Bioestadística, en: Características de los datos y de su recolección, revisión y computo. 1ª ed en español. Manual Moderno. 2004.
2. MacMahon B, Pugh TF. Epidemiology. Principles and methods. Boston: Little Brown & Co., 1970.
3. Moreno-Altamirano A. Salud Pública. Vol. 42. No 4 Julio/Agosto 2000 pp 337-348.

CAPÍTULO 4

¿CÓMO PLANEAR LA ELABORACIÓN DE UNA BASE DE DATOS?

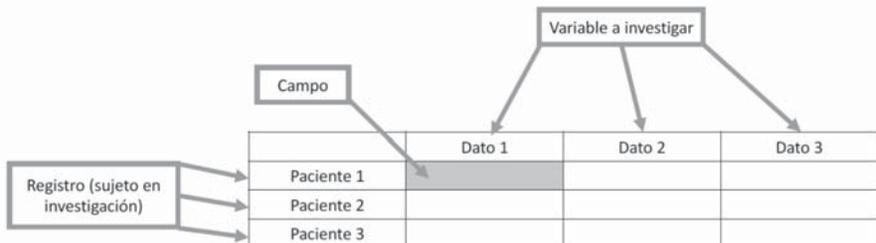
Pedro Luis Bazán

Introducción

El correcto manejo de la información es fundamental para cualquier tipo de actividad laboral del ser humano. La actividad médica científica no escapa a esta regla y es por eso que debemos conocer y planificar la mejor manera posible cómo vamos a recolectar, almacenar y analizar la información para resolver la pregunta que origino nuestra investigación.

Una manera sencilla es crear archivos en planillas de cálculo (tablas) que contengan en cada archivo un tópico. La unidad funcional de estas tablas son las celdas de la hoja, también llamados campos. En la primera celda de cada columna se pondrá el título de la variable a evaluar, que será el mismo para cada paciente ingresado. Cada fila representa un registro, que representa la información ingresada de cada paciente (Fig. 1). Se puede crear otra hoja que recabe otra información u otro documento.

Figura 1 – Ejemplo de componentes de una planilla de cálculo



Las planillas de cálculos actuales permiten realizar distintas funciones para el análisis estadístico de los datos. Manejar la información de esta forma, puede mostrar algunas dificultades, relacionadas principalmente a la complejidad del trabajo científico planificado y en la necesidad de contar con varias hojas, de esta manera podemos encontrar información redundante, inconsistente y no utilizable; siendo difícil manejar el error de ingreso de la información, uno de los errores más frecuentes en la investigación científica y la pérdida de datos.

Otras dificultades en la utilización de archivos en el manejo de los datos, consisten en la complejidad de: modificar dinámicamente la estructura de los archivos, consultar, actualizar y proteger los datos.

Otra forma de manejar la información es la creación de bases de datos, que facilite la gestión de los mismos. La Real Academia Española, define la base de datos como “un conjunto de datos organizados de tal modo que permita obtener con rapidez diversos tipos de información”.

Para su creación y manejo podemos utilizar diversos gestores de base de datos (Software) que, diferenciándose en su complejidad, todos intentan guardar en forma estructurada los datos para disminuir el error de ingreso y facilitar el análisis de la información, ejemplos de estos gestores son Access de Microsoft, EpiInfo, IBM DB2, etc.

Estos gestores de base de datos se caracterizan por: la independencia de los datos sin tener relación con las aplicaciones que los usen o donde se almacenen; los usuarios de estas bases de datos pueden acceder mediante consultas; los datos se guardan en forma centralizada e independiente tornándolos consistentes, completos, fiables y seguros.

Las bases de datos se almacenan en archivos, que contienen datos (números y caracteres) que pueden abrirse, guardarse, combinarse, editarse, agregar o quitar.

La estructura general de los archivos que contienen bases de datos incluye tablas que pueden relacionarse entre sí.

Podemos clasificar a las bases de datos en: **primarias**, son las diseñadas de acuerdo con las necesidades específicas de información (estudio, registro, generador de indicadores, etc.), y por tanto, con un diseño correcto se ajustan a la respuesta a nuestro interrogante inicial; **secundarias**, es una base de datos utilizada para un fin distinto del que motivó la recolección de los datos que contiene, la rapidez y economía de su uso son sus principales ventajas, pero los datos pueden ser de

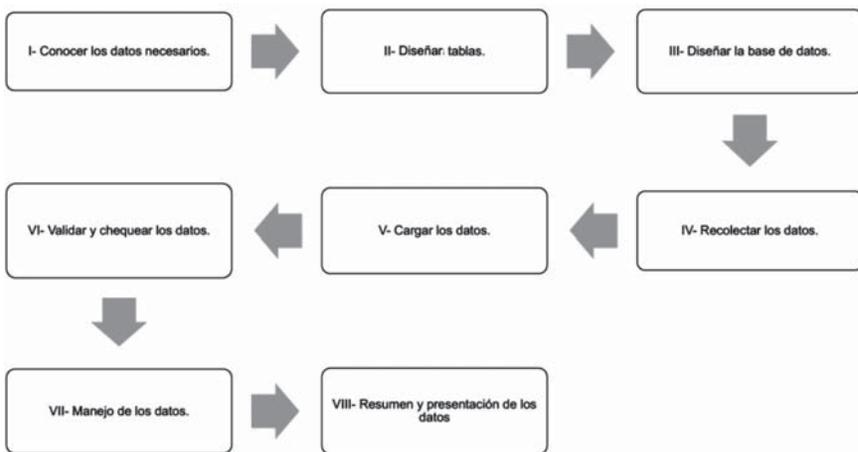
baja calidad, requiriendo muchas veces adaptar los objetivos a los datos; **terciarias**, son las creadas con datos provenientes de distintas bases de datos combinadas de datos reportados previamente.

Planificación de una base de datos

El origen de toda investigación científica es conocer el motivo (pregunta) que estimulo la misma. Una vez cumplido este requisito, el investigador debe planificar todo lo concerniente a trabajo científico, sustento bibliográfico, tipo de estudio, criterios de inclusión y exclusión, los datos necesarios para resolver el tema a investigar y poner plazos para todos estos momentos.

Con respecto a lo concerniente a la base de datos propiamente dicha, también se deben complementar con una serie de pasos para el correcto manejo de información y en cada uno de ellos debemos conocer al responsable del mismo. (Figura 2).

Figura 2 – Diagrama de flujo en la planificación y utilización una base de datos



Conocer los datos necesarios

Antes de comenzar a recabar datos, es imprescindible determinar qué datos son importantes para la investigación, esto permitirá recabar la totalidad de la información y solo lo necesario.

No cumplir con esta premisa puede causar que la información no sea suficiente para responder a los objetivos trazados. En el análisis de los resultados podemos darnos cuenta que esto sucede y estaríamos obligados a modificar el material evaluado o el método de analizarlos, habiendo perdido tiempo en vano o determinando el abandono de la investigación.

Tampoco es razonable tomar datos innecesarios, que no sirven para contestar los objetivos de la investigación y solo consumen tiempo y recursos de más.

Por estos y otros motivos se puede concluir que, como en toda actividad, uno de los pasos más importantes es la planificación sobre qué vamos a investigar, qué necesitamos para hacerlo, cómo lo haremos, con qué o quienes lo haremos, etc; en pocas palabras, de la correcta planificación depende en parte el éxito de la investigación.

Diseñar las tablas

Como expresamos con anterioridad, una base de datos es un conjunto de tablas interrelacionadas entre sí, para facilitar el manejo de los datos.

Al conocer cuáles son los datos que necesitaremos para nuestra investigación, pasaremos a decidir la estructura de nuestra base de datos. Dependiendo de la complejidad de los datos a cargar se determina la complejidad de la base de datos a crear.

Los distintos programas gestores de bases de datos nos brindan la posibilidad de editar a nuestra necesidad la forma de ingresar la información y el modo que se interrelacionan las distintas tablas.

En cada campo seleccionado (Figura 3), se le otorga un nombre a esta variable y le dan distintas características para los datos a cargar, que nos permitirán disminuir el error de carga y que todos los elementos compartan las mismas características y así permitir el correcto análisis.

Figura 3 – Diseño de la tabla. Observese la selección del tipo de dato para cada variable en el cuadro superior y las validaciones que se realizan en cada campo, en el inferior



Existen distintos tipos de **variables**:

- **Texto:** Contienen datos de texto y números, pero con estos últimos no se pueden realizar operaciones matemáticas. Son útiles para guardar información como nombres y direcciones.
- **Número:** Se las usa para almacenar información numérica. Pueden contener datos continuos o categóricos, números enteros o decimales. Esto debe especificarse al armar la estructura. Se pueden realizar operaciones matemáticas con ellos.
- **Lógicas; Si/No; Verdadero/Falso:** Se almacenan datos que tienen sólo dos posibilidades de respuesta, y se los llama también binarios.
- **Fecha/Hora:** Se almacenan fechas. Las mismas pueden ser cortas (año con las últimas dos cifras), largas (año de cuatro cifras), asociada o no con la hora. Con ellos se pueden realizar cálculos entre fechas y la mayoría de los programas permiten cargar solo fechas válidas.

Cada uno de estos campos corresponderá a la variable a estudiar, y se representa en la vista de hoja de datos como columnas, las filas corresponden a cada sujeto y cada casilla en particular representa el valor específico de la variable a estudiar en cada sujeto (Figura 4).

Figura 4 – Vista en hoja de datos de la tabla diseñada en la figura anterior

HISTORIA CL.	APELLIDO Y	SEXO	EDAD	OCUPACIÓN	MAREOS	VÉRTIGO	DERIVACIÓN	PTE	CM	ORL
539379		F	44	EMPLEADA DOP	<input type="checkbox"/>					
13027871		F	56	AUXILIAR EN IA	<input type="checkbox"/>					
24151952		F	38		<input type="checkbox"/>					
27512676		M	33	Administrativo	<input type="checkbox"/>					
27742034		F	34	ADC	<input type="checkbox"/>					
92707115		F	29	ESTUDIANTE	<input type="checkbox"/>					

Diseñar la base de datos

Al conocer las distintas tablas que podemos requerir, determinaremos como será nuestra base de datos.

Se puede desglosar este paso en tres etapas:

1. Diseño conceptual: en esta etapa planifica la estructura de la base de datos sin basarse en la tecnología ni en el software a utilizar. Se realiza un esbozo de lo que necesitamos para trabajar, basándonos en los datos que necesitamos evaluar. En este punto decidiremos el tipo de base de datos a utilizar (planas, relacional, centrada en objetos, jerárquicas, redes, relaciones operacionales, etc.).

2. Diseño lógico: es el momento de llevar la idea a la realidad basándonos en la tecnología a utilizar, es decir con que gestor de base de datos trabajaremos y cuáles son las claves y variables que jerarquizamos.

3. Diseño físico: se trata de relacionar la estructura de la base de datos con el hardware necesario para conseguir una mayor eficacia en la carga, almacenamiento y consulta de los datos, por ejemplo,

tener en cuenta la capacidad de memoria interna o externa para el almacenamiento de los datos a cargar, copias de seguridad, etc. Principalmente si además de los datos se almacenaran imágenes.

Es útil al finalizar esta etapa, realizar varias pruebas de ingreso, consulta y modificación de datos virtuales para evaluar el rendimiento de la base de datos y reconocer fallas de programación antes de comenzar la carga de los datos.

Recolectar los datos

Antes de comenzar el ingreso de pacientes, debemos realizar una planilla que servirá para la recolección de la información requerida y deben ser recolectada en el mismo orden que serán cargados en la base de datos.

Cada planilla debe ser rotulada con la misma clave del registro de la base de datos, para poder acceder a ella en forma rápida si existiera un error de entrada o faltase algún dato.

Carga de datos

La información puede ser cargada en nuestra base de datos de dos maneras posibles, tomando como base la tabla ya cargada. Una de ellas es viendo la misma tabla como hoja de datos que creamos al comienzo (Figura 4). La dificultad es que al cargar los datos el cursor debe mantener la misma fila (paciente) y moverse de columna (variable a evaluar) para no cometer el error de cargar un dato en otro registro.

Otra manera de registrar la información en una base de datos, es creando un formulario tomando como base el diseño de la tabla (Figura 5), de esta manera vemos en pantalla solo los campos a llenar con datos de un solo registro (paciente) y aceptar el dato cargado se pasa al siguiente campo del mismo paciente. Al finalizar la carga se pasa a otro registro.

Figura 5 – Formulario de entrada de datos con los campos a llenar

The screenshot shows a web-based data entry form titled "COLUMNA CERVICAL Y MAREOS". The form is displayed in a browser window with the address bar showing a file path. The form includes the following fields and options:

- HISTORIA CLINICA: [Text input field]
- APELLIDO Y NOMBRE: [Text input field]
- SEXO: EDAD: [Text input field] OCUPACIÓN: [Text input field]
- CONSULTA: MAREOS: VÉRTIGO:
- DERIVACIÓN: PTE: CM: ORL: NEURO: OTROS: [Text input field]
- TEC: FECHA: [Text input field] TRAUMA CERVICAL: FECHA: [Text input field]
- CERVICALGIA: BRAQUIALGIA: CEFALEA: ACUFENC: NISTAGMUS:
- ROMBERG: ROMBERG SENSIBILIZADA: GIRO TOTAL: GIRO CABEZA FIJA:
- Rx: [Text input field]
- RMN: [Text input field]

The form is part of a database application, as indicated by the browser's address bar and the "Herramientas de base de datos" menu item. The status bar at the bottom shows "Registro: 4 de 7" and "Sin filtro".

Como existen opciones de validación automática del formato de carga o de la obligatoriedad de la misma el programa nos avisa la dificultad para llevar a cabo el ingreso y de esta manera disminuimos el error de ingreso.

Validar y checar los datos

Este paso es de fundamental importancia para el correcto análisis de la información y la obtención de los resultados del trabajo.

La validación de los datos se debe realizar metodológicamente, en cada traspaso de información (recolección, carga y almacenamiento), de no realizarlo nos daremos cuenta de datos perdidos o inconsistentes en una etapa final de la investigación haciendo fracasar la misma o haciendo invertir tiempo extra en el control del error.

Cuando se investiga en grupo, lo recomendable es que la carga de los datos tenga un solo responsable y que sea quien valide y chequee el proceso de carga.

Manejo de los datos

Un punto crucial en el manejo de la información, es el análisis de los datos y será tema de otro capítulo.

Cuando los datos han sido cargados, se puede ingresar a la base de datos a realizar consultas e informes, donde se decide que operación estadística se utilizará.

Resumen y presentación de los datos

Los sistemas gestores de base de datos son capaces, luego de analizar los mismos entregar un informe resumido, tablas de doble entrada y gráficos que ilustran los resultados.

Responsables

Para evitar errores en la gestión de los datos (toma, carga, almacenamiento y análisis) cada paso debe tener un responsable conocido.

Todos los autores de la investigación son los responsables de la toma de decisión tanto de la planificación de la investigación como de reconocer los datos necesarios para la misma.

La planificación de la base de datos y de la planilla de recolección de la información debe tener uno o dos responsables; si uno de ellos es un técnico en estadística es mejor.

Para la recolección de los datos y análisis de los resultados, todos los participantes deben estar involucrados.

En la carga de datos se debe contar con solo un responsable, que al mismo tiempo sea en evaluador que todos los datos necesarios se hayan recolectado.

Un técnico en estadística o uno de los autores familiarizados con el análisis de los datos debe ser el responsable de la selección y análisis de las pruebas estadísticas. (Figura 6).

Figura 6 – Informe de análisis de datos realizados por EpiInfo, con tabla de doble entrada y resultados

Epi Info

Librería de resultados

Vista actual: C:\Users\Pedro\Documents\Trabajos\HIGA SM\En curso\Modic\PACIENTES.mdb:PTES

Número de registros: 100

Fecha: 09/08/2010 11:06:02 p.m.

TABLES DISCO_1 INESTABILIDAD

Siguiente procedimiento

INESTABILIDAD				Adelante	
DISCO_1	0	1	TOTAL		
0	48	17	65	?	?
% Fila	73.8	26.2	100.0		
% Columna	63.2	70.8	65.0		
1	28	7	35	?	?
% Fila	80.0	20.0	100.0		
% Columna	36.8	29.2	35.0		
TOTAL	76	24	100		
% Fila	76.0	24.0	100.0		
% Columna	100.0	100.0	100.0		

Análisis de tabla simple

	Point Estimación	95% Intervalo de Confianza	
		L. Inferior	L. Superior
PARAMETROS: Basados en OR			
Odds Ratio (producto cruzado)	0.7059	0.2607	1.9111 (T)
Odds Ratio (EMV-MLE)	0.7083	0.2465	1.9004 (M)
		0.2201	2.0773 (F)
PARAMETROS: Basados en el riesgo			
Razón de Riesgos (RR)	0.9231	0.7408	1.1502 (T)
Diferencia de Riesgos (DR)	-6.1538	-23.1763	10.8686 (T)

(T-Series Taylor;C-Cornfield;M-P-Media;F-Fisher)

TEST ESTADÍSTICOS	Chi cuadrado	P de 1 cola	p de 2 colas
Chi cuadrado: sin corregir	0.4723		0.4919156556
Chi cuadrado: Mantel-Haenszel	0.4676		0.4940887145
Chi cuadrado: Corrección de Yates	0.1952		0.6586242660
P-media exacta			0.2548047339
Test exacto de Fisher			0.3335295358

Utilidades de una base de datos

Si la planificación realizada ha sido exitosa, la base de datos creada no solo nos permite ingresar datos por medio de un formulario, que tiene la utilidad del manejo en la pantalla.

También es posible realizar consultas por diversos parámetros que nos permite evaluaciones parciales cada vez que se cargan los datos, relacionando más de una tabla del total.

Se puede realizar informes de cada registro en forma individual, como así también de las consultas. La diferencia entre los informes y los formularios en la visualización de cada registro es que la visualización del formulario es solamente en la pantalla, en cambio el informe es la forma de visualización impresa que puede servir.

Conclusiones

La gestión de los datos, incluye la carga, almacenamiento, consulta y análisis de los datos es muy importante para la investigación médica.

Las bases de datos son herramientas imprescindibles para gestionar la información de la investigación; su planificación requiere el análisis y conocimiento de los datos necesarios para llevar a cabo nuestro trabajo, y debemos basarnos en ellos.

Antes de crear una base de datos, debería hacerse las siguientes preguntas:

- ¿Qué datos deseo almacenar y cuál es la mejor forma de hacerlo? (Tablas y sus relaciones).
- ¿Cómo introduzco los datos relativos a la investigación? (Formulario).
- ¿Cómo realizo evaluaciones parciales de los datos? (Consultas).
- ¿Cómo evalúo los datos? (Operaciones estadísticas).

LECTURA COMPLEMENTARIA

1. Bleisch, S: Geographic Information Technology Training Alliance (GITTA) presents: Databases.
2. Deserno, TM; Haak, D; Brandenburg, V; Deserno, V; Classen, C; Specht, P: Integrated Image Data and Medical Record Management for Rare Disease Registries. A General Framework and its Instantiation to the German Calciphylaxis Registry. *J Digit Imaging* (2014) 27:702–713 DOI 10.1007/s10278-014-9698-8.
3. Gómez Ballester, E; Martínez Barco, P; Moreda Pozo, P; Suárez Cueto, A; Montoyo Guijarro, A; Saquete Boro, E: Bases de Datos I. Introducción a las bases de datos. fundamento de diseño de bases de datos. Departamento de Lenguajes y Sistemas Informáticos Escuela Politécnica Superior. Universidad de Alicante.
4. Haak, D; Samsl, C; Gehlen, J; Jonas, S; Deserno, TM: simplifying electronic data capture in clinical trials: workflow embedded image and biosignal file integration and analysis via web services. *J Digit Imaging* (2014) 27:571–580. DOI 10.1007/s10278-014-9694-z.
5. López Martínez A, C; Sánchez Brea, LM; Gómez, O; González Del Campo Rodríguez Barbero, MAV. Editorial McGraw-Hill/Interamericana de España. 2002.
6. Marcus, DS; Harms, MP; Snyder, AZ; Jenkinson, M; Wilson, JA; Glasser, MF; et al. Human Connectome Project Informatics: quality control, database services, and data visualization *Neuroimage*. 2013; October 15; 80. .doi:10.1016/j. neuroimage. 2013.05.077.
7. Musick, BS; Robb, SL; Burns, DS; Stegenga, K; McCorkle, KJ: Development and Use of a Web-based Data Management System for a Randomized Clinical Trial of Adolescents and Young Adults. *Comput Inform Nurs*. 2011 June; 29(6): 337–343. doi:10.1097/NCN.0b013e3181fcbc95.
8. Quintana Rondón, Y; Dominguez, LC; Diaz Berenguer, A: Diseño de la Base de Datos para Sistemas de Digitalización y Gestión de Medias. *Revista de Informática Educativa y Medios Audiovisuales*. Vol 8 (15). 17-25. 2011.
9. Robbins, RJ: Database Fundamentals. Johns Hopkins University.
10. Sanchez, J: Principios sobre Bases de Datos Relacionales.
11. Silberschatz, A; Korth, HF; Sudarshan, S; Fundamentos de bases de datos. Cuarta edición. Traducido de la cuarta edición en inglés por Sáenz Pérez, F; García Cordero.

CAPÍTULO 5

DETERMINANDO EL TAMAÑO DE LA MUESTRA

Emiliano Vialle

Introducción

Definir el número de participantes de una investigación siempre ha sido un tema crucial y de difícil solución y, a menudo, descubrimos que el número de participantes era insuficiente tras el término del proyecto. Parece obvio que un mayor número de sujetos incluidos en una investigación producirá resultados más fiables y próximos a la realidad. Tiempo, recursos y frecuencia del tema para el que buscamos respuesta hacen con que no se pueda incluir pacientes en un estudio indefinidamente, y la pregunta crucial pasa a ser ¿cuál el número mínimo de pacientes en cada grupo de investigación? Además, el cálculo muestral ha sido solicitado por comités de ética para aprobación de proyectos de investigación.

En este capítulo, se tratará de definir, a través de ejemplos prácticos, las maneras más comunes y prácticas de llegar a ese “número mágico”. Es importante subrayar que las fórmulas para cálculo de la muestra son sensibles a errores, y variaciones en los parámetros pueden llevar a grandes diferencias en el tamaño de la muestra. Por esto, antes de comenzar a buscar el tamaño de nuestra muestra, habiendo elaborado el proyecto de investigación, se deben definir algunos ítems:

1) Error del tipo I (alfa)

En estudios clínicos se evalúa solamente una parte de la población, probando la hipótesis de que hay diferencia entre muestras (tasa de infección en diabéticos, por ejemplo), en el resultado de tratamiento entre grupos (artrodesis de columna en diabéticos versus no diabéticos), o en el resultado de tratamientos en un grupo (artrodesis versus tratamiento conservador en pacientes diabéticos). La hipótesis nula (H_0) parte del principio de que los pacientes (muestras) que se están comparando no son diferentes. La hipótesis alternativa (H_1) considera

que los grupos son diferentes. En ese momento, el cálculo muestral define en qué momento la inclusión de pacientes en un estudio dejará de afectar el resultado.

En este proceso pueden ocurrir dos tipos de error (tipo I y tipo II, Tabla 1)

Tabla 1 – Panorama de errores en investigación clínica

		Población	
		No hay diferencia	
No hay diferencia			Resultado falso negativo Error tipo II (beta)
	Hay diferencia	Resultado falso positivo Error tipo I (alfa)	Poder (1-beta)

El error de tipo 1 (alfa) mide la probabilidad de que, a partir de la hipótesis nula (H_0) de que las muestras vienen de la misma población, las diferencias encontradas en el estudio son verdaderas. Representa la posibilidad de que rechazemos erróneamente la H_0 y elijamos un resultado falso positivo. Alfa generalmente se fija en 0.05, lo que significa que el investigador desea un riesgo inferior al 5% de presentar resultados falsamente positivos.

2) Error del tipo II (beta)

En vez de llegar a una conclusión falsamente positiva, el investigador podrá llegar a un falso negativo, concluyendo que no existe diferencia en un tratamiento entre grupos cuando, en verdad, existe diferencia. En otras palabras, aceptan la H_0 como verdadera, lo que se denomina error del tipo II (beta). Tradicionalmente, beta se define como 0.2, lo que significa que el investigador desea una chance inferior al 20% de llegar a una conclusión falsamente negativa. Para calcular el tamaño de la muestra, se debe determinar la fuerza del estudio. Esta fuerza estadística refleja la capacidad del estudio de encontrar un efecto en la población, basándose en una muestra de la misma (verdadero positivo). La fuerza estadística es el complemento del error tipo II (1-

beta). Por lo tanto, si el error tipo II es de 0.2, la fuerza del estudio será de 0.8, o del 80%. Esta es la probabilidad de rechazar la hipótesis nula correctamente, o evitar una conclusión falsamente negativa. Siendo más objetivo, sería la probabilidad de que al presentar una diferencia en un estudio, esta sea verdadera.

3) Diferencia mínima relevante:

Es la diferencia mínima que el investigador desea identificar entre los grupos estudiados. Debe ser una diferencia clínicamente relevante y biológicamente plausible.

4) Variabilidad:

El cálculo de la muestra se basa en la variación poblacional para la variable estudiada, a través de la desviación estándar. Este valor generalmente se obtiene a través de un estudio piloto, o a partir de estudios previos. La diferencia mínima y la variabilidad pueden, algunas veces, ser representadas como un múltiplo de la desviación estándar de las observaciones. Fórmula:

$$\text{Diferencia estandarizada} = \frac{\text{diferencia entre las medias entre dos grupos de tratamiento}}{\text{desviación estándar de la población}}$$

5) ¿Cuál el tamaño de la población?

Muchas veces este dato es difícil de obtener, principalmente si su proyecto implica una investigación muy específica. Imagínese que su estudio implica una comparación entre tratamiento quirúrgico y conservador de fracturas de la columna tóraco-lumbar. Su población se obtendría a través de algún estudio epidemiológico, que ni siempre se aplica a su realidad, o a través de una revisión de las bases de datos de atención de su hospital.

6) ¿Cuál el margen de error (intervalo de confianza) deseado?

Partiendo del principio de que su muestra no será perfecta, se debe definir cuánto error se aceptará. Generalmente se utiliza el valor de un 5%.

7) ¿Cuál el nivel de confianza deseado?

Los intervalos de confianza más utilizados son los valores del 90%, 95%, y 99%. Este valor determinará la posibilidad de que su resultado esté dentro del intervalo de confianza.

8) ¿Cuál la desviación estándar?

¿Cuánta variación espera para sus resultados? Como esta es una respuesta de difícil solución, la mayoría de las investigaciones utiliza el valor de 0.5.

Teniendo en manos estos datos que acabamos de mencionar, pero sin saber el tamaño de la población en estudio, pasamos al cálculo de la muestra. Su nivel de confianza corresponde a un Z-score, valor constante utilizado en el cálculo de la muestra:

90% - Z-score = 1.645

95% - Z-score = 1.96

99% - Z-score = 2.576

Existen tablas para los valores del Z-score en internet, pero los valores arriba son los más comunes para aplicación de la fórmula:

Tamaño de la muestra = (Z-score)² x DP x (1-DP) / (margen de error)²

Digamos que usted haya elegido un nivel de confianza del 95%, una desviación estándar de .5, y un intervalo de confianza de +/- 5%. Dejemos de lado el tamaño de la muestra, o consideremos que el estudio involucra a una parte significativa de la población.

Tamaño de la muestra = $((1.96)^2 \times .5(.5)) / (.05)^2$

$(3.8416 \times .25) / .0025$

$.9604 / .0025$

384.16

385 participantes serían necesarios.

Si este número no es alentador, es posible reducir el intervalo de confianza o aumentar el margen de error. La fórmula arriba representa

una de las maneras más sencillas de determinar muestras de investigación, y se utiliza a menudo para encuestas de opinión, por ejemplo, donde no es necesaria la comparación entre grupos. La Tabla abajo ilustra cómo la definición de los valores de alfa y beta influye en el tamaño de la muestra.

Tabla 2 – Tamaño de muestra relativa para niveles diferentes de alfa y poder Alfa (error tipo 1) Poder (1-beta)

	Alpha (type I error)		
	0.05	0.01	0.001
Power (1-beta)			
0.80	100	149	218
0.90	134	190	266
0.99	234	306	402

Cuando tratamos de estudios clínicos, donde se investigan diferencias entre grupos de pacientes, se hace necesaria alguna otra información.

Consideremos un estudio comparando tratamiento quirúrgico y conservador para fractura de columna toráco-lumbar, donde los participantes se distribuirán aleatoriamente en dos grupos homogéneos, y el resultado final sería el nivel de dolor, seis meses tras el inicio del tratamiento. A partir de estudios previos, se sabe que las fracturas tratadas conservadoramente (grupo A) presentan mejoría de la incapacidad de 15 puntos, y que las quirúrgicas (grupo B) mejoran 20 puntos.

Estudios previos muestran que la variación de la incapacidad puede ser de 5 puntos (desviación estándar de la muestra).

Para aclarar si el tratamiento quirúrgico lleva a una mejoría significativa en relación al conservador, se puede utilizar la prueba *t* de Student para muestras no pareadas, y para asegurar que el estudio no tendrá un número insuficiente o excesivo de participantes, se hace necesario planificar el tamaño de la muestra. Esto requiere una especificación de la fuerza estadística (probabilidad de identificar una diferencia real con la prueba, generalmente un 80 o 90%) y del nivel de significancia de la prueba estadística (probabilidad de obtener un resultado significativo cuando no hay diferencia real, generalmente del 2.5% para pruebas unicaudales). A partir de estos datos comienzan a surgir fórmulas estadísticas que no son familiares a la mayoría de los médicos, en el caso de la prueba *t* abajo:

$$n \approx \left[\frac{2(z_{\text{Power}} + z_{1-\alpha})}{2(\mu_1 - \mu_2)/\delta} \right]^2$$

En la fórmula arriba, los grupos ($n_1=n_2$) y las desviaciones estándar son iguales. La diferencia promedio entre las dos poblaciones es de $\mu_1-\mu_2$. La fuerza (Z-power) está dada como 0.8 o 80% y el nivel de significación es alfa. N es la suma de los dos grupos (n_1+n_2). Los valores a insertar en las fórmulas se obtienen a partir de tablas disponibles en internet. El trabajo de Noordzij et al (2010) suministra los valores que se utilizan con más frecuencia. .

Evidentemente, esta fórmula no es alentadora, pero claramente no puede ser interpretada únicamente por un estadístico. El motivo de su descripción es justamente informarle al cirujano los datos necesarios para realización del cálculo de tamaño de la muestra, y permitir un trabajo conjunto para elegir la mejor prueba estadística y la muestra más viable de acuerdo a la realidad clínica en que se realizará el estudio.

Para la prueba *t* demostrada arriba, es necesario estimar los valores promedio en los dos grupos comparados, así como las desviaciones estándar.

Para la prueba de Fisher, se debe suministrar una estimativa de proporción de eventos (por ejemplo, una reacción adversa) en los dos grupos.

La prueba de Mann-Whitney requiere una estimativa de la probabilidad de que la variable estudiada será menor en un grupo que en el otro.

Estudios más simples o menos pretensiosos podrán utilizar gráficos para estimar el tamaño de la muestra, partiendo de presuposiciones sobre el tipo de dato y prueba estadística que se debe utilizar, como publicado por Altman (1980).

Ejemplo de estudio experimental

Para determinar el tamaño de la muestra de un estudio acerca del efecto del trasplante celular sobre la degeneración discal en conejos, se utilizó el valor obtenido en estudio previo comparando discos normales y degenerados.

El cálculo del tamaño de la muestra se realizó en el programa WinPEPI (*Programs for Epidemiologists for Windows*) versión 11.43 y se basó en el estudio de Vialle et al (2012). Para un nivel de significación del 5%, poder del 80% y un tamaño de efecto mínimo de 1,2 desviaciones estándar entre los grupos, se obtuvo un total mínimo de 12 animales en cada grupo.

Lamentablemente, muchas etapas de la determinación son subjetivas, como la determinación de la fuerza de la prueba estadística, y otras son externas, como la frecuencia de la enfermedad estudiada, tiempo de reclutamiento, recursos disponibles. Es posible hacer adaptaciones en consecuencia de estas limitaciones, reduciéndose la fuerza de la prueba, o el intervalo de confianza.

A pesar de las dificultades impuestas por las etapas de la determinación del tamaño de la muestra, y de la posibilidad de error, ella es siempre mejor que elegir una muestra de manera arbitraria.

LECTURA COMPLEMENTARIA

1. Florey CD. Sample size for beginners. *BMJ*. 1993;306(6886):1181-1184. doi:10.1136/bmj.306.6886.1181.
2. Gupta KK, Attri JP, Singh A, Kaur H, Kaur G. Basic concepts for sample size calculation: Critical step for any clinical trials! *Saudi J Anaesth*. 2016;10(3):328-331. doi:10.4103/1658-354X.174918.
3. Noordzij M, Tripepi G, Dekker FW, Zoccali C, Tanck MW, Jager KJ. Sample size calculations: Basic principles and common pitfalls. *Nephrol Dial Transplant*. 2010;25(5):1388-1393. doi:10.1093/ndt/gfp732.
4. Röhrig, B. et al., 2010. Sample size calculation in clinical trials: part 13 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt international*, 107(31-32), p.552-6. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2933537&tool=pmcentrez&rendertype=abstract>.
5. Tobergte DR, Curtis S. No Title No Title. *J Chem Inf Model*. 2013;53(9):1689-1699. doi:10.1017/CBO9781107415324.004.

CAPÍTULO 6

¿CÓMO ELEGIR LAS PRUEBAS ESTADÍSTICAS?

Pedro Luis Bazán

Introducción

La estadística es la ciencia que utiliza conjuntos de datos numéricos para obtener a partir de ellos, inferencias basadas en el cálculo de probabilidades; se pueden distinguir dos tipos de inferencias, las que se conocen como pruebas de hipótesis y la que se ocupa a la estimación de intervalos.

La importancia del análisis estadístico se fundamenta en la decisión si se acepta o se rechaza la hipótesis del estudio a realizar.

Se puede dividir en dos grandes áreas: **estadística descriptiva**, es la que se dedica a la descripción, visualización y resumen de datos originados a partir de los fenómenos de estudio; **estadística inferencial**, se dedica a la generación de los modelos, inferencias y predicciones asociadas a los fenómenos en cuestión teniendo en cuenta la aleatoriedad de las observaciones.

De esta forma podemos analizar distintas variables, que se definen como las cualidades, rasgos, atributos, características o propiedades de diferentes valores, magnitudes o intensidades valoradas en un grupo de elementos. De esta forma tenemos las **variables cualitativas** y estas pueden ser: binarias (p.e. si o no; femenino o masculino; verdadero o falso), ordinales (p.e. estadificación de tumores) y nominal (p.e. tipo de fractura toracolumbar); y las **cuantitativas**, que otorgan un valor, pudiendo ser continuas (p.e. peso, edad, talla) o discontinuas (p.e. número de infección de sitio quirúrgico).

Otra forma de clasificar las variables estadísticas es; variables independientes (antecedentes, causa, etc.) y variables dependientes (efecto, resultado, etc.)

Un error frecuente es encontrarnos al final de nuestra investigación con una gran cantidad de información y decidir en ese momento cómo los analizaremos; en este momento nos damos cuenta que no sabemos

que hacer con esos datos y que muchas veces son erróneos o incompletos y no nos permiten responder a la pregunta que origina nuestra investigación.

La planificación en conjunto del proceso de investigación es fundamental para el éxito de nuestro trabajo. Debe ser realizado en forma completa antes de comenzar con la carga de los datos y tiene distintos pasos a tener en cuenta:

1. Plantear la pregunta que origina la investigación.
2. Decidir los objetivos del trabajo que responderán la pregunta. Aquí se decide las variables a investigar.
3. Definir la hipótesis nula, con un índice de confianza entre 0.05 y 0.01. Este paso permite conocer como se interpretaran los resultados.
4. Decidir el tipo de estudio y la muestra necesaria. De aquí se infieren las pruebas estadísticas a utilizar.
5. Recolectar los datos en plantillas que describan cada variable.
6. Introducir los datos en los formularios electrónicos.
7. Hacer una lista codificadas de las variables a analizar.
8. Corregir las bases de datos utilizando gestores de datos y planillas de cálculos.
9. Describir los datos en forma gráfica y numérica (estadística descriptiva). Esta es la ultima posibilidad para confirmar la calidad de los datos.
10. Inferir las características de la población basandose en la muestra (estadística inferencial).
11. Interpretar los resultados de acuerdo a los objetivo y la hipótesis nula.

Todos los pasos son importantes pero los primeros siete pasos son la base del éxito del análisis estadísticos de los datos.

Formas de seleccionar las distintas pruebas estadística

No existe una receta magistral para seleccionar cual es la prueba estadística más acorde a nuestra investigación, sino que hay distintas manera de elegir la adecuada y esto trataremos de analizar.

La selección de la prueba a utilizar deber ser resuelta, incluso antes de comenzar con la carga de datos.

El primer paso para es el test de hipotesis, para confirmar o descartar la misma y por otro lado el intervalo de confianza.

Tipos de pruebas estadísticas

Las pruebas estadísticas se pueden clasificar en dos grandes grupos, las pruebas paramétricas y las pruebas no paramétricas. La utilización de estas pruebas depende de varios factores, que se pueden analizar desde distintos puntos de vista: distribución de los datos, las observaciones a estudiar, sus variables, la centralización, dispersión y conteo de los datos y el tamaño de la muestra. (Tabla 1).

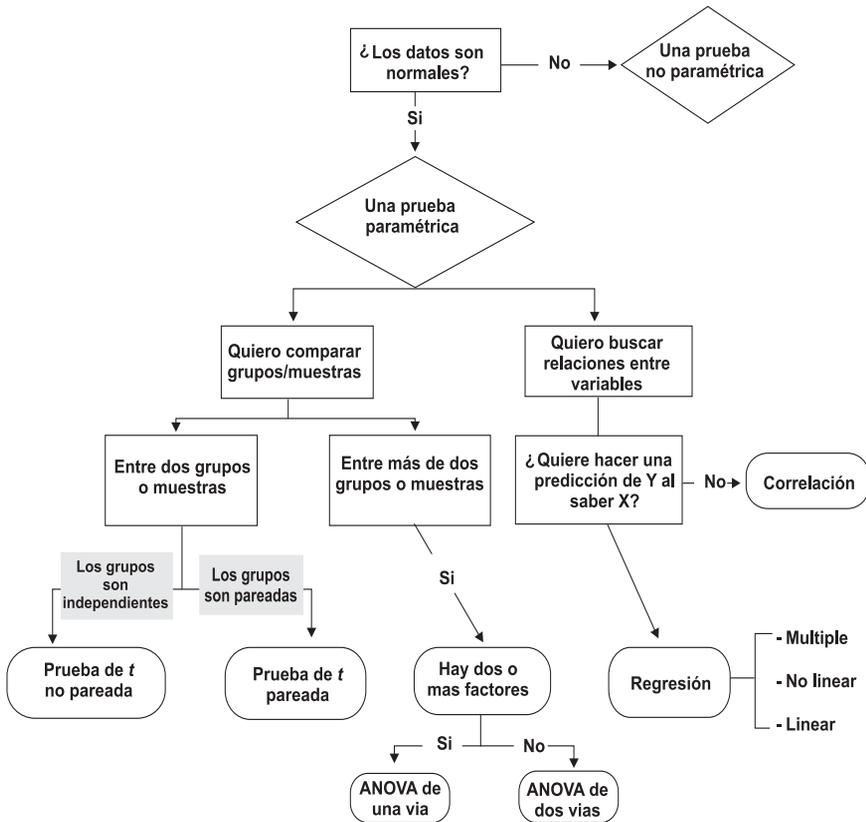
Tabla 1– Diferentes características entre las pruebas paramétricas y no paramétricas

Características	Pruebas paramétricas	Pruebas no paramétricas
Distribución	Normal y variancias homogéneas	Asimétrico y varianzas heterogeneas
Observaciones	Reales	Reales o convertidos a rangos
Variables	Proporcionales o intervalicas	Nominales, ordinales escala de intervalos
Centralización	Promedio	Medianas, modas
Dispersión	Varianzas	Rangos
Conteos	Datos que muestran frecuencias	Deben ser transformados
Tamaño de la muestra	$n > 30$	$n < 30$

En las pruebas paramétricas se pueden nombrar: prueba del valor Z de la distribución normal, prueba de *t* de Student para datos relacionados (muestras dependientes), prueba de *t* de Student para datos no relacionados (muestras independientes), prueba *t* de Student-Welch para dos muestras independientes con varianzas no homogeneas, prueba de χ^2 (ji) de Bartlett para demostrar homogeneidad de varianza t prueba F (análisis de varianza o ANOVA). En la toma de decisión

sobre las pruebas paramétricas, las podemos esquematizar con el diagrama de flujo que se muestra en la figura 1.

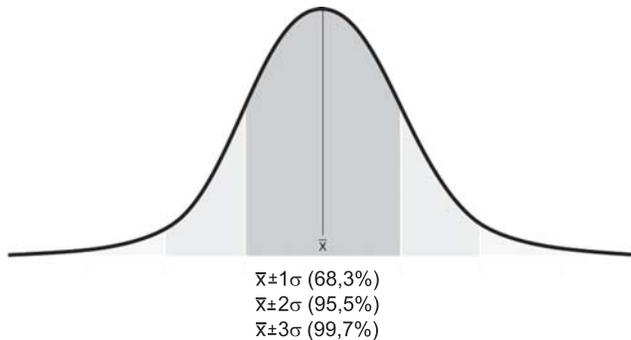
Figura 1 – Diagrama de flujo sobre pruebas paramétricas. Tomado de la Guía de José Carlos Herrera F y Laura E, Carse



Cuando no se cumplan los requisitos paramétricos, en primer lugar se deben transformar los datos para normalizar la distribución y disminuir el sesgo, para ello podemos utilizar pruebas de raíz cuadrada, logarítmicas, arcoseno o chi invertido. Si a pesar de esta transformación se deben realizar las pruebas no paramétricas.

Entonces, en la toma de decisión de cuál prueba utilizar, la primera pregunta a responder es si los datos se distribuyen en forma normal y se expresan en promedio y desvíos estándar (donde se encuentra el 99,7% de la población) siguiendo por ejemplo un gráfico según campana de Gauss (Figura 2) y, si la respuesta es positiva, la prueba a utilizar es paramétrica. La próxima decisión es si quiero comparar distintas muestras o si quiero relacionar distintas variables. Si deseo comparar dos muestras de datos independientes, utilizare la prueba *t* de Student no pareadas; si los grupos son dependientes, se utiliza la prueba *t* de Student pareada. Si son mas de dos muestras utilizaremos el ANOVA de una o dos vías según sean dos o mas factores.

Figura 2– Gráfico correspondiente a una campana de Gauss, que muestra la distribución normal de los datos con sus respectivos desvíos estandar



Si queremos encontrar relaciones entre variables y si quiero conocer un valor, si conozco otro, utilizaremos pruebas de regresión y por el contrario una de correlación.

Cuando la distribución de los datos no es normal, se expresa en términos de modo y la amplitud, la mediana y los percentiles y se utilizarán las pruebas no paramétricas, se pueden nombrar: prueba χ^2 de Pearson, prueba binomial, prueba de Anderson-Darling, prueba de Cochran, prueba de Cohen Kappa, prueba de Fisher, prueba de Friedman, prueba de Kendall; prueba de Kolmogórov-Smirnov, prueba de Kruskal-Wallis, prueba de Kuiper, prueba de Mann-Whitney (Wilcoxon), prueba de McNemar, prueba de la mediana, prueba de Siegel-Tukey, prueba de los signos, Coeficiente de correlación de Spearman, Tablas de contingencia, prueba de Wald-Wolfowitz y la prueba de los rangos con signo de Wilcoxon.

Las ventajas de utilizar pruebas no paramétricas es que son: sencillas; fáciles de aplicar y usar, se utilizan a los grupos mayores de población, se pueden usar con datos ordinales o nominales y se pueden utilizar en muestras pequeñas.

Tamaño de la muestra

Otro punto importante para decidir cual o cuales son las variables a utilizar, es el tamaño de la muestra (n). Por ello muchos paquetes estadísticos no permiten conocer cuanto datos debemos poseer para que los resultados puedan ser extrapolables y sus reportes consistentes desde el punto de vista de la investigación.

Distintos autores no llegan a una conclusión definitiva, pero existe un consenso que cuanto más grande es la muestra, más se puede disminuir el error aleatorio. El volumen mínimo para considerar una muestra como representativa podría ser 20 o 30.

En resumen, el tamaño de la muestra es el necesario para realizar un estudio con el menor número posible de elementos que sean suficientes para obtener los resultados con la precisión fijada.

Tipo de variable

Un primer paso para el análisis de los datos es conocer las variables con la que se trabaja.

Las pruebas estadísticas tienen mayor potencia si la variable es cuantitativa, ya que contiene más información que si fuese cualitativa.

De acuerdo al tipo de variables y su correlación, podemos seleccionar distintas pruebas. Si son dos variables cualitativas utilizaremos pruebas no paramétricas como el test χ^2 . Si una es cualitativa y la otra cuantitativa podremos seleccionar entre algunas paramétricas (t de Student y análisis de varianza) y no paramétricas (U de Mann-Whitney; prueba de rangos de Wilcoxon y la prueba de Kruskal-Wallis. Si ambas variables son cuantitativas, también podremos seleccionar entre paramétricas (coeficiente de correlación de Pearson) y no paramétricas (Coeficiente de Correlación de Spearman). Tabla 2.

Tabla 2 – Selección de pruebas de acuerdo a distintas variables

Variable 1	Variable 2	Prueba paramétrica	Prueba no paramétrica
Cualitativa	Cualitativa	t de Student	Test de la χ^2
Cuantitativa	Cualitativa Cuantitativa	Student para datos pareados Análisis de varianza	U de Mann-Whitney Prueba de los rangos de Wilcoxon Prueba de Kruskal-Wallis
Cuantitativa		Coeficiente de Correlación de Pearson	Coeficiente de Correlación de Spearman

Otra forma que la relación entre las distintas variables nos puede indicar el tipo de prueba a elegir podría ser:

- a. Dos variables continuas con “n” grande, se utiliza coeficiente de Pearson.
- b. Dos continuas con “n” mediana o pequeña; coeficiente de correlación de Spearman o Coeficiente de correlación de Kendall.
- c. Continua vs. ordinal; Coeficiente de correlación de Kendall o el Test de tendencia de Cuzick.
- d. Continua vs. nominal; ANOVA paramétrico.
- e. Continua vs. dicotómica; test t o Test Mann-Whitney.
- f. Dos variables ordinales; Coeficiente de correlación de Kendall; Test de tendencia de Cuzick o χ^2 .
- g. Ordinal vs. nominal; Test de χ^2 .
- h. Ordinal vs. dicotómica; χ^2 de tendencia lineal.
- i. Dos variables nominales; χ^2 .
- j. Nominal vs. dicotómica; χ^2 sin tendencia lineal.
- k. Dos variables dicotómicas; Test de Fisher; Riesgo Relativo o Odds Ratio.

Pregunta

Como lo denota la frase acuñada por Albert Einstein: “La formulación de un problema es más importante que su solución”, también la elección de la prueba se relaciona con la pregunta que motivó el trabajo.

Si es un solo grupo de análisis, la prueba a elegir puede ser el intervalo de confianza o la prueba t. Si pregunta se observa que se

detalla que nos enfrentamos a dos muestras independientes, con una variable cuantitativa, en escala continua y deseamos saber si existen diferencias estadísticamente significativas, debemos utilizar el test *t* Student no apareado. Si por el contrario son dos grupos dependientes, con variables cuantitativas, escalas discontinuas, las pruebas pueden ser: test de Student apareados, Wilxicom o Correlación de Pearson. Si son tres o más grupos independientes, las pruebas pueden ser ANOVA o Chi². Si son tres o más grupos dependientes las prueba de Cochran o el ANOVA con observaciones repetidas. Si por el contrario son grupos multivariado, las pruebas seran logaritmicas lineales o regresión.

Tipo de estudio

Podemos seleccionar la prueba estadística a utilizar relacionandola al tipo de estudio que planificamos, por eso desde la planificación de nuestra experiencia pordemos saber las pruebas que vamos a requerir.

Para los estudios de cohorte, estudios transversales y los ensayos clinicos, se puede utilizar el riesgo relativo; para los de casos y controles se utiliza el Odds Ratio.

Tipos de datos

Los grupos de datos pueden ser independientes o dependientes.

Son independientes cuando los grupos no tienen relación y el tamaño de cada uno de ellos es distinto.

Son dependientes, cuando se relacionan directamente, son grupos causa-efecto. La "n" de cada grupo son similares.

Conclusión

El plan de análisis estadístico es una etapa definida antes de la recolección de los datos.

Es consecuencia de un buen marco teórico del tema a investigar.

El diseño del formulario y el tamaño de la muestra depende de la complejidad del análisis estadístico establecido.

Manuel Gomez Gomez y coladoradores, nos ilustran en este tema concluyendo que la elección de la prueba estadística depende de:

1. Del tipo de diseño de la investigación.
2. Del objetivo u objetivos del trabajo.

3. Del interés particular de los autores.
4. De la distribución de los datos.
5. De la manera como se recopilan los datos y.
6. Del potencial de sesgo y confusión de los datos obtenidos.

LECTURA COMPLEMENTARIA

1. Castañeda, MB; Cabrera, AF; Navarro, Y; de Vries, W: Análisis de los datos I (Cap 3; 26 - 54) y II (Cap 4; 55 - 70) en Procesamiento de datos y análisis estadísticos utilizando SPSS. Un libro práctico para investigadores y administrativos educativos. Ed. Ediprucs, 2010.
2. Cazau, P: Fundamentos de estadísticas. UBA; 2006.
3. Gomez-Gomez, M; Danglot-Banck, C; Vega Franco, L: Cómo seleccionar una prueba estadística. Primera de dos partes. Rev Mex Pediatr 2013; 80 (1); 30-34.
4. Fernandez, P; Livacic-Rojas, P; Vallejo, G: Cómo elegir la mejor prueba estadística para analizar un diseño de medidas repetidas. Int J Clin Health Psychol; 7 (1): 154-175.
5. Herrera F, JC; Carse, LE: Guía de aplicación de pruebas estadísticas en el programa Systat 7.0 para ciencias biológicas y forestales. Sección 4.
6. Tetitla Munive, JM: Cómo elegir la prueba estadística. En Metodología de la investigación para el área de la salud. Ed McGraw Hill. 2013; Cap 9: 71-79.
7. Universidad de Alcalá - Departamento de ecología: Métodos de análisis de datos en ecología.

CAPÍTULO 7

BIOESTADÍSTICA DESCRIPTIVA

Alberto Aceves Pérez

Cuando se habla de estadística, habitualmente se piensa en una organización de datos numéricos presentados de forma ordenada y sistemática. Esta idea es consecuencia de un concepto de uso común que existe sobre el término y que cada vez se ha extendido debido a la influencia de nuestro entorno, ya que actualmente es prácticamente imposible que cualquier medio de difusión no nos hable acerca de cualquier tipo de información estadística.

De este modo ya hablando específicamente en el campo de la investigación como la medicina, se comienza a percibir la estadística como una entidad diferente, convirtiéndose en una herramienta que permite obtener resultados y por tanto beneficios, en cualquier tipo de estudio, cuyos movimientos y relaciones, por su naturaleza variable, no pueden ser abordadas desde una perspectiva determinante. Se puede definir a la estadística como una ciencia que estudia cómo debe emplearse la información y como dar una guía de acción en situaciones prácticas que tienen como base una pregunta.

La Estadística se ocupa de los métodos y procedimientos para recoger, clasificar, resumir, hallar regularidades y analizar los datos, siempre y cuando la variabilidad e incertidumbre sea una causa intrínseca de los mismos; así como de realizar inferencias a partir de ellos, con la finalidad de ayudar a la toma de decisiones y en su caso formular predicciones.

También considerada como el arte de realizar conclusiones a partir de una serie de datos imperfectos. Los datos son comúnmente imperfectos en el sentido que aun cuando posean información útil no nos cuentan la historia completa. Es necesario contar con métodos que nos permitan extraer información a partir de los datos observados para comprender mejor las situaciones que los mismos representan.

Todos, aún los expertos en hacer investigación, tienen problemas al enfrentarse con listados de datos. Existen diversos métodos estadísticos

cuyo propósito es ayudar a poner de manifiesto las características sobresalientes e interesantes de los datos que pueden ser usados en casi todas las áreas del conocimiento.

Los métodos estadísticos pueden y deberían ser usados en todas las etapas de un proceso de investigación, desde el comienzo hasta el final. Existe el convencimiento de que la estadística trata con el análisis de los datos, pero este punto de vista excluye aspectos vitales relacionados con el diseño de las investigaciones. Es de suma importancia tomar conciencia que la elección del método de análisis para un problema, se basa tanto en el tipo de datos disponibles como en la forma en que fueron recolectados.

¿Para qué sirve y por qué estudiar bioestadística?

La respuesta pudiera considerarse simple; porque los datos estadísticos y las conclusiones obtenidas aplicando el método científico ejercen una profunda influencia en prácticamente todos los campos de la actividad humana. En particular, la estadística invade cada vez más cualquier investigación realizada en el campo de la medicina. Este crecimiento, probablemente relacionado con el interés por aumentar la credibilidad y confiabilidad de las investigaciones, no garantiza que en todos los casos la metodología estadística haya sido correctamente utilizada, o peor aún, que sea válida.

Entonces la importancia de la aplicación correcta de la bioestadística se preocupa:

1. Porque las conclusiones pueden ser incorrectas.
2. Porque no todos los lectores tienen la capacidad y la conciencia de detectar el error.

El estudio de la bioestadística y el modo de pensamiento que se genera a partir del mismo, capacita a la persona para evaluar objetiva y efectivamente si la información que recibe (vía tablas, gráficos, porcentajes, tasas, etc.) es relevante, adecuada, veraz y objetiva. Por supuesto, la interpretación de cualquier problema requiere, no sólo de conocimientos metodológicos sino también, de un profundo conocimiento del tema. Aún cuando una persona no esté interesada

en especializarse en estadística, un entrenamiento básico en el tema permite una mejor comprensión de la información.

Áreas de la bioestadística

A continuación se describe brevemente cada una de las áreas en que puede dividirse la bioestadística:

- Diseño: Planificación y desarrollo de investigaciones.
- Descripción: Resumen y exploración de los datos.
- Inferencia: Realizar predicciones acerca de las características específicas de una muestra en base a su información.

Diseño

Es una actividad fundamental y consiste en definir como se desarrollará la investigación para dar respuesta a las preguntas que motivaron la misma. La recolección de los datos requiere en general de un gran esfuerzo, por lo que, dedicar especial cuidado a la etapa de planificación de la investigación ahorra trabajo en las siguientes etapas. Un estudio bien diseñado resulta simple de analizar y las conclusiones suelen ser obvias. Un experimento pobremente diseñado o con datos inapropiadamente recolectados o registrados puede ser incapaz de dar respuesta a las preguntas que motivaron la investigación, más allá de lo sofisticado que sea el análisis estadístico.

Aún en los casos en que se estudian datos ya registrados, en que estamos restringidos a la información existente, los principios del buen diseño de experimentos, pueden ser útiles para ayudar a seleccionar un conjunto razonable de datos que esté relacionado con el problema de interés.

Descripción

Los métodos de la estadística descriptiva o análisis exploratorio de datos ayudan a presentar los datos de modo tal que sobresalga su estructura. Hay varias formas simples e interesantes de organizar los datos en gráficos que permiten detectar tanto las características sobresalientes como las características inesperadas. El otro modo de describir los datos es resumirlos en uno o dos números que pretenden caracterizar el conjunto con la menor distorsión o pérdida de información posible.

Explorar los datos, debe ser la primera etapa de todo análisis de datos. ¿Por qué no analizarlos directamente? En primer lugar porque las computadoras no son demasiado hábiles (sólo son rápidas), hacen aquello para lo que están programadas y actúan sobre los datos que les ofrecemos. Datos erróneos o inesperados serán procesados de modo inapropiado y ni usted, ni la computadora se darán cuenta a menos que realice previamente un análisis exploratorio de los datos.

Inferencia

Hace referencia a un conjunto de métodos que permiten hacer predicciones acerca de características de un fenómeno sobre la base de información parcial acerca del mismo.

Los métodos de la inferencia nos permiten proponer el valor de una cantidad desconocida o decidir entre dos teorías contrapuestas cuál de ellas explica mejor los datos observados. El fin último de cualquier estudio es aprender sobre las poblaciones.

Pero es usualmente necesario y más práctico, estudiar solo una muestra de cada una de las poblaciones.

Por lo tanto la estadística se puede clasificar en inferencial y descriptiva, la primera cuando el objeto de estudio consiste en derivar las conclusiones obtenidas a un conjunto de datos más amplio y la segunda que nos compete en este capítulo es cuando los resultados del análisis no pretenden ir mas allá de un conjunto de datos.

Estadística descriptiva: Describe, analiza y representa un grupo de datos utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos.

Estadística inferencial: Apoyándose en el cálculo de probabilidades y a partir de datos muestrales, efectúa estimaciones, decisiones, predicciones u otras generalizaciones sobre un conjunto mayor de datos.

LECTURA COMPLEMENTARIA

1. A. Martín Andrés, J.D. Luna del Castillo, Bioestadística para las Ciencias de la salud. Norma, Granada, 1994.
2. D. Peña Sánchez de Rivera, Estadística: Modelos y Métodos, Alianza Universidad Textos, Madrid, 1994.
3. E. Cobo, P. Muñoz y J.A. González.(2007) Bioestadística para no estadísticos. Ed. Elsevier/Masson.
4. P. Armitage, G. Berry, Estadística para la Investigación Biomédica. Doyma, Barcelona, 1992.
5. J. Montanero Fernández, Manual de bioestadística. Universidad de Extremadura. Departamento de matemáticas. 2001.
6. J. S. Milton. Estadística para Biología y Ciencias de la Salud. Ed. Interamericana. McGraw-Hill.
7. Peña Sánchez de Rivera, D. (1987): "Estadística. Modelos y Métodos. Volumen 2". Alianza Editorial. Madrid.

CAPÍTULO 8

BIOESTADÍSTICA ANALÍTICA O INFERENCIAL

José María Jiménez Avila

Cuando se inicia una investigación, en ocasiones no es posible abarcar toda la población, por lo que regularmente se selecciona una muestra, con la esperanza de que esta pueda ser representativa de la población que se va a estudiar, aunque esto no siempre ocurre, ya que en ocasiones no reflejan la realidad del universo de donde surge, ya que existen condiciones como el azar, las cuales tienen un papel muy importante en el proceso de la investigación y por consecuencia un efecto en su resultado.

La rama de las matemáticas que estudia los resultados del azar, es la teoría de la probabilidad y este es el núcleo de la inferencia estadística.

El análisis inferencial o analítico, se realiza después de haber hecho el análisis descriptivo, donde se evidencia como es y cuáles son las características del grupo de población y las características particulares de nuestra muestra.

A manera de definición, la inferencia estadística, es el procedimiento estadístico, que permite establecer conclusiones, teniendo en cuenta los efectos de probabilidad (resultados) y regularmente es en esta etapa, donde se aceptan o se rechazan las hipótesis, así como el estimar los patrones y la intensidad de la asociación entre las variables, enmarcando el resultado dentro del concepto: “existen o no diferencias” entre los grupos.

Cuando se realizan experimentos, se deriva una serie de probabilidades con resultados posibles y mediante la probabilidad del evento, se puede expresar la confianza de que el evento ocurra al observar el experimento y se puede representar del “0” al “1”, de manera que el “1” indica que el evento ocurrirá con toda seguridad, mientras que el “0” corresponde a un evento que con toda seguridad NO ocurrirá.

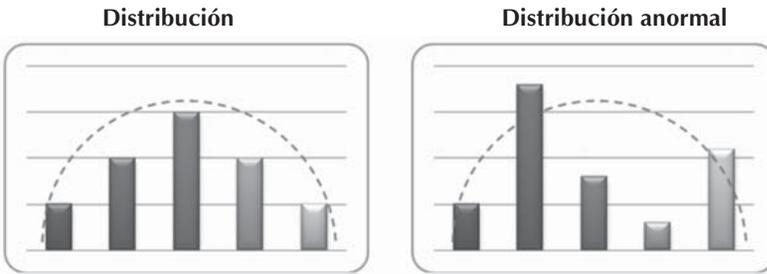
Existen 2 tipos de inferencia estadística: la Prueba de Hipótesis y la Estimación de Intervalo. (se revisarán en capítulos posteriores).

Tipos de pruebas

Los procedimientos estadísticos pueden clasificarse en 3 grupos:

- Aquellos que suponen la Distribución Normal, que en conjunto se reconocen como pruebas paramétricas.
- Aquellos que No consideran la distribución normal de los datos, que se identifican como procedimientos libres de distribución. (distribución binomial y de Poisson).
- Aquellos que no se refieren a parámetros de población que se conocen como pruebas no paramétricas. (χ^2 , prueba exacta de Fisher). Figura 1.

Figura 1 – Distribución normal y anormal de los datos



Nota: En algunos textos de estadística, se agrupa tanto a los procedimientos no paramétricos como a los libres de distribución y se presenta con el título de “no paramétrico”.

Se cuenta con una gran cantidad de pruebas que podemos clasificarlas en 2 tipos: Pruebas paramétricas y no paramétricas. Figura 2.

Figura 2 – Tipos de pruebas estadísticas

Pruebas paramétricas en cirugía

T de Student
ANOVA
Correlación de Pearson

Pruebas no paramétricas en cirugía

χ^2
Prueba de Wilcoxon
U de Mann-Whitney
Kruskal-Wallis
Correlación Spearman
McNemar
Regresión lineal
Probabilidad de Poisson
Odds Ratio

Paramétricas

Las pruebas paramétricas son las que se consideran de mayor potencia estadística y las que proporcionan mayor información, lo que permite establecer estimaciones de los parámetros de la población a través de muestras estadísticas.

Requisitos:

1. Variables con distribución normal, misma que puede determinarse con la prueba de Kolmogorov-Smirnov y la prueba de Shapiro-Wil.
2. No se usa para variables ordinales.
3. Se requiere un mínimo de 30 personas por grupo en la muestra, ya que mientras más grande es la muestra, más exacta será la estimación probabilística.
4. La hipótesis se realiza sobre los valores numéricos, especialmente el promedio de la población.

No paramétricas

Las pruebas no paramétricas o de distribución libre, son de fácil aplicación y estas se pueden utilizar para realizar un análisis de variables ordinales o normales, esto cuando no se requiere establecer o plantear una inferencia sobre cuáles son los parámetros de la población.

Requisito:

1. Variables de tipo ordinal o nominal.
2. Cuando 2 series de observaciones se derivan de distintas poblaciones.
3. La muestra puede ser menor o igual a 20 personas.
4. La hipótesis se realiza sobre los rangos, mediana o frecuencia de los datos.

Es importante conocer todos estos requisitos y familiarizarse con todas y cada una de ellas, ya que la selección de la prueba estadística dependerá del tipo de variable, de la presencia o ausencia de normalidad de la distribución de los grupos que se van a analizar y de cómo se formuló la pregunta de investigación (PICOT). Figura 3.

Figura 3 – Clasificación de acuerdo al objetivo y la variable

Prueba Estadística según Objetivo y Variable					
Tipo de variable	Tipo de muestra	Diferencia		Mostrar relación	Predicir 1 variable***
		2 Grupos	3 Grupos	2 Variables	Variables desenlace
Cuantitativa (distribución normal)	NR	t Student *	Anova 1 factor	Pearson	Regresión lineal
	R	t Student**	Anova 1 factor		
Cualitativa ordinal	NR	U Mann-Whitney	Kruskal-Wallis	Spearman	
	R	Wilcoxon	Friedman		
Cualitativa dicotómica	NR	X2 (Fisher)	X2	Coeficiente phi	Regresión logística
	R	McNemar			
NR= no relacionada, R= relacionada					
*t de Student para muestras independientes					
**t de Student para muestras relacionadas					
***La variable predictiva puede ser cuantitativa, dicotómica u ordinal (transformar a variables tipo dummy)					

Interpretación

Este tipo de pruebas, permite tener herramientas en las que el cirujano ante los pacientes, pueda pronosticar la sobrevida o las posibilidades de que una cirugía se llegará a infectar o la probabilidad de tener cualquier tipo de complicación, en este caso el cirujano basará su estimación, en el conocimiento que tenga de lo escrito en la literatura, de las características del paciente así como su expectativa y su evolución, considerando su propia experiencia y los recursos de que dispone para la atención del enfermo y posible desenlace.

Esto permite aplicar los conceptos fundamentales de la “Medicina Basada en la Evidencia”, la cual se considera como el uso consciente, explícito y juicioso de la “mejor evidencia científica” disponible, para tomar decisiones sobre los pacientes. Figura 4.

Figura 4 – ¿Qué es hacer Medicina Basada en Evidencia?

LECTURA COMPLEMENTARIA

1. Celis de la Rosa A. Bioestadística. 1ª ed en español. Manual Moderno. 2004.
2. Daniel WW: Estadística con aplicaciones a las Ciencias Sociales y a la Educación. 1ª ed en español. McGraw-Hill, México. 1988.
3. Gómez Gómez M, Danglot Banck C, Velásquez Jones L. Sinopsis de pruebas estadísticas no paramétricas. Cuando usarlas. Revista Mexicana de Pediatría. Vol. 70 No.2 2003.
4. Gómez Gómez M, Danglot Banck C, Velásquez Jones L. Bases para la revisión de artículos médicos. Revista Mexicana de Pediatría. Vol. 68 No. 4. Julio – Agosto 2001.
5. Pita Fernández, S. Tipos de estudios clínico epidemiológicos. Atención Primaria de la Red. 2001.
6. Siegel S, Castellan NJ: Nonparametric statistics for the behavioral sciences. 2ª ed. McGraw-Hill, New York, 1988.
7. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA 1992; 268: 2420-5.

CAPÍTULO 9

MEDIDAS DE TENDENCIA CENTRAL

Alberto Aceves Pérez

Tras la elaboración de la tabla y su representación gráfica, en la mayoría de las ocasiones resulta más eficaz “condensar” dicha información en algunos números que la expresen de forma clara y concisa.

Los fenómenos biológicos no suelen ser constantes, por lo que será necesario que junto a una medida que indique el valor alrededor del cual se agrupan los datos, se asocie una medida que haga referencia a la variabilidad que refleje dicha fluctuación.

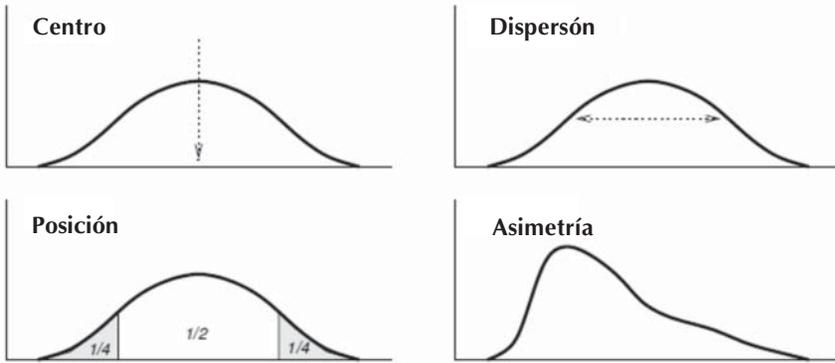
Por tanto el siguiente paso y objeto de este capítulo consistirá en definir algunos tipos de medidas (estadísticos o parámetros) que los sintetizan aún más.

Es decir, dado un grupo de datos organizados en una distribución de frecuencias (o bien una serie de observaciones sin ordenar), pretendemos describirlos mediante dos o tres cantidades sintéticas.

En este sentido pueden examinarse varias características, siendo las más comunes:

1. La tendencia central de los datos.
2. Los datos de ocupan ciertas posiciones.
3. La simetría de los datos.
4. La forma en la que los datos se agrupan.
5. La dispersión de los datos en relación al centro.

Figura 1 – Medidas representativas de un conjunto de datos estadísticos



En este capítulo y siguiendo este orden, iremos estudiando los estadísticos que nos van a orientar sobre cada uno de estos niveles de información: valores alrededor de los cuales se agrupa la muestra, la mayor o menor fluctuación alrededor de esos valores, nos interesaremos en ciertos valores que marcan posiciones características de una distribución de frecuencias así como su simetría y su forma.

El centro es fácil de identificar si la distribución es simétrica, pero es difícil si la distribución es asimétrica. Por esta razón, no hay una única medida de posición para resumir una distribución. Si la distribución es simétrica diferentes medidas conducirán a similares resultados. Si la distribución es claramente asimétrica diferentes propuestas apuntarán a distintos conceptos de “centro” y por lo tanto los valores serán diferentes.

Las tres medidas más usuales de tendencia central son:

1. Media.
2. Mediana.
3. Moda.

En ciertas ocasiones estos tres estadísticos suelen coincidir, aunque generalmente no es así. Cada uno de ellos presenta ventajas e inconvenientes que precisaremos más adelante. En primer lugar vamos a definir los conceptos anteriores.

La media

Es la medida de posición más frecuentemente usada. Para calcular la media aritmética o promedio de un conjunto de observaciones se

suman todos los valores y se divide por el número total de observaciones.

Si tenemos una muestra de n observaciones y denotadas por X_1, X_2, \dots, X_n , definimos la *media muestral* \bar{X} del siguiente modo:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n}$$

Ejemplo:

$$X_1 = 10 \quad X_2 = 14 \quad X_3 = 12 \quad X_4 = 11 \quad X_5 = 12 \quad X_6 = 13$$

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_6}{n} = \frac{10 + 14 + 12 + 11 + 12 + 13}{6} = \frac{72}{6} = 12$$

Características de la media

1. Se usa para datos numéricos.
2. Representa el centro de gravedad o el punto de equilibrio de los datos.
3. La suma de la distancia de los datos a la media es cero.
4. Es muy sensible a la presencia de datos atípicos.

Aunque la media es la medida más simple de tendencia central, otras medidas nos ofrecen mayor información y ocasionalmente son más apropiadas.

La mediana

La mediana es el dato que ocupa la posición central en la muestra ordenada de menor a mayor.

$$\left(\frac{n+1}{2} \right)$$

$$X_1 = 10 \quad X_2 = 14 \quad X_3 = 12 \quad X_4 = 18 \quad X_5 = 11 \quad X_6 = 23$$

Ordenamos los datos:

$$10 \quad 11 \quad 12 \quad 14 \quad 18 \quad 23$$

$$\text{Posición de la mediana} \Rightarrow \frac{6+1}{2} = 3.5$$

Obtenemos la mediana

promediando el tercer y cuarto dato:

$$\tilde{X} = \frac{12+14}{2} = 13.$$

Ejemplo:

Características de la mediana:

1. Como medida descriptiva, tiene la ventaja de no estar afectada por las observaciones extremas, ya que no depende de los valores que toma la variable, sino del orden de las mismas. Por ello es adecuado su uso en distribuciones asimétricas.
2. Es de cálculo rápido e interpretación sencilla.
3. A diferencia de la media la mediana es siempre un valor de la variable que estudiamos.

La moda

La moda es el dato que ocurre con mayor frecuencia en el conjunto.

Es una medida de poca utilidad salvo para datos categóricos en los que suele interesar identificar la categoría con mayor cantidad de datos. En una muestra de datos numéricos, puede ocurrir que la moda

Figura 2 – Resumen de las medidas de tendencia central

MEDIDAS DE TENDENCIA CENTRAL																					
DATOS SIN AGRUPAR	DATOS AGRUPADOS																				
(ordenados) x_1, x_2, \dots, x_N	<table border="1"> <thead> <tr> <th>Interv.</th> <th>x_i</th> <th>n_i</th> <th>N_i</th> </tr> </thead> <tbody> <tr> <td>l_0-l_1</td> <td>x_1</td> <td>n_1</td> <td>N_1</td> </tr> <tr> <td>l_1-l_2</td> <td>x_2</td> <td>n_2</td> <td>N_2</td> </tr> <tr> <td>...</td> <td>...</td> <td>...</td> <td>...</td> </tr> <tr> <td>$l_{k-1}-l_k$</td> <td>x_k</td> <td>n_k</td> <td>N_k</td> </tr> </tbody> </table>	Interv.	x_i	n_i	N_i	l_0-l_1	x_1	n_1	N_1	l_1-l_2	x_2	n_2	N_2	$l_{k-1}-l_k$	x_k	n_k	N_k
Interv.	x_i	n_i	N_i																		
l_0-l_1	x_1	n_1	N_1																		
l_1-l_2	x_2	n_2	N_2																		
...																		
$l_{k-1}-l_k$	x_k	n_k	N_k																		
MEDIA $\bar{x} = \frac{x_1 + \dots + x_n}{N}$	$\bar{x} = \frac{n_1 x_1 + \dots + n_k x_k}{N}$																				
MEDIANA Primera observación que deja debajo de sí estrictamente a las $[N/2]$ observaciones menores: $x_{[N/2]+1}$	$M_{ed} = l_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} \cdot a_i$																				
MODA $M_{oda} = x_i$ de mayor frecuencia	$M_{oda} = l_{i-1} + \frac{n'_i - n'_{i-1}}{(n'_i - n'_{i-1}) + (n'_i - n'_{i+1})} a_i$																				

sea un valor que se repite un cierto número de veces, pero que no es típico.

Características de la moda:

1. Es muy fácil de calcular.
2. Puede no ser única.

Relación entre la media, mediana y moda

En el caso de distribuciones unimodales, la mediana está con frecuencia comprendida entre la media y la moda (incluso más cerca de la media).

En distribuciones que presentan cierta inclinación, es más aconsejable el uso de la mediana. Sin embargo en estudios relacionados con propósitos estadísticos y de inferencia suele ser más apta la media.

LECTURA COMPLEMENTARIA

1. A. Martín Andrés, J.D. Luna del Castillo, Bioestadística para las Ciencias de la salud. Norma, Granada, 1994.
 2. E. Cobo, P. Muñoz y J.A. González.(2007) Bioestadística para no estadísticos. Ed. Elsevier/Masson.
 3. D. Peña Sánchez de Rivera, Estadística: Modelos y Métodos, Alianza Universidad Textos, Madrid, 1994.
 4. J. Montanero Fernández, Manual de bioestadística. Universidad de Extremadura. Departamento de matemáticas. 2001.
 5. J. S. Milton. Estadística para Biología y Ciencias de la Salud. Ed. Interamericana. McGraw-Hill.
 6. P. Armitage, G. Berry, Estadística para la Investigación Biomédica. Doyma, Barcelona, 1992.
- Peña Sánchez de Rivera, D. (1987): "Estadística. Modelos y Métodos. Volumen 2". Alianza Editorial. Madrid.

CAPÍTULO 10

MEDIDAS DE DISPERSIÓN

(Rango, Varianza, Desviación Estándar, Coeficiente de Varianza)

Delio Martins

Un aspecto importante en el estudio descriptivo de los datos de un trabajo es la determinación de la variabilidad o dispersión de esos datos con relación a la medida de ubicación del centro de la muestra.

Como ejemplo, mencionamos un estudio sobre escoliosis que evaluó el efecto de un determinado tratamiento quirúrgico A comparado a otro tratamiento quirúrgico B en niños con diferentes grados de curvatura. El promedio de la magnitud de las curvas fue de 60 grados (Cobb), tanto en el grupo operado cuanto en el grupo control y ambos grupos presentaron moda y mediana iguales. ¿Esto bastaría para decir que esos grupos son similares?

Veamos algunos conceptos importantes.

Rango

Es la medida entre el mayor y menor valor en un conjunto de datos.

Por lo tanto, si consideramos el ejemplo anterior acerca del grado de curvatura de la escoliosis de los niños y observamos los 2 grupos con las siguientes magnitudes de curvas:

Tratamiento A: 40, 42, 44, 50, 65, 65, 65, 68, 70, 70, 81

Tratamiento B: 48, 48, 50, 54, 65, 65, 65, 66, 66, 68

El rango corresponde a la diferencia entre el valor más alto y el más bajo observados. En ese caso, vemos que en el tratamiento A, el rango fue de 41 grados mientras que en el tratamiento B fue de solamente 20 grados, demostrando una gran variación entre los grupos.

Sin duda, el rango es la medida de dispersión más fácilmente visualizada, sin embargo, se basa solamente en dos valores y no considera la distribución de los datos.

Varianza

Veamos otra medida de dispersión, la Varianza.

Para simplificar nuestra comprensión, vamos a reducir nuestra muestra:

Tratamiento A: 40, 50, 65, 81

Tratamiento B: 48, 54, 66, 68

En ambos casos, tenemos un promedio de 59 grados Cobb.

$$\mathbf{A:} (40 + 50 + 65 + 81 / 4) = 59$$

$$\mathbf{B:} (48 + 54 + 66 + 68 / 4) = 59$$

Los rangos continúan diferentes:

$$\mathbf{A} = 81 - 40 = 41$$

$$\mathbf{B} = 68 - 48 = 20$$

Observamos que el grupo del tratamiento A es más disperso que el grupo del tratamiento B.

Para comenzar a entender la desviación estándar calcularemos, primero, en el grupo A, la distancia a que se encuentra cada observación del promedio del grupo. Tendremos:

$$(40 - 59, 50 - 59, 65 - 59, 81 - 59)$$

De esa forma, tenemos los valores:

$$(-19, -9, 6, 22)$$

Ahora, una prueba: calcule el promedio aritmético de esos números....

¿Lo ha calculado???

¿Ha observado que la suma de todos los desvíos con relación al promedio aritmético es igual a cero? ¡Esa es una propiedad aritmética que siempre se repetirá!

De esa forma, ¡los valores negativos siempre anularán a los positivos y no es posible obtener una medida de dispersión utilizando este cálculo algebraico!

Entonces, lo intentaremos de otra forma...

Como en el cálculo nos importa solamente la magnitud del desvío y no si son positivos o negativos, ¡lo haremos de una forma en que todos los números sean positivos!

¡Eleve al cuadrado todos los números!

$$(-19)^2, (-9)^2, (6)^2, (22)^2$$

O sea:

$$361, 81, 36, 484$$

De esa forma, el promedio de los cuadrados de los números se vuelve una medida razonable de variabilidad.

Sin embargo, para calcular el promedio de esos valores de tal forma que tenga propiedades que funcionen con todas las teorías (aquí, créanme: eso es más de lo que le gustaría saber sobre ese tema), no dividimos por el número total de casos (aquí: $n=4$), sino que dividimos por el número total $- 1$ ($n - 1$) que en este caso: ($4-1=3$).

De esa forma:

$$361 + 81 + 36 + 484 / 3 = 320,6$$

Así, calculamos la variancia de la muestra. Ahora, pasemos al próximo paso...

Desviación Estándar

Es probable que el desvío estándar sea la media de dispersión más utilizada en estadística, pero también es probable que sea la menos entendida. Sin embargo, ahora que entendemos la variancia, esto nos resultará más fácil.

Como lo dice su nombre, es un estándar de desvío del promedio. Representa la distancia normal de cualquier punto en el conjunto de datos hasta el promedio.

La desviación estándar no es nada más que la raíz cuadrada de la variancia. Con eso, evitamos tener que trabajar con cuadrados de valores.

Entonces, en nuestro ejemplo, la desviación estándar es

$$\sqrt{320,6} = 17,90.$$

La desviación estándar de toda la población se representa con la letra griega σ pero como casi nunca conocemos toda la población y generalmente trabajamos con una muestra, la desviación estándar de una muestra de la población se designa con la letra S .

Más importante que calcular el desvío estándar es entender su significado. Una desviación estándar pequeña significa que los valores del conjunto de datos están, en promedio, próximos al centro del conjunto. Una desviación estándar alto significa que los valores del conjunto de datos están, en promedio, alejados del centro.

Una desviación estándar alto no significa, necesariamente, algo malo. Solamente refleja una gran variabilidad dentro del grupo evaluado.

Otro cuidado que se debe tener es observar, siempre, la unidad en que está la desviación estándar. Por ejemplo, si se considera una desviación estándar de 2 años en una muestra determinada, se lo puede alterar a una desviación estándar de 24, si los datos están en meses.

En una curva de distribución normal, presente en la mayoría de los trabajos científicos, tenemos como estándar que el 68% de la muestra está alejada en hasta 1 desviación estándar del promedio y que el 95% de la muestra está alejado en hasta 2 desviaciones estándar del promedio. Así, por ejemplo, si tenemos un cálculo de la cantidad de proteoglicano presente en el disco intervertebral, recibimos la información de que el promedio de proteoglicano era de 250 mg/g y que la desviación estándar era de 8,75 mg/g en esa muestra, sabremos que el 68% de los discos presentan dosis de proteoglicanos dentro de 1 desvío estándar (8,75 mg/g) del promedio (250 mg/g). Por lo tanto, el 68% de los discos contenían entre $250 - 8,75$ mg/g e $250 + 8,75$ m/g de proteoglicano, o sea, el 68% de los discos de la muestra contenía entre 241,25 y 258,75 mg/g de proteoglicano. De la misma manera, el 95% de las muestras medidas, tiene entre $250 - 2(8,75$ mg/g) y $250 + 2(8,75$ mg/g), o sea, el 95% de las muestras contiene entre 232,5 e 267,5 mg/g de proteoglicanos.

Algunas propiedades importantes de la desviación estándar:

- Nunca puede ser un número negativo (debido a la forma en que se calcula y porque mide una distancia, y las distancias nunca son números negativos).
- El menor valor posible es cero y eso sucede solamente en situaciones planificadas, donde cada número del conjunto de datos es igual. No hay variación.
- Sufre la influencia de valores discrepantes dentro del conjunto (o muy altos o muy bajos). Recuerde que esto se debe a que la desviación estándar se basa en la distancia de los datos en relación al promedio.
- La desviación estándar tiene la misma unidad de los datos originales.

Sin la desviación estándar no es posible saber si los datos se encuentran cercanos al promedio o muy esparcidos. Por ejemplo, si el reembolso de una cirugía que dure 1 hora es, en promedio, \$7.000,00, usted podrá pensar: ¡Oh! ¡Qué maravilla! pero y si la desviación estándar para los reembolsos de cirugía fuera de \$ 2.000,00 y utilizáramos la regla empírica de distribución normal, el reembolso de esa misma cirugía podría ser de un valor entre \$ 3.000,00 y \$ 11.000,00 (o sea, \$ 7.000,00 mayor o menor que las desviaciones estándar. Por lo tanto, podemos observar que existe una variabilidad muy grande en ese valor de reembolso y solamente el promedio informado no refleja tan bien el valor que se puede esperar de ese procedimiento.

Coeficiente de Varianza

Por último, vamos a entender el coeficiente de varianza. Cuando comparamos la variabilidad de una medida entre dos grupos, la desviación estándar se puede emplear adecuadamente. No obstante, si los grupos tienen promedios muy discrepantes, la comparación directa de los desvíos estándar podrá inducir a conclusiones equivocadas, principalmente cuando se espera que el grupo con mayor promedio tenga mayor variabilidad.

Por ejemplo, si comparamos hipotéticamente, la altura de vértebras de adultos que, supongamos, tienen una desviación estándar de 5,9 mm, con la altura de vértebras de recién nacidos, que tienen una desviación estándar de 4,0 mm. La desviación estándar es mayor en

los adultos, lo que sugiere que su variabilidad es superior a la de los recién nacidos. Sin embargo, los adultos tienen vértebras mucho más grandes, supongamos que, de alrededor de 50 mm, mientras que los recién nacidos tienen vértebras mucho menores, supongamos que de alrededor de 25 mm. Siendo así, será mejor comparar la variación de altura de las vértebras en relación al promedio. Así, se puede emplear el coeficiente de varianza que es, simplemente, la desviación estándar expresado como un porcentaje del promedio.

Entonces, la fórmula del coeficiente de varianza (CV) es:

$$CV = \frac{\text{Desviación estándar} \times 100}{\text{Promedio}}$$

El CV es independiente de la unidad de medida y se expresa en porcentaje. En este ejemplo, el CV de los adultos es del 11,8% y el CV de los recién nacidos es del 16%. Lo que demuestra que la altura de la vértebra del recién nacido es más variable que la de un adulto, aunque menos variable en términos absolutos.

En suma, el estudio de los datos descriptivos de una muestra tiene que evaluar de manera conjunta las medidas de tendencia central y las medidas de dispersión. Es eso lo que hace posible un análisis real de los datos.

LECTURA COMPLEMENTARIA

1. Blair, RC; Taylor, RA. Bioestatística para ciências da saúde. Pearson Education do Brasil, 2013.
2. Daly, LE; Bourke, GJ. Interpretação e aplicações da estatística em medicina. Instituto Piaget. 2000.
3. Griffiths, D. Use a cabeça! Estatística. Alta Books Editora. Rio de Janeiro. 2008.
4. Levine, DM; Berenson, ML; Stephan, D. Estatística: Teoria e Aplicações usando Microsoft Excel em Português. Rio de Janeiro: LTC, 2000.
5. Rumsey, DJ. Estatística para leigos. Alta Books Editora. Rio de Janeiro. 2012.
6. Triola, MF. Introdução à Estatística. 7ª Ed. Rio de Janeiro: LTC, 1999.

CAPÍTULO 11

PRESENTACIÓN DE LOS DATOS

José María Jiménez Avila

La presentación de los datos es una de las partes del manuscrito más utilizadas en la estadística, ya que a partir de diferentes tipos de gráficos se puede evidenciar y analizar los resultados de una manera rápida.

Un cuadro o una figura bien diseñada, muestra con orden, el mayor número de ideas en la menor cantidad de espacio, es importante que imite el formato de cuadros o figuras de la revista que haya elegido para su próxima publicación (títulos, líneas, colores, variables y números decimales), también numere los cuadros y figuras de manera consecutiva y redacte el título siendo breve, informativo y preciso, algo que no puede olvidar es la elaboración de pies de cuadros y figuras con notas y abreviaturas.

Al realizar el análisis de los resultados, un detalle que no se debe olvidar es el de mantener el objetivo de la investigación, por lo que debe seleccionar los datos que por sí solos reflejen el resultado del “mensaje que se quiere dar”, que sea fácil de entender y que permita comprender la relación que guardan los datos entre sí.

Deberá incluir:

- a) Título.
- b) Cuerpo de la figura.
- c) Notas explicativas.

Regularmente el número total recomendado para la utilización de cuadros y figuras en los artículos científicos es de 6, por ejemplo 2 y 4, 3 y 3, etc.

No sea reiterativo con la información que presenta esto es, no repita en el texto lo que se describe en los cuadros y/o las figuras.

La calidad de la presentación de los resultados es de mucha importancia para la adecuada comprensión e interpretación de los resultados que se obtienen de la investigación.

Tipos de Representación Gráfica

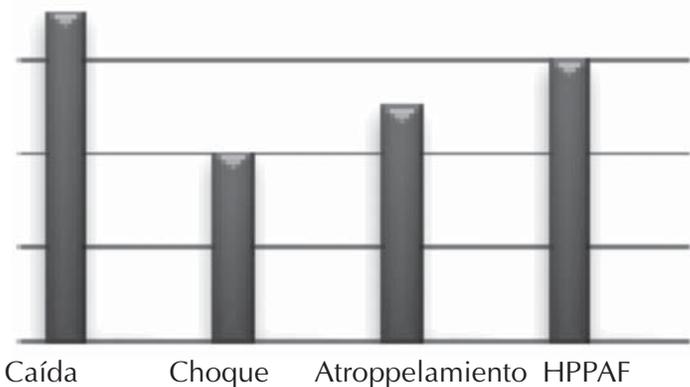
Diagrama de barras

Este tipo de representación gráfica es muy útil para evidenciar variables de tipo cuantitativas discretas y cualitativas en escala nominal.

Se sugiere que las barras tengan la misma anchura y estas deben estar separadas por un espacio, el cual no debe ser mayor que el espesor de las barras.

La sugerencia, es que se presenten de mayor a menor a reserva que exista una secuencia predeterminada que deba asignarse de esa manera.

Ejemplo:



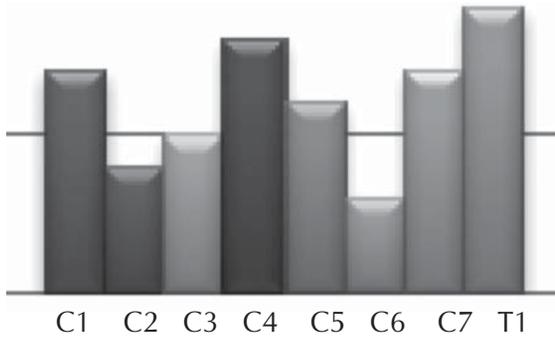
Aplicación en la cirugía:

Distribución de 60 pacientes que han sufrido fractura de columna lumbar según el mecanismo de la lesión. Jalisco, 2014-2015.

Histograma

Este tipo es adecuado para variables cuantitativas continuas, esto porque puede agrupar las frecuencias de presentación de las variables en intervalos

Ejemplo:



Aplicación en la cirugía

Distribución de 100 pacientes con fractura cervical y torácica alta. Jalisco. 2014-2015.

Diagrama de pizza

Este tipo de representación es adecuado para variables cualitativas en escala nominal.

Ejemplo:



Aplicación en la cirugía

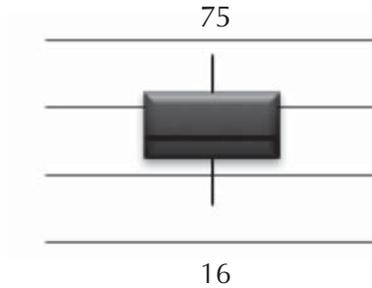
Distribución de 60 pacientes con fractura de columna según su ubicación anatómica. Jalisco. 2014-2015.

Nota: Tener cuidado cuando se utilice este tipo de gráfica, ya que puede mostrar evidencia muy obvia, que en algún momento podría omitirse dentro del manuscrito, cuando no forma parte del mensaje principal.

Diagrama de caja

Este es adecuado para variables cuantitativas en las que queremos demostrar la desviación de la distribución normal, regularmente el extremo superior de la caja corresponde al denominado percentil 75, el extremo inferior al percentil 25 y la línea horizontal que va a dividir a la caja corresponderá a la media o promedio.

Ejemplo:



Aplicación en la cirugía

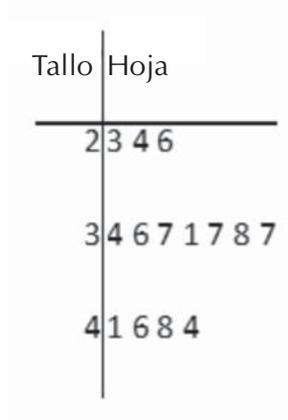
Distribución de las edades en un grupo de 60 pacientes con fractura de columna. Jalisco. 2014-2015,

Diagrama de tallos y hojas

Este es de utilidad para variables cuantitativas que tienen pocos valores, esto es debido a que en este tipo de presentación cada uno de los valores de las variables representadas se agrupan en relación a su distribución, puede tener similitud con el histograma.

Los valores aleatorios se ordenan de mayor a menor y la hoja de cada columna corresponde al último valor de cada lectura y se anota, cuantas veces se repite, formando el equivalente de la columna del histograma y el tallo corresponde al resto de la lectura y sólo se escribe una vez para cada región.

Ejemplo:



Aplicación en la cirugía

Distribución de tipo de fractura (subescala) en un grupo de 14 pacientes. Jalisco, 2014-2015.

Regularmente la primera representación debe incluir los datos demográficos del grupo de estudio, donde se observe de manera objetiva las características generales, esto es el análisis descriptivo de la muestra.

LECTURA COMPLEMENTARIA

1. Camel F: Estadística Médica y Planificación de la Salud. Tomo I. 1ra edición Universidad de los Andes. Consejo de Publicaciones. Mérida, Venezuela, 1991.
2. Gómez Gómez M, Danglot Banck C, Velásquez Jones L. Sinopsis de pruebas estadísticas no paramétricas. Cuando usarlas. Revista Mexicana de Pediatría. Vol. 70 No.2, 2003.
3. Talavera J. O, Rivas Ruiz R. Pertinencia de la prueba estadística seleccionada. Revista Médica del Instituto Mexicano del Seguro Social. Vol. 49 No. 2. 2011.

CAPÍTULO 12

PRUEBA DE HIPÓTESIS (IC) Y VALOR DE “p”

José María Jiménez Avila

Descripción de la prueba

Una prueba de hipótesis se define como una prueba estadística que puede utilizarse para determinar si existe una evidencia suficiente en una **muestra** de datos para hacer “inferencia”, que cierta condición pueda ser válida para toda la **población**.

Esto es, una prueba puede examinar 2 hipótesis opuestas: la hipótesis nula (H_0) y la hipótesis alterna (H_a). La primera es la que se probará, por lo general la hipótesis nula establece un enunciado de que “no hay efecto” o “no hay diferencia”, y la hipótesis alterna es el enunciado que se desea poder concluir que es verdadero, “hay diferencias”.

En virtud de los datos de la muestra, la prueba puede determinar si se debe rechazar la hipótesis nula y para tomar la decisión se utiliza un valor denominado “p”, si este es $<$ (menor) que o igual al nivel de significancia, que es un punto de corte que uno define, entonces puede rechazar la hipótesis nula.

Un error común de percepción, es que las pruebas estadísticas de hipótesis se diseñan para seleccionar que es lo más probable de 2 hipótesis, esto significa que una prueba mantendrá validez de la hipótesis nula, hasta que exista suficiente evidencia (resultados) a favor de la hipótesis alterna.

Ejemplo:

La cirugía por mínima invasión tiene mejores resultados que una cirugía convencional abierta

Para ello se puede establecer mediante 2 tipos de pruebas:

- a) Test de Hipótesis.
- b) Estimación (Intervalo de confianza).

Prueba de Hipótesis

Una vez que se formula una hipótesis donde se quiere comparar a 2 poblaciones y no es posible estudiar al total del universo, se toma una muestra de cada grupo a partir de inferir los parámetros de ambos grupos, esto con el objeto de que la información sea suficiente para tomar la decisión correcta sobre la "igualdad" o la "diferencia" de los 2 grupos.

Para ello se deben de asumir varios aspectos:

1. Ambas poblaciones son idénticas y no hay diferencias entre los grupos.
2. Selección del nivel de significancia, mediante una prueba estadística que puede ser el valor de "p".
 - a) $p > 0.05$: Se acepta la hipótesis nula. Esto es, "no hay diferencia", es igual de efectiva (Efecto 0).
 - b) $p < 0.05$: Se rechaza la hipótesis nula, esto es "hay diferencias", y se define como la probabilidad de que las diferencias sean significativas y no sean debidas al azar (Efecto 1).

Notas de interés

Cuando se asume la hipótesis nula, la reflexión es que no hay más diferencias que las debidas al azar.

La probabilidad implica incertidumbre, por más pequeño que sea el valor de "p", nunca podemos estar seguros de contar con la verdad absoluta, haber obtenido una $p < 0.0000001$, solamente nos indica que la probabilidad de que H_0 sea cierta es mínimo, pero jamás será nula.

Desde el punto de vista médico quirúrgico, el que una prueba sea "estadísticamente significativo, no es sinónimo de clínicamente importante".

Estimación (Intervalo de confianza)

Se encuentra asociado a lo que se conoce como intervalo de confianza (IC), el cual es el intervalo en el que sabemos que se encuentra un parámetro, con un nivel de confianza específico.

Se traduce como la probabilidad de que el parámetro a estimar se encuentre en el intervalo de confianza (más utilizados 95% ó 99%).

Se interpreta por ejemplo si se tuviera una población, en el que el promedio de edad se encontrara entre 40.2 y 53.6 y seleccionamos 100 muestras; 95 cubrirán el valor verdadero de la media de la población y 5 no lo harán.

Esto es, se establece una estimación de aproximación del parámetro correspondiente a la población de la cual se extrajo la muestra.

Tipos de error

Existen ciertos riesgos que pueden provocar algunos errores, al realizar e identificar las inferencias, los cuales se denominan: error tipo I y tipo II.

- a. Error de tipo I: Rechazar la hipótesis nula cuando la hipótesis nula es verdadera: La probabilidad de cometer este tipo de error es justamente α . Por eso, a este error se lo denomina error α .
- b. Error de tipo II: Aceptar la hipótesis nula cuando la hipótesis nula es falsa. Conocido como error de aceptación o error β :

Ejemplos:

1. El acusado era inocente y el jurado lo declaro inocente: OK esta todo bien: buen trabajo: el tipo sale libre y no hay problema.
2. El acusado era culpable (el la mató!!), y el jurado (muy astutos ellos!!), lo declaran culpable: cadena perpetua: y todo OK (merecido se lo tenía!!).

Pero el jurado puede cometer dos tipos de errores:

3. Que declare al acusado culpable y el pobre tipo era inocente!!!: Error de tipo I.
4. Que el jurado declare al acusado inocente y el muy desgraciado era culpable: Error de tipo II.

1. Si el valor de "p" nos da que la probabilidad de haber encontrado ese valor por azar es muy baja, rechazaremos la hipótesis nula: diremos que la diferencia es estadísticamente significativa: Si nos equivocamos estaremos cometiendo un error de tipo I.

2. Si el valor de p nos da alto: diremos que la probabilidad de haber encontrado ese valor por azar es alta. Por lo tanto, aceptaremos la hipótesis nula: diremos que las diferencias

encontradas no eran estadísticamente significativas: si nos equivocamos (y SI había diferencias): estaremos cometiendo un error de tipo II.

El valor de “p” y el nivel de significancia

Es la probabilidad de que el azar por si solo pueda producir una diferencia entre los grupos comparados.

Si el valor de “p” es grande ($p > 0.10$ ó $p > 0.25$), la diferencia observada entre las incidencias bien puede deberse al azar y por lo tanto se considerará a la asociación como “no significativa”.

Si el valor de “p” es pequeño ($p < 0.05$ ó $p < 0.01$) la diferencia observada tiene una probabilidad muy pequeña de deberse al azar y casi seguramente se debe a una diferencia real entre las incidencias de las poblaciones en estudio.

Ejemplo en la cirugía de columna

Usted desea determinar si la cirugía por mínima invasión tiene un efecto beneficioso sobre la cirugía convencional abierta, en los pacientes que sufren de dolor lumbar provocado por una hernia de disco.

1º Formular la pregunta de investigación (PICOT).

2º Establecer una postura en base a la posible comparación de las 2 opciones.

2º Transformar la pregunta en 2 hipótesis.

Hipótesis nula: H_0 : Los resultados son iguales.

Hipótesis alterna: H_a : Los resultados son diferentes.

3º Obtener el valor de “p” mediante pruebas estadísticas.

Este arrojará 2 posibilidades $p > 0.05$.

$p < 0.05$.

Interpretación

La manera más común es comparar el valor de “p” con el nivel de significancia (α). α es la probabilidad de rechazar la H_0 , cuando la H_0 es verdadera.

El valor de "p", indica el grado en que la evidencia de la muestra apoya el rechazo de H_0 , por lo general, mientras más pequeño sea el valor de "p", mayor influencia tendrá la evidencia de la muestra para rechazar H_0 .

Más específicamente, el valor de "p", es el menor valor de α que conduce al rechazo de H_0 .

Para cualquier valor de $\alpha >$ al valor "p", no puede rechazar H_0 y para cualquier valor de $\alpha \leq$ al valor de "p", rechaza H_0 .

En el ejemplo de cirugía por mínima invasión, en donde se trata de comparar algunos aspectos relacionados con la mejoría en comparación con la cirugía convencional abierta, en el caso hipotético que en promedio 25 pacientes presentaran mejores resultados, el estadístico de prueba dependerá de la media y si el valor de "p" fuera por ejemplo .026, indica que el 2.6% de las muestras con mejoría en el resultado extraídas de la población donde $\mu = 25$, producirá una media que proporciona una evidencia tan (o más) fuerte que la muestra actual de que μ no es igual a 25, ahora pregúntese qué es más probable: ¿que $\mu = 25$ y usted simplemente seleccionó una muestra muy poco común o que μ no es igual a 25?

Tradicionalmente, el valor de "p" se compara con valores de α menores que .05 o .01, dependiendo del campo de estudio.

En el ejemplo, supongamos que el valor de α es .05. El valor de "p" de .026, esto indica que la media (no solo la media de los demás pacientes incluidos en el estudio) probablemente no sea igual a 25. Una forma más correcta de señalar esto en términos estadísticos es: "a un nivel de significancia de .05, la media de mejoría por este tipo de cirugía parece ser significativamente diferente de 25".

Esto se interpreta de la siguiente manera:

"Mejoría de resultados con el uso de la cirugía mínima invasiva" -
- - $p < 0.05$.

El uso de la cirugía mínima invasiva tiene una probabilidad mayor de obtener mejores resultados que la cirugía convencional abierta.

Para lo cual se puede asumir que pudiera existir una asociación que no se debe al azar y que probablemente genere el resultado mencionado.

Usar los valores de "p" resulta sencillo si se conocen dos datos clave: los valores de α que son aceptables en su campo y las hipótesis nula y alterna para las pruebas que se usarán.

LECTURA COMPLEMENTARIA

1. Celis de la Rosa A. Bioestadística. 1ª ed en español. Manual Moderno. 2004.
2. Daniel WW: Estadística con aplicaciones a las Ciencias Sociales y a la Educación. 1ª ed en español. McGraw-Hill, México. 1988.
3. Gómez Gómez M, Danglot Banck C, Velásquez Jones L. Sinopsis de pruebas estadísticas no paramétricas. Cuando usarlas. Revista Mexicana de Pediatría. Vol. 70 No.2 2003.
4. Talavera J. O, Rivas Ruiz R. Pertinencia de la prueba estadística seleccionada. Revista Médica del Instituto Mexicano del Seguro Social. Vol. 49 No. 2. 2011.

CAPÍTULO 13

TEST DE KOLMOGOROV-SMIRNOV

Nelson Astur Neto
Maximiliano Gimenez Gigón
Rodrigo Góes Medéa de Mendonça

Descripción

El test de Kolmogorov-Smirnov (KS) tiene por objetivo, determinar si dos conjuntos de datos difieren significativamente.

El test de KS tiene la ventaja de no hacer suposiciones sobre la distribución de los datos. Técnicamente es un test no-paramétrico y sin distribución. Por otro lado, su desventaja es que otros testes pueden ser más sensibles si los datos contemplan los requerimientos del test.

Dentro de la estadística se aplican en la investigación los testes o pruebas paramétricos y no paramétricos. Entre los testes no paramétricos que comúnmente se utilizan para verificar si una distribución se ajusta o no a una distribución esperada, en particular a la distribución normal se encuentran el test de Kolmogorov-Smirnov. El test de Kolmogorov-Smirnov es bastante potente con muestras grandes. El nivel de medición de la variable y su distribución son elementos que intervienen en la selección del test que se utilizará en el procesamiento posterior. De hecho, si la variable es continua con distribución normal, se podrán aplicar técnicas paramétricas. Si es una variable discreta o continua no normal, solo son aplicables técnicas no paramétricas pues aplicar las primeras arrojaría resultados de dudosa validez.

Formula

Hipótesis a contrastar:

H_0 : Los datos analizados siguen una distribución M .

H_1 : Los datos analizados no siguen una distribución M .

Comparar la distribución empírica de la muestra con la distribución propuesta bajo H_0

Si esta comparación revela diferencias significativas, se rechaza H_0 .

$$D = \sup_{1 \leq i \leq n} |\hat{F}_n(x_i) - F_0(x_i)|$$

Estadístico de contraste:

Donde:

- Es el i -ésimo valor observado en la muestra (cuyos valores se han ordenado previamente de menor a mayor).
- Es un estimador de la probabilidad de observar valores menores o iguales que x_i .
- Es la probabilidad de observar valores menores o iguales que x_i cuando H_0 es cierta.

Así pues, D es la mayor diferencia absoluta observada entre la frecuencia acumulada observada $F_n(x)$ y la frecuencia acumulada teórica $F_0(x)$, obtenida a partir de la distribución de probabilidad que se especifica como hipótesis nula.

Si los valores observados $F_n(x)$ son similares a los esperados $F_0(x)$, el valor de D será pequeño.

Cuanto mayor sea la discrepancia entre la distribución empírica $F_n(x)$ y la distribución teórica, mayor será el valor de D .

Por tanto, el criterio para la toma de la decisión entre las dos hipótesis será de la forma₃:

Si $D \leq D_\alpha \Rightarrow$ Acptar H_0
Si $D > D_\alpha \Rightarrow$ Rechazar H_0

Aplicación

Aplicaciones del test:

Contrastar si un conjunto de datos muestrales pueden considerarse procedentes de una distribución determinada.

Alternativa al test χ^2 cuando el modelo propuesto bajo la hipótesis nula es de tipo continuo y el tamaño muestral es pequeño.

Ventajas del test Kolmogorov-Smirnov frente al test χ^2 :

No requiere la agrupación de los datos en clases

Es aplicable a muestras pequeñas.

Inconvenientes del test Kolmogorov-Smirnov frente al test χ^2 :

Solo es válido para modelos de tipo continuo.

Desarrollo del test:

Sea X_1, X_2, \dots, X_n una m.a.s. de una v.a. X con distribución de tipo continuo.

Contraste:

H_0 : X sigue la distribución F

H_1 : X no sigue la distribución F

Realizar un test de Kolmogorov – Smirnov, a nivel $\alpha = 0.1$, para contrastar si puede suponerse que los 10 datos:

10.5, 8, 15, 12.1, 12.1, 4.1, 12.1, 8, 10.5, 16

proceden de una distribución normal $N(10.84, 3.5)$

1. Ordenados los datos de la muestra, construimos la tabla con los valores D_i

$x_{(i)}$	$\hat{F}_n(x_{(i)})$	$F(x_{(i)})$	D_i
4.1	0.1	0.027	0.073
8	0.3	0.209	0.109
10.5	0.5	0.641	0.161
12.1	0.8	0.640	0.160
15	0.9	0.882	0.082
16	1	0.930	0.070

$\max\{0.027, |0.1 - 0.027|\}$
 $\max\left\{\begin{matrix} |0.1 - 0.209| \\ |0.3 - 0.209| \end{matrix}\right\}$

2. $D_{exp} = \max\{D_i, i = 1, 2, \dots, n\} = 0.161$

3. Región Crítica, $C = [D_{1-\alpha}, +\infty[= [0.368, +\infty[$

4. Conclusión: $0.161 < 0.368$, por tanto, **no se rechaza** que los datos procedan de una distribución $N(10.84; 3.5)$

Aplicación en la cirugía de columna

En un trabajo reciente, en 2009, publicado pela revista *Spine*, pudimos verificar la utilización de dicho test KS.

Se estudió la variación entre la forma anatómica de las facetas articulares y sus variaciones durante el crecimiento de niños eutróficos dinamarqueses.

El autor, Masharawi y sus colaboradores de la Universidad de Tel Aviv en conjunto con la Universidad del Sur de Dinamarca, realizó mediciones de todas las facetas lumbares (L1-S1) de 100 niños (51 varones y 49 mujeres) utilizando resonancia magnética, tanto en largo como ancho de las mismas.

Tales medidas fueron tomadas a los 13 años y 3 años después, con software indicado para tales casos.

Más de 4400 medidas fueron hechas. Se desprende de esto, la importancia de un test que pueda verificar si estamos delante de una muestra normal. Aquí quedaría patente, la importancia del test KS.

En dicho trabajo, se concluyó que las facetas de aquellos niños continuaban creciendo luego de los 12 años. Cualquier asimetría debería considerarse fuera de la normalidad y pone de manifiesto la necesidad de fabricación de implantes más personalizados, teniendo en cuenta las franjas etarias para esa población.

LECTURA COMPLEMENTARIA

1. Daniel WW. *Biostatistics. A foundation for analysis in the health sciences*. 7th ed. New York: John Wiley and Sons Inc, 1999: 658-736.
2. Kolmogorov AN. Sulla determinazione empirical di una legge di distribuzione. *Giornale Inst Ital Altuari* 1933; 4: 83-91.
3. Masharawi YM, Kjaer P, Bendix T, Manniche C, May H, Mirovsky Y, Anekshtein Y, Jensen TS, Hershkovitz I. Lumbar facet and interfacet shape variation during growth in children from the general population: a three-year follow-up MRI study. *Spine (Phila Pa 1976)*. 2009 Feb 15;34(4):408-412.
4. Smirnov NV. Estimate of deviation between empirical distribution functions in two independent samples. *Bull Moscow University* 1939; 2: 3-16.

5. Siegel S, Castellan NJ. Estadística no paramétrica aplicada a las ciencias de la conducta. 4ª ed. México: Editorial Trillas, 1995: 151-157.
6. Song-Hee Kim and Ward Whitt, 2013. The Power of Alternative Kolmogorov-Smirnov Tests Based on Transformations of the Data. ACM Trans. Model. Comput. Simul. V, N, Article A (January YYYY), 18 pages. DOI:<http://dx.doi.org/10.1145/0000000.0000000>.

PRUEBA DE SHAPIRO-WILK

Nelson Astur Neto

Rodrigo Góes Medéa de Mendonça

Definición de la prueba

La prueba de Shapiro-Wilk es una prueba de normalidad en las estadísticas frecuentistas. Lo publicaron, en 1965, Samuel Shapiro Sanford y Martin Wilk. La distribución Normal es una de las más importantes distribuciones de probabilidades de la estadística, conocida también como Distribución de Gauss o Gaussiana. La prueba rechaza la hipótesis de normalidad cuando el valor de p es menor o igual a 0,05.

Para verificar la Normalidad por medio de la prueba de *Shapiro-Wilk (S-W)*, existen varias herramientas y programas estadísticos, entre los cuales el *Statistical Package for Social Sciences (SPSS)* versión 23.0 (SPSS, 2016). El SPSS se utiliza para análisis de datos y permite manipular, transformar, crear tablas y gráficas que resuman la información obtenida. Sus potencialidades van más allá del análisis de procedimientos más avanzados como inferencia estadística, pruebas de hipótesis y estadísticas multivariadas para datos cualitativos y cuantitativos.

Fórmula

La prueba Shapiro-Wilk, propuesta en 1965, se basa en la estadística W dada por:

$$W = \frac{\left(\sum_{i=1}^n a_i x_{(i)}\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

donde $an-i+1$ son constantes generadas por los promedios, variancias y covariancias de las estadísticas de orden de una muestra de tamaño n de una distribución Normal.

Ejemplos en la cirugía de columna

Como ejemplo de la prueba de Shapiro-Wilks aplicada a la columna, tenemos el trabajo desarrollado en la Universidad de Pensilvania en conjunto con los grupos de ortopedia y radiología, además del departamento de estadística y publicado en 2014 por la Academia Americana de Fisiatría y Rehabilitación. Billy y colaboradores estudiaron la posibilidad de que existieran diferencias entre las alturas y los diámetros de los discos lumbares, antes y después de periodos prolongados en la posición sentada y sin intervalos, durante cuatro horas consecutivas. El diseño del estudio fue prospectivo, en pacientes ambulatoriales, utilizando Resonancia Magnética para las medidas. El primer día, los pacientes se analizaron al inicio y al final del día de trabajo, y tras pasar 4 horas sentados. El segundo día, los mismos 12 pacientes se estudiaron también al comienzo de sus actividades, pero se sometieron a alteraciones y protocolo de estiramiento a cada 15 minutos hasta el final de las mismas 4h.

En este ejemplo, la prueba de Shapiro-Wilks fue decisiva para saber si la muestra se encontraba en la normalidad, pues siendo normal, el próximo paso sería realizar una Prueba T pareada, si no, se utilizarían las pruebas de Wilcoxon y Rank.

Aún acerca de los resultados de este trabajo, se encontró una diferencia estadística entre las alturas de los discos en aquellos que permanecieron sentados todo el tiempo. La aplicabilidad clínica de esta investigación radica en el hecho de que pequeñas alteraciones en la altura discal pueden estar correlacionadas a la mejoría de la lumbalgia y, por consiguiente, de sus consecuencias.

Aplicación en cirugía

Cuando escuchamos el nombre específico de alguna prueba estadística, como Shapiro-Wilk, es difícil imaginar cómo podríamos utilizarlos a nuestro favor, aún más cuando hacemos investigaciones con poblaciones sometidas a alguna forma de intervención, como cirugía de columna.

Como ejemplo ilustrativo de esa aplicación, tenemos la investigación realizada por Farjoodi et al en el Hospital Johns Hopkins,

en los Estados Unidos, donde el efecto del volumen quirúrgico del hospital y del cirujano se evaluó en las complicaciones tras cirugía en la columna lumbar. El artículo se publicó en el periódico Spine en 2011.

En ese caso, la función de la prueba de Shapiro-Wilks fue la de verificar la normalidad de las variables, y separar aquellas que necesitaban ajustes para no volverse factores de confusión, como las variables demográficas. Se estudiaron 232.668 procedimientos hospitalarios listados como descompresión posterolateral con artrodesis y/o exploración/descompresión del canal vertebral.

Tras la aplicación de esa herramienta estadística, los autores concluyeron que la mortalidad y la tasa de complicaciones asociadas a la cirugía de columna lumbar fueron menores cuando los pacientes fueron tratados por cirujanos y hospitales con alto volumen.

Interpretación

La prueba de *S-W* ofrece el parámetro valor de prueba (valor-*p*, *p*-value o significancia), que se puede interpretar como la medida del grado de acuerdo entre los datos y la hipótesis nula (H_0), correspondiendo H_0 a la distribución Normal. Cuanto menor el valor-*p*, menor será la consistencia entre los datos y la hipótesis nula. Por lo tanto, la regla de decisión adoptada para saber si la distribución es Normal o no, es rechazar H_0 : (i) si H_0 se rechaza, se constata que el conjunto de datos en discusión no puede tener distribución Normal; si no se rechaza H_0 , la distribución Normal es una distribución posible para el conjunto de datos en discusión.

Si falla, la prueba de normalidad permite afirmar con un 95% de confianza que los datos no se encajan en la distribución normal. Pasar la prueba de normalidad solamente permite declarar que no se encontró desvío significativo de la normalidad.

LECTURA COMPLEMENTARIA

1. Shapiro, S. S.; Wilk, M. B. (1965). "An analysis of variance test for normality (complete samples)". *Biometrika* 52 (3-4): 591-611. p. 593.
2. Lopes M. M, Castelo Branco V. T. F. Soares J. B. Utilização dos testes estatísticos de Kolmogorov- Smirnov e Shapiro Wilk para verificação da normalidade de pavimentação. TRANSPORTES v. 21, n. 1 (2013) p. 59-66.
3. <http://www.portalaction.com.br/inferencia/64-prueba-de-Shapiro-Wilk>.
4. Billy GG, Lemieux SK, Chow MX. Changes in lumbar disk morphology associated with prolonged sitting assessed by magnetic resonance imaging. *PM R*. 2014 Sep;6(9):790-5.
5. Farjoodi P, Skolasky RL, Riley LH. The effects of hospital and surgeon volume on postoperative complications after LumbarSpine surgery. *Spine (Phila Pa 1976)*. 2011 Nov 15;36(24):2069-75.
6. <<http://www.variation.com/da/help/hs141.htm>>.

Unidad 2

CAPÍTULO 15

Prueba t de Student

Barón Zarate

Francisco López Meléndez

CAPÍTULO 16

Prueba de ANOVA

Juan Pablo Guyot

CAPÍTULO 17

Correlación de Pearson

Alisson R. Teles

Asdrubal Falavigna

CAPÍTULO 15

PRUEBA t DE STUDENT

Barón Zarate
Francisco López Meléndez

Lo primero que tendríamos que explicar para este tema es que existen dos grandes grupos de pruebas estadísticas, que se basan en la distribución de la muestra o de la población, que son pruebas paramétricas y no paramétricas; la diferencia entre pruebas PARAMÉTRICAS y NO PARAMÉTRICAS, es que las pruebas paramétricas asumen los parámetros de la distribución de una variable (media y varianza), y su distribución es normal; y las pruebas no paramétricas **NO** asumen nada acerca de la distribución, ni se preocupa por el tipo de distribución de la muestra, solo trabaja con el ordenamiento y recuento de los valores de la variable sin importar la distribución (por lo que también se llaman de distribución libre).

Características de la muestra para saber qué tipo de prueba se puede realizar.

Paramétricas

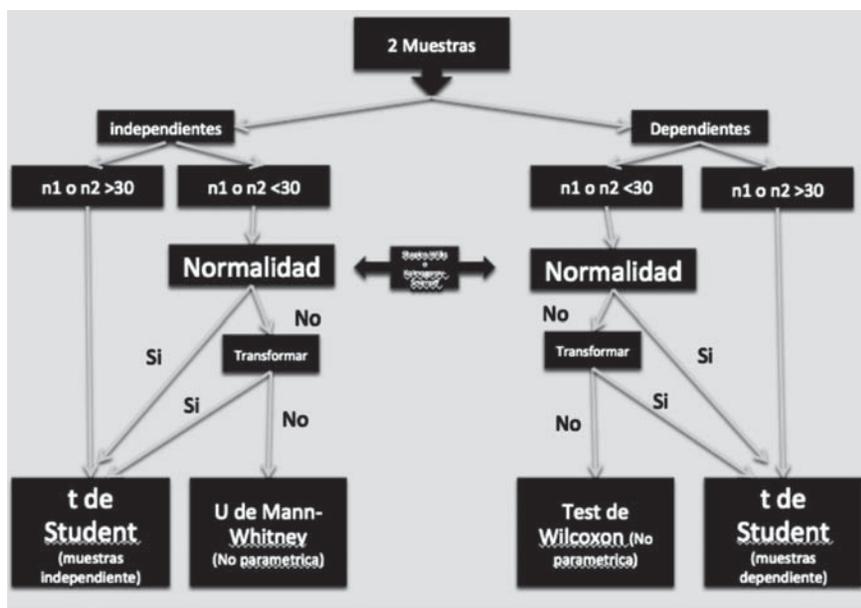
- Las variables tienen que ser cuantitativas
- Usar escalas de intervalo o razón
- Las variables deben de seguir una distribución normal
- Las varianzas deben de ser iguales
- Muestras grandes ($n > 30$)

No paramétricas

- Variables de distribución libre (no normal)
- Si son variables cuantitativas, ordinales o nominales
- Grupos de varianzas grandes
- Muestras pequeñas ($n < 30$ o < 10)

En bioestadística una de las cuestiones más frecuentes es valorar la significancia estadística de las diferencias entre muestras de individuos, sean de un mismo grupo (ej. *Antes y después*), o de distintos grupos (ej. *Tratamiento 1 Vs Tratamiento 2*), y a esto le llamaremos muestras emparejadas o de muestras independientes respectivamente. (Figura 1).

Figura 1 – Diagrama de flujo para decidir el tipo de prueba estadística para hacer contraste de hipótesis en dos muestras (dependientes o independientes)



Student es el pseudónimo de William Sealy Gossett, quien fue un estadista y matemático inglés que trabajaba para la cervecería Guinness, ideó la prueba de t o t de Student en 1908 para valorar la distribución de muestras pequeñas y esta basada en el teorema de pequeñas muestras y en el teorema del límite central.

La prueba de Student o t de Student generalmente se utiliza para la comparación de muestras continuas distribuidas de forma normal (prueba paramétrica) entre dos grupos o entre grupos; por lo que existen dos tipos de t de Student, uno para muestras independientes y otro para muestras dependientes o emparejadas.

En esta ocasión hablaremos de la *t* de Student para muestras emparejadas (**muestras relacionadas**). Se dice que hay emparejamiento cuando:

- Las mediciones son hechas en un mismo individuo (ej. Antes y después de una intervención) y se llama autoemparejamiento.
- Se comparan entre si dos parejas naturales (ej. Gemelos) y es emparejamiento natural.
- Se busca un control para cada caso, lo que se llama Casos y controles, y en estos se buscan con características iguales (sexo, edad, etc.) y se llama emparejamiento artificial.

NOTA: La selección de una prueba estadística para contrastar hipótesis es diferente cuando se trata de muestras independientes (ej. Dos tipos de individuos) y de muestras emparejadas (ej. Mismo individuos observados dos veces), debido a que la variabilidad aleatoria entre individuos (inter individual) no existe o es menor en las muestras relacionadas (variabilidad intra individual), lo mismo pasa si es de distribución normal y no normal.

La formula de *t* de Student es la siguiente:

$$T_{n-1} = \mu \text{ DIF} / \text{EEMDIF}$$

$$t = \frac{\bar{X}_D - \mu_0}{s_D / \sqrt{n}}$$

A continuación, se desarrollará el con un ejemplo la aplicación practica de una prueba de *t* de Student para datos relacionados.

Interpretación en la cirugía de columna

Un ejemplo cotidiano en donde se pudiera utilizar una prueba *t* de Student para muestras relacionadas, seria evaluar la calidad de vida (en una escala continua validada de 0 a 100, en donde mas alto es el valor mejor calidad de vida), en un numero *n* pacientes que se sometieron a un procedimiento quirúrgico (por ejemplo: una artrodesis L4-L5), en donde se realiza una medición preoperatoria y otra medición post operatoria, para posterior evaluar si existe cambio postquirúrgico.

(objetivo: evaluar si existe un cambio estadísticamente significativo en la calidad de vida posterior al tratamiento quirúrgico). (Tabla 1).

Tabla 1

Paciente	Preoperatorio (SF-36)	Postoperatorio (SF-36)
1	43.2	56.3
2	46.2	46.4
3	38.6	65.3
4	34	43
5	36.4	39.2
6	16.9	49.5
7	43	56
8	18	45.3
9	36.4	62.1
10	49.1	63.2
11	38.6	49
12	43.2	45
13	42	55
14	38.6	56

Nuestra hipótesis nula (H_0) sería que no existe diferencia significativa entre la medición preoperatoria y la medición postoperatoria.

$$H_0 \cong \mu \text{ antes} = \mu \text{ después}$$

Y nuestra hipótesis alterna (H_1) sería que si existe diferencia estadísticamente significativa entre ambas mediciones.

$$H_1 \cong \mu \text{ antes} \neq \mu \text{ después}$$

Pasos a seguir para realizar una prueba t Student para muestras relacionadas.

1. Lo primero que realizaremos será crear una nueva variable (**DIF**), que es la diferencia entre el valor de "antes" y "después", y esta nueva variable nos dice cuanto ha cambiado un individuo entre las dos mediciones, después hay que calcular la media (*m*) de la variable "DIF". (**Tabla 2**).

Tabla 2

Paciente	Preoperatorio	Postoperatorio	"DIF"
1	42	50	8,00
2	37.3	52.3	15,00
3	26.4	38	11,60
4	29.7	48	18,30
5	24.9	49.1	24,20
6	44.6	55	10,40
7	33.5	49	15,50
8	44.6	67	22,40
9	26.5	39	12,50
10	18	56	38,00
11	33.4	47	13,60
12	42	65	23,00
13	37.5	57	19,50
14	24.9	67	42,10
			MEDIA:19.58

2. Si la muestra es menor de 30, debemos comprobar si la nueva variable sigue una *distribución normal*, en caso de no tener distribución normal, se puede transformar de manera logarítmica, o hacer una prueba no paramétrica (prueba de Wilcoxon). En el programa SPSS se realizan las dos pruebas de normalidad Kolmogorov-Smirnov y Shapiro-Wilks.

NOTA: La prueba de Shapiro-Wilks está considerada como la prueba más potente para muestra inferiores a 30 casos.

3. El siguiente paso después de calcular la media de la variable (*DIF*), hay que calcular el error estándar de la misma (**EEMDIF**= error estándar de la media de la variable "diferencia").

La fórmula para calcular el EEMDIF es dividiendo la desviación estándar de la población (σ) entre la raíz cuadrada del tamaño de la muestra (n).

$$\mathbf{EEMDIF} = \sigma / \sqrt{n}$$

EJEMPLO:

$$\mathbf{EEMDIF} = \sigma / \sqrt{n}$$

$$= 9.98 / \sqrt{14}$$

$$= 9.98 / 3.74$$

$$\mathbf{EEMDIF} = 2.66$$

4. Después hay que calcular los grados de libertad (**GL**= $n-1$) en donde n es el número total de individuos de la muestra y 1 es constante para prueba t de Student de muestras relacionadas.

$$\mathbf{GL} = n-1$$

$$\mathbf{GL} = 14-1 = 13$$

5. Calcular la t de Student para muestras relacionadas; esta es el resultado que se obtiene de dividir la media (μ) de la variable *DIF* entre su error estándar (**EEMDIF**).

$$\mathbf{T}_{n-1} = \mu \mathbf{DIF} / \mathbf{EEMDIF}$$

EJEMPLO:

$$t_{13} = m \mathbf{DIF} / \mathbf{EEMDIF} = \mu \mathbf{DIF} / (\sigma / \sqrt{n})$$

$$t_{13} = 19.58 / 2.66$$

$$\mathbf{t}_{13} = 7.36$$

6. El siguiente paso es comparar y buscar en las tablas el valor obtenido de t y sus grados de libertad, para obtener el valor de p ; también se puede utilizar **EXCEL** en lugar de ver las tablas, con la siguiente fórmula:

$$= \mathbf{DISTR.T.2} (\text{valor de } t; \mathbf{GL}; 2)$$

el primer número será el valor obtenido de la t , el segundo son los grados de libertad, y el tercero puede ser 2, que es una prueba a dos colas o 1 si es una prueba a una cola.

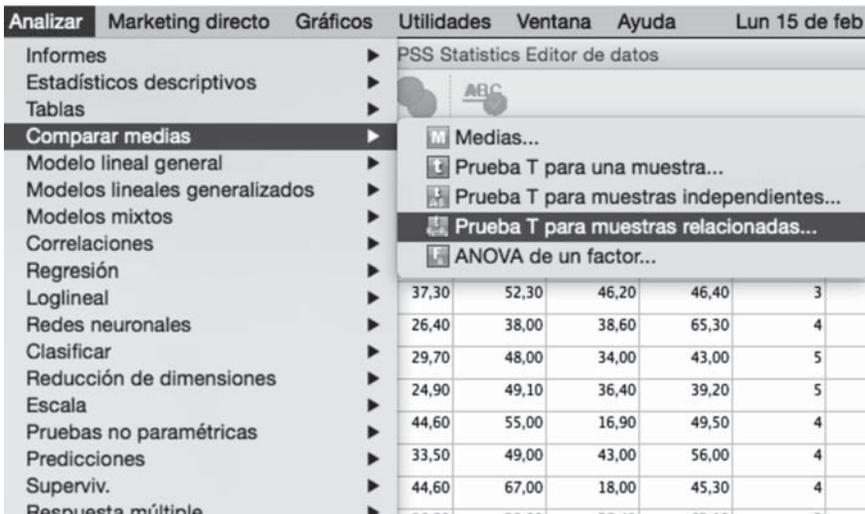
7. Siempre se debe de acompañar el resultado de p obtenido, junto con el intervalo de confianza para la diferencia de medias.

Resultado del ejemplo: en este ejemplo obtenemos una t de 7.36, con 13 GL, lo que equivale a una $p < 0.0000$, con un IC de 95% a dos colas. Lo que significa que existe evidencia suficiente para rechazar la hipótesis nula; es decir que, si existe un aumento significativo en la escala de calidad de vida en los pacientes pos operados, y que hubo un aumento promedio de 19.58 puntos en la escala, con una confianza de 95% que el aumento estará entre 13.8 y 25.3 puntos.

Ejemplo y forma de hacerlo con SPSS

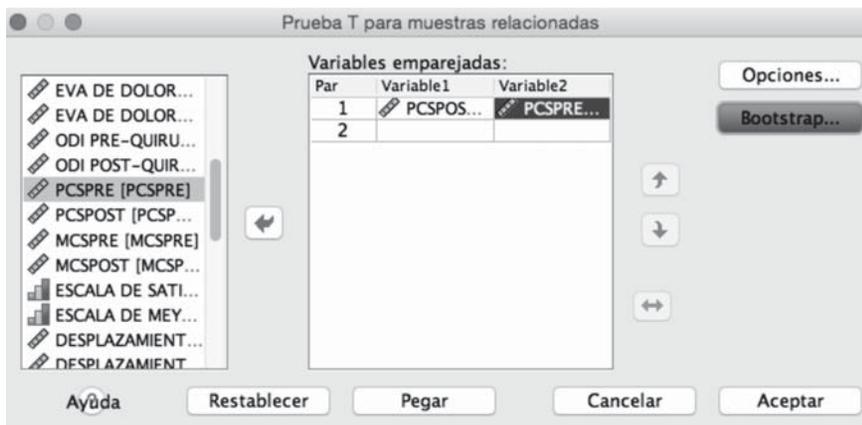
1. Lo primero que tenemos que hacer en el SPSS es tener nuestras variables a contrastar ordenadas, después se ira al menú "analizar" \Rightarrow "comparar medias" \Rightarrow "prueba de T para muestras relacionadas..." \Downarrow !. (Figura 2)

Figura 2 – Prueba T Muestras relacionadas



2. Aparecerá un recuadro en donde aparece introducir las dos variables a contrastar, en este caso introduciremos primero la variable "después y posterior la variable "antes" \Rightarrow aceptar \Downarrow !. (Figura 3).

Figura 3 – Prueba T Muestras relacionadas



3. Resultados: El ordenador nos dará tres tablas (Figura 4); la primera tabla son los estadísticos descriptivos de las variables a contrastar, la segunda tabla nos da un coeficiente de correlación (*Pearson*), entre las variables; y la mas importante para interpretar es la **tercera tabla**, que nos muestra en la primera columna la media de la "diferencia" (el cambio de hubo de antes y después); en la segunda columna la desviación estándar de la media de la diferencia, y posterior el error estándar de la media (**EEMDIF**), la siguiente columna el intervalo de confianza (IC) con el limite inferior y el limite superior, lo que sigue es el valor de la **t**, después los grados de libertad y por ultimo la significancia estadística.

Figura 4 – Tablas de resultados de SPSS. (V. 19)

Estadísticos de muestras relacionadas				
Par	Media	N	Desviación tip.	Error tip. de la media
Par 1 PCSPOST	52,8143	14	9,14506	2,44412
PCSPRE	33,2357	14	8,42347	2,25127

Correlaciones de muestras relacionadas			
Par	N	Correlación	Sig.
Par 1 PCSPOST y PCSPRE	14	,357	,210

Prueba de muestras relacionadas								
	Diferencias relacionadas				t	gl	Sig. (bilateral)	
	Media	Desviación tip.	Error tip. de la media	95% intervalo de confianza para la diferencia				
				Inferior				Superior
Par 1 PCSPOST - PCSPRE	19,57857	9,97938	2,66710	13,81665	25,34049	7,341	13	,000

NOTA: se utilizaron los mismos datos para ambos ejemplos: la forma manual y con SPSS. La decisión de rechazar la hipótesis nula es cuando $p = < 0.05$.

Tabla 3 – Tabla para ver la t de Student

Tabla de la t de Student.
 Contiene los valores t tales que $P(|T| > t) = \alpha$,
 donde n son los grados de libertad.



n \ α	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,1584	0,3249	0,5095	1,0000	1,9626	3,0777	6,3137	12,7062	31,8210	63,6559	636,5776
2	0,1421	0,2887	0,4447	0,8165	1,3862	1,8856	2,9200	4,3027	6,9645	9,9250	31,5998
3	0,1366	0,2767	0,4242	0,7649	1,2498	1,6377	2,3534	3,1824	4,5407	5,8408	12,9244
4	0,1338	0,2707	0,4142	0,7407	1,1896	1,5332	2,1318	2,7765	3,7469	4,6041	8,6101
5	0,1322	0,2672	0,4082	0,7267	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321	6,8685
6	0,1311	0,2648	0,4043	0,7176	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074	5,9587
7	0,1303	0,2632	0,4015	0,7111	1,1192	1,4149	1,8946	2,3646	2,9979	3,4995	5,4081
8	0,1297	0,2619	0,3995	0,7064	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554	5,0414
9	0,1293	0,2610	0,3979	0,7027	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498	4,7609
10	0,1289	0,2602	0,3966	0,6998	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693	4,5868
11	0,1286	0,2596	0,3956	0,6974	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058	4,4359
12	0,1283	0,2590	0,3947	0,6955	1,0832	1,3562	1,7823	2,1788	2,6810	3,0545	4,3178
13	0,1281	0,2586	0,3940	0,6938	1,0795	1,3502	1,7709	2,1604	2,6503	3,0123	4,2209
14	0,1280	0,2582	0,3933	0,6924	1,0763	1,3450	1,7613	2,1448	2,6245	2,9768	4,1403
15	0,1278	0,2579	0,3928	0,6912	1,0735	1,3406	1,7531	2,1315	2,6025	2,9467	4,0728
16	0,1277	0,2576	0,3923	0,6901	1,0711	1,3368	1,7459	2,1199	2,5835	2,9208	4,0149
17	0,1276	0,2573	0,3919	0,6892	1,0690	1,3334	1,7396	2,1098	2,5669	2,8982	3,9651
18	0,1274	0,2571	0,3915	0,6884	1,0672	1,3304	1,7341	2,1009	2,5524	2,8784	3,9217
19	0,1274	0,2569	0,3912	0,6876	1,0655	1,3277	1,7291	2,0930	2,5395	2,8609	3,8833
20	0,1273	0,2567	0,3909	0,6870	1,0640	1,3253	1,7247	2,0860	2,5280	2,8453	3,8486
21	0,1272	0,2566	0,3906	0,6864	1,0627	1,3232	1,7207	2,0796	2,5176	2,8314	3,8193
22	0,1271	0,2564	0,3904	0,6858	1,0614	1,3212	1,7171	2,0739	2,5083	2,8188	3,7922
23	0,1271	0,2563	0,3902	0,6853	1,0603	1,3195	1,7139	2,0687	2,4999	2,8073	3,7676
24	0,1270	0,2562	0,3900	0,6848	1,0593	1,3178	1,7109	2,0639	2,4922	2,7970	3,7454
25	0,1269	0,2561	0,3898	0,6844	1,0584	1,3163	1,7081	2,0595	2,4851	2,7874	3,7251
26	0,1269	0,2560	0,3896	0,6840	1,0575	1,3150	1,7056	2,0555	2,4788	2,7787	3,7067
27	0,1268	0,2559	0,3894	0,6837	1,0567	1,3137	1,7033	2,0518	2,4727	2,7707	3,6895
28	0,1268	0,2558	0,3893	0,6834	1,0560	1,3125	1,7011	2,0484	2,4671	2,7633	3,6739
29	0,1268	0,2557	0,3892	0,6830	1,0553	1,3114	1,6991	2,0452	2,4620	2,7564	3,6595
30	0,1267	0,2556	0,3890	0,6828	1,0547	1,3104	1,6973	2,0423	2,4573	2,7500	3,6460
40	0,1265	0,2550	0,3881	0,6807	1,0500	1,3031	1,6839	2,0211	2,4233	2,7045	3,5510
80	0,1261	0,2542	0,3867	0,6776	1,0432	1,2922	1,6641	1,9901	2,3739	2,6387	3,4164
120	0,1259	0,2539	0,3862	0,6765	1,0409	1,2886	1,6576	1,9799	2,3578	2,6174	3,3734
∞	0,1258	0,2533	0,3855	0,674	1,036	1,282	1,645	1,96	2,326	2,576	3,291

LECTURA COMPLEMENTARIA

1. Altman DG, Bland JM. Statistics Notes: Transforming data. BMJ 1996; 312: 770.
2. Altman DG. Practical statistics for medical research. Londres: Chapman and Hall, 1991.
3. Altman DG, Bland JM. Detecting skewness from summary information, BMJ 1996; 313:1200.
4. Altman DG, Bland JM. Statistics Notes: The use of transformations when comparing two means. BMJ 1996; 312:1153.
5. Altman DG. Preparing to analyse data. En: Practical statistics for medical research. London: Chapman and Hall; 1991. p.132-145.
6. Berry G., Armitage P. Statistical Methods in Medical Research. 3rd. ed. Oxford: Blackwell Science; 1994.
7. Braitman LE. Confidence intervals asses both clinical significance and statistical significance. Ann Intern Med 1991; 114 (6): 515-517.

8. Gehlbach SH. Interpreting the medical literature: practical epidemiology for clinicians. 5th ed. USA: McGraw-Hill. Medical; 2006; p. 166.
9. Livingston EH. Who was student and why do we care so much about his t-test. *J Surg Research* 2004;118:58.
10. Moreno V, Vallescar R, Martín M. Las pruebas no paramétricas en el análisis estadístico de datos. *Aten Primaria* 1991; 8 (1): 58-60.
11. Moser CJ, Stevens MB, Homogeneity of variance in the two-sample means test, *The American Statistician* 46 (1992), 19-21.
12. Newcombe, RG. Confidence intervals for an effect size measure based on the Mann-Whitney statistic. Part 2: Asymptotic methods and evaluation. *Statistics in Medicine*, (2006) 25, 559-573.
13. Rajiv Gandhi, Holly N. Smith, Nizar N. Mahomed, Randy Rizek, MD, Mohit Bhandari. Incorrect Use of the Student t Test in Randomized Trials of Bilateral Hip and Knee Arthroplasty Patients. *The Journal of Arthroplasty* Vol. 26 N. 5. 2011.
14. Riegelman, RK. *Studying a Study and Testing a Test: How to Read the Medical Evidence*. Philadelphia: Lippincott Williams & Wilkins, 1999.

CAPÍTULO 16

PRUEBA DE ANOVA

Juan Pablo Guyot

La prueba de ANOVA es una prueba estadística utilizada para analizar 2 varianzas, de ahí el término proveniente del inglés (**AN**alysis **Of** **VA**riance). Su diseño inicial se remonta al año 1925 al aplicarlo sobre trabajos agropecuarios, sin embargo su difusión generó su empleo en muchos campos diferentes. El objetivo inicial consistía en observar la influencia de diferentes factores sobre el crecimiento de una determinada semilla que de no mediar distintos métodos de cultivo crecerían todas de la misma manera. En otras palabras intenta evaluar la importancia de uno o mas factores observando el impacto que tiene sobre una o mas poblaciones determinadas con varianzas constantes.

En el diseño experimental de una prueba ANOVA se intenta identificar “factores fijos” y “factores randomizados”. Se habla de los primeros cuando intentamos estudiar un grupo que de no presentar exposición se presentarían inalterados y nuestro interés pasa por analizar la respuesta a diferentes factores específicos. Por otra parte, se define a la prueba con factores randomizados a la elección al azar de factores con diferentes niveles de exposición con un número de niveles múltiples tendiendo al infinito. En este caso, de repetir las pruebas los factores al azar deben surgir de la elección sobre distintos niveles de elección y no siguiendo siempre un mismo patrón.

Si lo graficáramos con un ejemplo estaríamos definiendo a factores fijos en una comparación de medias entre diferentes poblaciones mientras que los randomizados se tratarían del análisis de diferencias al azar entre la misma población. Un ejemplo práctico de factores randomizados en la cirugía de columna estaría orientado en el estudio de la evolución del balance sagital de la población latinoamericana. El investigador tomará factores al azar pero siempre basándose en una distribución simétrica: rango de edades, diferencias por razas, etc.

Existen pruebas de ANOVA de una vía y de 2 vías en relación con los factores a analizar. Se determina ANOVA de una vía o de un factor al análisis que involucra el estudio sobre una sola variable mientras que se define como ANOVA de doble vía o doble factor cuando las variables involucradas provienen de 2 varianzas diferentes.

ANOVA de una sola vía

Se intenta determinar la influencia que presenta el resultado de una variable Y relacionada directamente por los distintos factores X (X1, X2, X3.....). Se utiliza cuando uno intenta analizar la varianza de varias variables independientes o varios niveles de una sola variable independiente, manteniendo un mismo grado de confianza.

Al diseñar un modelo experimental que utilice la prueba de ANOVA de una vía debemos descartar que se cumplan 4 variables:

- 1- Los resultados de las variables estudiadas deben ser absolutamente independientes. Es decir, el valor de una observación no debe influir en el resultado de otra variable.
- 2- Los datos se registran siguiendo un modelo aditivo incorporando toda la información tanto los efectos fijos como los errores al azar.
- 3- Se asume que la distribución de los errores sigue un patrón ordenado.
- 4- Los errores al azar siguen una varianza homogénea.

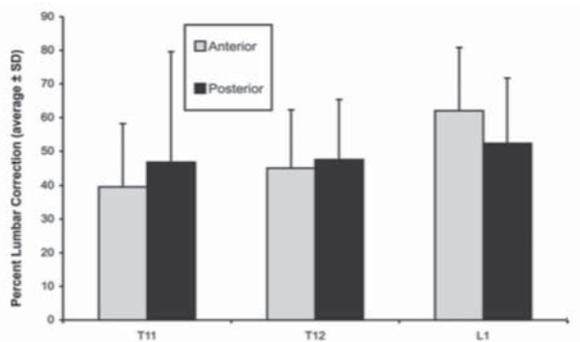
Un ejemplo de este tipo de prueba sería el estudio de diferentes tipos de tratamientos para pacientes con estenosis degenerativa del canal lumbar. En este caso la población a estudiar sería el factor fijo. Los pacientes con una estenosis del conducto lumbar se espera que evolucionen de una manera similar y estudiaremos el impacto que tiene diferentes procedimientos (factores) sobre su seguimiento. Se estudiarán la respuesta según los distintos tratamientos aplicados (descompresión simple, descompresión + fusión por vía posterior y descompresión indirecta por dispositivos intersomáticos colocados por vía lateral) utilizando el estudio de ANOVA al comparar diferentes aspectos de los mismos (tiempo quirúrgico, sangrado, corrección de deformidad, Oswestry pre y postoperatorio).

ANOVA de doble vía

El diseño de este tipo de experimento involucra el estudio de 2 o mas variables fijas (Y y Z) en relación con diferentes grados de exposición a los factores X (X1, X2, X3.....). La inclusión de otro factor no es inocente ya que este puede generar algunos problemas asociados ya que ambos factores pueden ser fijos, aleatorios o uno fijo y otro aleatorio. Al mismo tiempo, puede pasar que dichos factores actúen de manera independiente o que los mismo interactúen entre si potenciando o inhibiendo su potencia final. Por último, el diseño puede determinar 2 grupos de factores que sean equilibrados o no.

Interpretación

Un ejemplo de este tipo de diseño experimental vinculado a la cirugía de columna sería la evaluación del grado de corrección espontánea de las curvaturas lumbares en instrumentaciones selectivas de escoliosis torácicas tipo Lenke 1 comparando diferentes vías de abordaje.



En este caso, la variable a estudiar sería la corrección espontánea de la columna lumbar. Sin embargo, a diferencia del ejemplo presentado en el caso de 1 vía en este caso tenemos 2 variables fijas independientes entre sí (vía anterior vs vía posterior). Por último estudiaremos que influencia tiene sobre cada uno de los grupos la vértebra caudal instrumentada (T11, T12 o L1).

LECTURA COMPLEMENTARIA

1. Tejada Barreras M. Estadística Básica. En: Falavigna A, Jimenez Avila JM. Educación en investigación AOSPine Latin America: de la idea a la publicación. Caxias do Sul, RS: Educs. 2014, p. 149-165.
2. Larson MG. Analysis of variance. *Circulation*. 2008 Jan 1;117(1):115-21.
3. Dangelmajer S, Zadnik PL, Rodriguez ST, Gokaslan ZL, Sciubba DM. Minimally invasive spine surgery for adult degenerative lumbar scoliosis. *Neurosurg Focus*. 2014 May; 36(5):E7.
4. Sit V. Analyzing ANOVA Designs: Biometrics Information Handbook No. 5. Province of British Columbia, Ministry of Forests Research Program. Working paper 07/1995. Available at: <<http://www.for.gov.bc.ca/hfd/pubs/docs/Wp/Wp07.pdf>>. Accessed September 1, 2016.
5. Patel PN, Upasani VV, Bastrom TP, Marks MC, Pawelek JB, Betz RR, Lenke LG, Newton PO. Spontaneous lumbar curve correction in selective thoracic fusions of idiopathic scoliosis: a comparison of anterior and posterior approaches. *Spine(Phila Pa 1976)*. 2008 May 1;33(10):1068-73.

CORRELACIÓN DE PEARSON

*Alisson R. Teles
Asdrubal Falavigna*

Descripción de la prueba

Correlación es una medida del grado de dependencia lineal entre dos variables o entre una medida de intensidad de la asociación de esas variables. En otras palabras, la prueba de correlación evalúa la existencia y la fuerza de asociación entre dos variables cuantitativas. Por ejemplo, ¿existe correlación entre el grado de desequilibrio sagital (medido en centímetros) y la incapacidad funcional de los pacientes con escoliosis degenerativa (medida por la escala de incapacidad Oswestry)?, otro ejemplo sería la correlación entre la incidencia pélvica y la lordosis lumbar.

Se dice que dos variables están correlacionadas (varían juntas) cuando se logra demostrar que existe asociación lineal entre esas variables.

Existen básicamente dos formas de verificar correlación. La más sencilla es a través de una gráfica denominada diagrama de dispersión, también conocido como diagrama de puntos (Figura 1). Por convención, se denomina x la variable independiente o predictora, que se describe en el eje horizontal, e y la variable dependiente o explicada, descrita en el eje vertical. La otra manera de evaluar correlación es por un coeficiente. El coeficiente es superior al diagrama de dispersión, pues representa un valor absoluto, lo que facilita la interpretación de la correlación entre dos variables de medidas diferentes (ej.: desequilibrio sagital, verificado en centímetros, e índice de incapacidad Oswestry, medido en la escala entre 0 y 100 puntos).

El coeficiente de correlación de Pearson es la medida más conocida de correlación entre dos variables cuantitativas. Se debe aplicar para evaluar la correlación de variables continuas (no categóricas, por lo menos con medición intervalar) que presenten distribución normal (distribución gaussiana). O sea, el coeficiente de correlación de Pearson

no se debe utilizar para evaluar la correlación de variables asimétricas que no tengan distribución normal. En ese caso, se debe aplicar la prueba de correlación de Spearman.

Es importante recordar que el coeficiente de correlación mide una asociación, no una relación de causa y efecto. Por ejemplo, en un estudio con pacientes con enfermedades degenerativas lumbares se identificó que existe correlación entre pérdida sanguínea intraoperatoria y calidad de vida medido por el componente físico del SF36 1 año tras la cirugía. En este ejemplo, la correlación estadísticamente significativa no indica que la pérdida sanguínea sea la causa de peor calidad de vida 1 año tras el procedimiento. Otro concepto es que, en muestras grandes, aun coeficientes de correlación bajos pueden ser estadísticamente significativos. Eso corrobora la importancia de la evaluación del grado de correlación (ej: $n = 1300$, $r = 0,15$, $p = 0,001$, correlación débil).

Fórmula

La fórmula de cálculo del coeficiente de correlación (r) fue propuesta por Karl Pearson en 1896. El coeficiente de correlación es un valor absoluto, independiente de las unidades usadas para medir las variables x e y .

$$r = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}}$$

Actualmente, distintos softwares de análisis de datos generan el coeficiente de correlación, nivel de significancia de la correlación, grado de la correlación y coeficiente de determinación (Figura 2).

Interpretación

La interpretación del coeficiente de correlación (r) implica básicamente 4 pasos (Tabla 1). En primer lugar, se verifica si existe realmente la correlación a través de la prueba de hipótesis. El razonamiento de la prueba de hipótesis es básicamente el mismo de las comparaciones de promedios. Cuando se calcula la r en una muestra, extraída de una población particular, es necesario tener en mente que se está estimando la asociación verdadera entre las variables

x e y en la población, por lo tanto, es necesario verificar si la correlación identificada existe realmente o si está relacionada a un error de muestreo. A través de la prueba de hipótesis, se obtiene el valor p , que indica la probabilidad de que la correlación realmente exista (en términos estadísticos, aceptar la hipótesis alternativa). Generalmente, se acepta la existencia de correlación verdadera cuando $p < 0.05$; o sea: hay menos que un 5% de chance de que la correlación identificada no exista realmente (aceptar hipótesis nula, rechazar la hipótesis alternativa), por lo tanto, se acepta que ella es verdadera.

El segundo paso es determinar la fuerza de la correlación entre las variables. Usualmente, la correlación con coeficiente entre 0 – 0,3 se considera débil, entre 0,3 – 0,6 regular, 0,6 – 0,9 fuerte), 0,9 – 1 muy fuerte, y 1 correlación perfecta.

El tercer paso es analizar la señal de la correlación. El coeficiente de correlación (r) puede presentar una variación entre -1 y +1. Valores negativos expresan una correlación indirecta, esto es, cuando x aumenta, y , en promedio, disminuye. Valores positivos indican correlación directa, cuando x e y varían en el mismo sentido. Por ejemplo, en individuos normales, los valores de incidencia pélvica presentan correlación directa con la lordosis lumbar. En contrapartida, cuánto mayor la edad, menor es la lordosis lumbar (correlación indirecta).

Otra manera de evaluar el valor de la medida de asociación lineal entre dos variables es por el coeficiente de determinación (r^2). Elevando el coeficiente de correlación de Pearson al cuadrado, se obtiene el coeficiente de determinación, que representa la fracción de la variabilidad que es compartida entre dos variables. En otras palabras, representa el porcentaje de variación explicada por una de las variables en relación a la otra. El valor r^2 puede variar entre 0 y 1. Por ejemplo, si el coeficiente de correlación de Pearson (r) es de 0,734, el coeficiente de determinación (r^2) es de $0,734 \times 0,734 = 0,539$. O sea, el 53,9% de la variación de y se explica por la variable x . El coeficiente de determinación no se puede obtener a través del coeficiente de correlación de Spearman (pruebas no paramétricas), solamente a través del coeficiente de correlación de Pearson.

Ejemplo en el área de cirugía de columna

La correlación de Pearson es una de las pruebas más utilizadas en investigación médica. Para su adecuada aplicación, sin embargo, es necesario seguir los requisitos básicos: las variables deben ser continuas

(o, por lo menos, con distribución intervalar), y deben presentar distribución normal (gaussiana).

Uno de los ejemplos en cirugía de columna es la correlación entre el eje sagital vertical cervical, medido a través de la distancia entre la línea de plomada del punto promedio de C2 y C7. Varios estudios han demostrado que existe correlación negativa entre el eje sagital cervical y calidad de vida en el posoperatorio de pacientes que se someten a 3 o más niveles de laminectomía cervical y fusión posterior. Tang et al. demostraron que el coeficiente de correlación entre el eje sagital vertical C2-C7 y el Índice de Incapacidad Cervical (NDI - *Neck Disability Index*) es de 0.2957 ($p = 0.024$). El coeficiente de correlación entre el eje y el componente físico del SF36 (PC-SF36) -0.4262 ($p = 0.001$). En este ejemplo, se verifica que (1) existe correlación entre la medida radiográfica y los instrumentos de evaluación NDI y PC-SF36 (ambas pruebas de hipótesis demuestran $p < 0.05$); la correlación entre eje C2-C7 con NDI es débil ($r = 0.2957$) y con PC-SF36 es regular ($r = -0.4262$); (3) la correlación es positiva entre el eje cervical y el NDI (cuanto más grande la distancia entre el eje sagital vertical cervical, mayor es la incapacidad medida por el NDI) y negativa entre el eje y el PC-SF36 (cuanto más grande el eje, menor es la calidad de vida medida por el PC-SF36).

Figuras y tabla

Figura 1 – Ejemplos de diagramas de dispersión y los respectivos coeficientes de correlación (r)

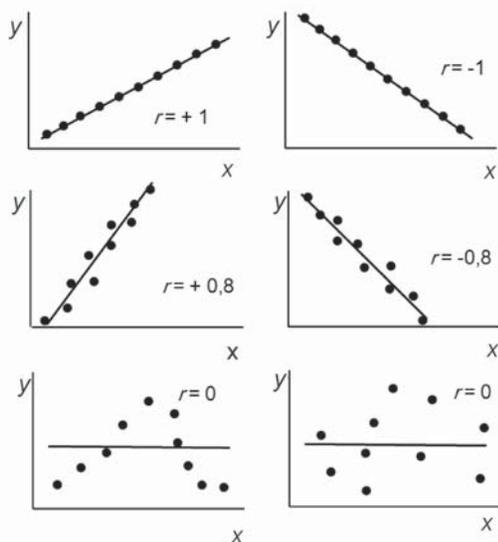
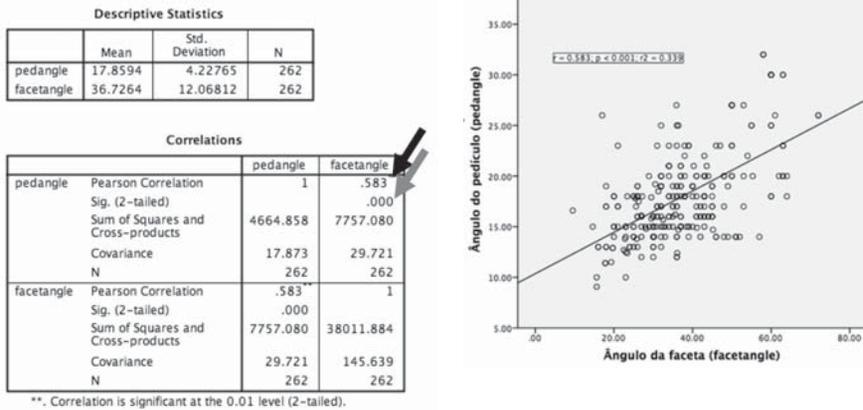


Figura 2 – Ejemplo de prueba de correlación entre dos variables usando el software SPSS



Nota: En el ejemplo arriba, se demuestra la prueba de correlación entre el ángulo del pedículo (pedangle) y el ángulo de la articulación interapofisaria (facetangle) de vértebras lumbares verificados en tomografía computadorizada.

Se observa que cuanto mayor es el ángulo pedicular en relación a la línea media, mayor es el ángulo de la articulación interapofisaria ($r = 0.583$, $p < 0.0001$). El software también genera el diagrama de dispersión entre las dos variables.

Tabla

Tabla – Interpretación de la prueba de correlación de Pearson (r)

1. ¿Existe correlación entre las variables?
 - Cuando $P < 0,05$ se considera que la correlación es estadísticamente significativa, o sea, no ha ocurrido al azar. Si $P > 0,05$ no existe correlación.

2. ¿Cuál es el grado de la correlación?

- Evaluar el valor de r (evaluación cualitativa de r cuanto a la intensidad):

0	Correlación nula
0 – 0,3	Correlación débil
0,31 – 0,6	Correlación regular
0,61 – 0,9	Correlación fuerte
0,91 – 1	Correlación muy fuerte
1	Correlación plena o perfecta

3. ¿Cuál la señal de la correlación?

- Valores positivos indican correlación directa, valores de x aumentan a medida que aumentan los valores de y , o viceversa.
- Valores negativos indican correlación indirecta, cuando los valores de x disminuyen a medida que aumentan los valores de y , o viceversa.

4. ¿Cuál es el coeficiente de determinación (r^2)?

- El r^2 indica el porcentaje de variación de x que se explica por la variable y .
-

LECTURA COMPLEMENTARIA

1. Peat J, Barton B. Medical Statistics: a guide to data analysis and critical appraisal. BMJ Books, 2008. ISBN: 978-0-7279-1812-3.
2. Peacock JL, Peacock PL. Oxford Handbook of Medical Statistics. Oxford University Press, 2011. ISBN: 978-0-19-9555128-6.
3. Tang JA, Scheer JK, Smith JS, Deviren V, Bess S, Hart RA, Lafage V, Shaffrey CI, Schwab F, Ames CP; ISSG. The impact of standing regional cervical sagittal alignment on outcomes in posterior cervical fusion surgery. *Neurosurgery*, 2015, 76 (Suppl 1): S14-21. Doi: 10.1227/01.neu.0000462074.66077.2b.
4. Teles AR. Basic statistical shortcomings can induce mistaken interpretations of neurosurgical articles. *World Neurosurgery*, 2016, 92 (577-578) Doi: 10.1016/j.wneu.2015.12.090.

Unidad 3

CAPÍTULO 18

Chi² (Ji²)

Alberto Aceves Pérez

CAPÍTULO 19

Test o prueba de rangos de Wilcoxon

Baron Zarate

CAPÍTULO 20

U de Mann-Whitney

José Manuel Pérez Atanasio

Victor Fernando Luján Celis

Maritza Belén Sandoval Rincón

CAPÍTULO 21

Prueba de Kruskal-Wallis

Esteban Araya Ramírez

Fabián Viquez Monge

CAPÍTULO 22

Coefficiente de correlación de Spearman

Fernando Alvarado

Miguel Farfán

CAPÍTULO 23

Prueba de McNemar

Esteban Araya Ramírez

Fabián Viquez Monge

CAPÍTULO 24

Regresión lineal simple

Alisson R. Teles

Asdrubal Falavigna

CAPÍTULO 25

Distribución de Poisson

José Acuña Pinilla

Margarita Gómez Chantaine

Elizabeth Heins

CAPÍTULO 26

Odds ratio

José Manuel Pérez Atanasio

Victor Fernando Luján Celis

Maritza Belén Sandoval Rincón

CAPÍTULO 18

Chi² (Ji²)

Alberto Aceves Pérez

Existen multitud de situaciones en el ámbito de la salud en el que las variables de interés, las cuales no pueden cuantificarse mediante cantidades numéricas, entre las que el investigador esté interesado en determinar posibles relaciones. Ejemplos de este tipo de variables pueden ser las complicaciones tras una intervención quirúrgica, el sexo, el nivel socio-cultural, etc. En este caso tendríamos, a lo sumo, las observaciones agrupadas en forma de frecuencia, dependiendo de las modalidades que presente cada paciente en cada una de las variables.

El objetivo de este tema es el estudio de este tipo de cuestiones en relación con las variables cualitativas. Estos son los contrastes asociados con el estadístico. En general este tipo de tests consisten en tomar una muestra y observar si hay diferencia significativa entre las frecuencias observadas y las especificadas por la ley teórica del modelo que se contrasta, también denominadas “frecuencias esperadas”.

Sin embargo, aunque éste sea el aspecto más conocido, el uso de esta prueba no se limita al estudio de variables cualitativas. Podríamos decir que existen tres aplicaciones básicas en el uso de este test, y cuyo desarrollo veremos en el transcurso de este capítulo:

Tres son los temas que abordaremos de esta manera:

1. Prueba de Bondad de Ajuste: consiste en determinar si los datos de cierta muestra corresponden a cierta distribución poblacional. En este caso es necesario que los valores de la variable en la muestra y sobre la cual queremos realizar la inferencia esté dividida en clases de ocurrencia, o equivalentemente, sea cual sea la variable de estudio, deberemos categorizar los datos asignado sus valores a diferentes clases o grupos.

2. Prueba de Homogeneidad de varias muestras cualitativas: consiste en comprobar si varias muestras de una carácter cualitativo proceden de la misma población (por ejemplo: ¿estas tres muestras de pacientes provienen de poblaciones con igual distribución de enfermedades? Es necesario que las dos variables medibles estén representadas mediante categorías con las cuales construiremos una tabla de contingencia.

3. Prueba de Independencia: consistente en comprobar si dos características cualitativas están relacionadas entre sí (por ejemplo: ¿la presencia de destrucción vertebral está relacionada a la presencia de un tumor?). Aunque conceptualmente difiere del anterior, operativamente proporciona los mismos resultados. Este tipo de contrastes se aplica cuando deseamos comparar una variable en dos situaciones o poblaciones diferentes.

Formula:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \approx \chi_{k-p-h}^2$$

Ejemplo:

Se desea saber si cierta enfermedad afecta del mismo modo a los hombres que a las mujeres. Para ello se considera una muestra de $n = 618$ individuos que padecen la enfermedad, y se observa que 341 son hombres y el resto son mujeres. ¿Qué conclusiones se obtiene de ello?

Si p es el porcentaje de hombres en la población de enfermos, podemos considerar el contraste:

$$\begin{cases} H_0 : p = 1/2 \\ H_1 : p \neq 1/2 \end{cases}$$

De la muestra obtenemos la siguiente estimación puntual del porcentaje de enfermos de sexo masculino: $p = 341/618 = 0.55178$

LECTURA COMPLEMENTARIA

1. P. Armitage, G. Berry, Estadística para la Investigación Biomédica. Doyma, Barcelona, 1992.
2. A. Martín Andrés, J.D. Luna del Castillo, Bioestadística para las Ciencias de la salud. Norma, Granada, 1994.
3. D. Peña Sánchez de Rivera, Estadística: Modelos y Métodos, Alianza Universidad Textos, Madrid, 1994.
4. J. Montanero Fernández, Manual de bioestadística. Universidad de Extremadura. Departamento de matemáticas. 2001.
5. J. S. Milton. Estadística para Biología y Ciencias de la Salud. Ed. Interamericana. McGraw-Hill.
6. E. Cobo, P. Muñoz y J.A. González.(2007) Bioestadística para no estadísticos. Ed. Elsevier/Masson.
7. Peña Sánchez de Rivera, D. (1987): "Estadística. Modelos y Métodos. Volumen 2". Alianza Editorial. Madrid.
8. J. Monge, Estadística no paramétrica. Secretaria de Educación y Universidades. www.uoc.edu. Consultado 3/9/2016.

CAPÍTULO 19

TEST O PRUEBA DE RANGOS DE WILCOXON

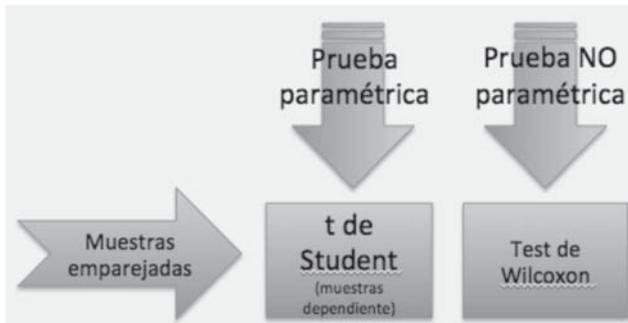
Baron Zarate

La prueba de Wilcoxon, es una prueba de contraste de hipótesis paralela (similar) a la prueba paramétrica de t de Student para muestras emparejadas, se utiliza para comparar o contrastar variables de 2 grupos relacionados que **NO** tienen una distribución normal o que sean variables **ordinales**; o cuando las variables no cumplen con los supuestos básicos de las pruebas paramétricas.

Se debe de usar de manera directa siempre en estudios con menos de 10 observaciones ($n < 10$).

En esta prueba estadística cuando la variable "DIF" no cumple los criterios de normalidad, y se trabaja con la diferencia de medianas. (primero hacer pruebas de normalidad K-S o S-W, en caso de no ser normal, se puede intentar transformarla de manera logarítmica, en caso de no lograr normalidad seguir con la prueba). (Figura 1).

Figura 1 – Pruebas de contraste de Hipótesis para muestras relacionadas



Los pasos a seguir en la prueba de Rangos de Wilcoxon, son los siguientes:

1. Arreglar las observaciones pareadas y obtener las diferencias de cada pareja.
2. Arreglar las diferencias en función de rangos como valores absolutos, sin importar el signo, pero de manera que los rangos conserven el signo correspondiente a la diferencia.
3. Obtener la sumatoria de los rangos cuyo signo es el menos frecuente, por ejemplo: si el signo es +, se considerará para efectuar sumatorias; sin embargo, la sumatoria mencionada finalmente pierde el signo.
4. Si se trata de muestras pequeñas, comparar el valor obtenido con los valores críticos de la tabla de Wilcoxon.
5. Distribuir las muestras mayores que 25 bajo la curva normal y por tanto, calcular el valor Z, en referencia al cual se debe consultar la probabilidad de diferir con respecto al promedio en la tabla de probabilidades asociadas.
6. Decidir si se acepta o rechaza la hipótesis.

Ejemplo: En este ejemplo se intenta decidir si tiene significancia estadística el cambio de la escala de mielopatía cervical de Nurick en pacientes operados de una descompresión cervical; Se utilizo una escala ordinal en del 0 al 5. Y se evaluo mediante la prueba de wilcoxon.

Nuestra hipótesis nula (H_0) es que no existe diferencia en grado mielopatía de los pacientes antes y después del tratamiento quirúrgico.

$$H_0 \cong \mu \text{ antes} = \mu \text{ después}$$

La hipótesis alterna (H_1) es que si existe una diferencia estadísticamente significativa (positiva) en la mejoría de la disfunción medular.

$$H_1 \cong \mu \text{ antes} \neq \mu \text{ después}$$

$$\text{Suma de rangos: } n(n+1) / 2$$

1. El primer paso sería sacar la diferencia de los valores del "antes y el "después", pero en este caso la n se tomaría en cuenta solo en los valores que la diferencia sea DIFERENTE de 0. (Tabla 1).

Tabla 1

Caso	Nurick pre	Nurick post	DIF	DIF. Absoluta	Orden	Rangos	Sum Postos +	Sum Postos -
1	4	4	0		0			
2	3	1	2	1	1	3	3	
3	4	1	3	1	2	3	3	
4	3	3	0	1	3	3	3	
5	1	1	0	1	4	3	3	
6	5	3	2	1	5	3	3	
7	4	3	1	2	6	7	7	
8	3	3	0	2	7	7	7	
9	3	2	1	2	8	7	7	
10	3	3	0	3	9	9	9	
11	4	2	2					
12	3	2	1					
13	3	2	1					
14	5	4	1					
15	2	2	0					
							45	0

En este caso la n sería 9. Se toma en cuenta solo los valores de la "DIF" que sea diferentes a 0.

2. El siguiente paso sería ordenar el valor absoluto (no se toma en cuenta el valor del signo +o-) de menor a mayor, posterior asignar el rango medio de cada intervalo. (Tabla 2).

Tabla 2

Caso	Nurick pre	Nurick post	DIF	DIF. Absoluta	Orden	Rangos	Sum Postos +	Sum Postos -
1	4	4	0		0			
2	3	1	2	1	1	3	3	
3	4	1	3	1	2	3	3	
4	3	3	0	1	3	3	3	
5	1	1	0	1	4	3	3	
6	5	3	2	1	5	3	3	
7	4	3	1	2	6	7	7	
8	3	3	0	2	7	7	7	
9	3	2	1	2	8	7	7	
10	3	3	0	3	9	9	9	
11	4	2	2					
12	3	2	1					
13	3	2	1					
14	5	4	1					
15	2	2	0					
							45	0

3. Ordenar y sumar los rangos positivos y negativos. (Tabla 3).

Tabla 3

Caso	Nurick pre	Nurick post	DIF	DIF. Absoluta	Orden	Rangos	Sum Postos +	Sum Postos -
1	4	4	0		0			
2	3	1	2	1	1	3	3	
3	4	1	3	1	2	3	3	
4	3	3	0	1	3	3	3	
5	1	1	0	1	4	3	3	
6	5	3	2	1	5	3	3	
7	4	3	1	2	6	7	7	
8	3	3	0	2	7	7	7	
9	3	2	1	2	8	7	7	
10	3	3	0	3	9	9	9	
11	4	2	2					
12	3	2	1					
13	3	2	1					
14	5	4	1					
15	2	2	0					
							45	0

En este caso el valor obtenido de la suma de rangos mayor (45) se compara con los valores críticos de la tabla T en pruebas de rangos señalados de pares iguales de Wilcoxon, y se puede apreciar que para ser significativo (es decir, por debajo de 0.05, que fue el nivel de significancia) debe ser menor a 0.05. (Tabla 4).

Tabla 4 – Valores críticos de Tabla de Wilcoxon

se rechaza la hipótesis nula si el mayor de las sumatorias de rangos es \geq que:

n	Con $\alpha = 0.01$	Con $\alpha = 0.05$
6		21
7		26
8	36	33
9	44	40
10	52	47
11	61	56
12	71	65
13	82	74
14	93	84
15	105	95
16	117	107
17	130	119
18	144	131
19	158	144
20	173	158
21	189	173
22	205	187
23	222	203
24	239	219

En este caso el valor es de 45 con una n de 9 por lo que el valor es menor de 0.01 por lo que se debe de rechazar la Hipótesis Nula (H_0).

NOTA: Cuando el valor de n sea igual o mayor a 25, se debe de hacer el calculo manual de Z. (Tabla 5)

$$Z = \frac{\Sigma(R+) - (n(n+1)/4)}{\sqrt{n(n+1)(n2+1)/24}}$$

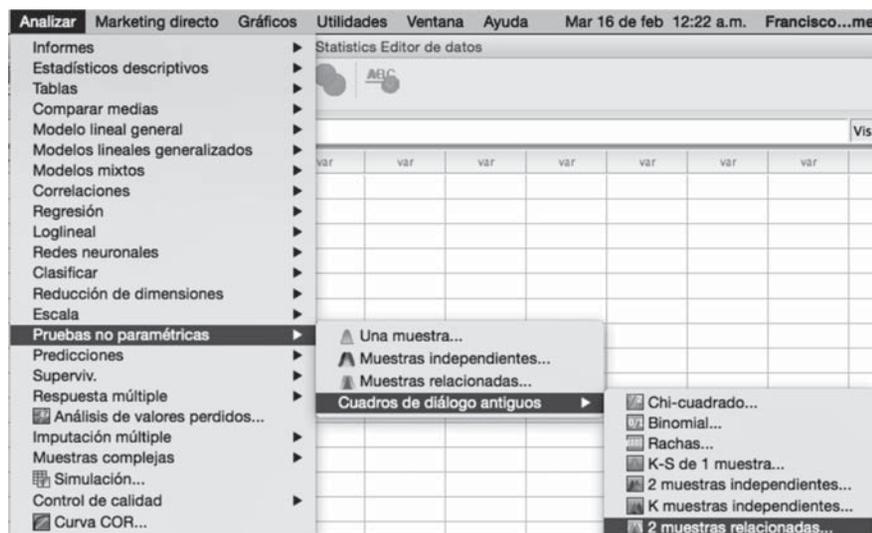
Tabla 5 – Valores críticos de Z y su significancia (a una o a dos colas).

Percentiles de la distribución normal						
	Prueba Unilateral			Prueba bilateral		
Significancia	0.05	0.01	0.001	0.05	0.01	0.001
Percentil	95	99	99.9	97.5	99.5	99.95
z	1.6449	2.3263	3.0902	1.9600	2.5758	3.2905

Para realizarlo con SPSS:

1. Tener ordenadas las variables en la introducción de datos, posterior en el menú "analizar" ⇒ "pruebas no paramétricas" ⇒ "cuadros de dialogo antiguos" ⇒ "dos muestras relacionadas" (Figura 2).

Figura 2



2. Aparecerá un cuadro de dialogo en el que hay que introducir las dos variables a contrastar, introducir la variable "antes" y posterior la variable "después", en esta prueba por automático esta seleccionada la casilla "Wilcoxon", posterior pulsar aceptar. (Figura 3).

Figura 3



3. Posterior aparecerán dos tablas. (Interpretación).

La primera nos da las variables introducidas (η) y la suma de rangos (negativos, positivos, nulos) . Y en la segunda tabla (llamada estadísticos de contraste), nos da el valor de z y la significancia a dos colas.

Figura 4

Prueba de los rangos con signo de Wilcoxon

		Rangos		
		N	Rango promedio	Suma de rangos
ESCALA DE NURICK POSTQUIRURGICA - ESCALA DE NURICK PREQUIRURGICA	Rangos negativos	9 ^a	5,00	45,00
	Rangos positivos	0 ^b	,00	,00
	Empates	6 ^c		
	Total	15		

a. ESCALA DE NURICK POSTQUIRURGICA < ESCALA DE NURICK PREQUIRURGICA

b. ESCALA DE NURICK POSTQUIRURGICA > ESCALA DE NURICK PREQUIRURGICA

c. ESCALA DE NURICK POSTQUIRURGICA = ESCALA DE NURICK PREQUIRURGICA

Estadísticos de contraste^a

	ESCALA DE NURICK POSTQUIRURGICA - ESCALA DE NURICK PREQUIRURGICA
Z	-2,724 ^b
Sig. asintót. (bilateral)	,006

a. Prueba de los rangos con signo de Wilcoxon
b. Basado en los rangos positivos.

La interpretación sería con una p de 0.006 lo que indica que hay evidencia suficiente para rechazar la hipótesis nula H_0 , ($p < 0.05$), y aceptar nuestra hipótesis alterna (H_1); que en este caso es direccional hacia la disminución de la escala de Nurick en el postoperatorio, lo que indicaría significancia estadística en la mejoría de la función con la cirugía de descompresión.

Nota: se utilizaron los mismos datos para ambos ejemplos: la forma manual y con SPSS. La decisión de rechazar la hipótesis nula es cuando $p = < 0.05$.

LECTURA COMPLEMENTARIA

1. Altman DG, Bland JM. Statistics Notes: Transforming data. BMJ 1996; 312: 770.
2. Altman DG. Practical statistics for medical research. Londres: Chapman and Hall, 1991.
3. Altman DG, Bland JM. Detecting skewness from summary information, BMJ 1996; 313:1200.
4. Altman DG, Bland JM. Statistics Notes: The use of transformations when comparing two means. BMJ 1996; 312:1153.

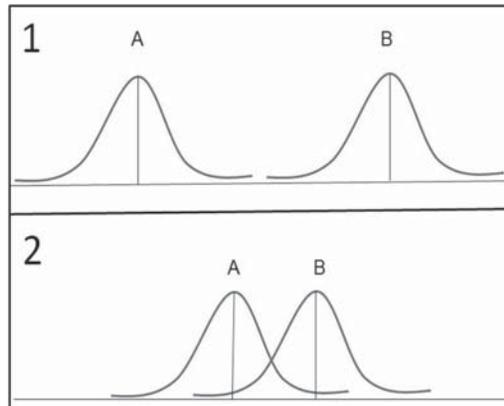
5. Altman DG. Preparing to analyse data. En: Practical statistics for medical research. London: Chapman and Hall; 1991. p.132-145.
6. Berry G., Armitage P. Statistical Methods in Medical Research. 3rd. ed. Oxford: Blackwell Science; 1994.
7. Braitman LE. Confidence intervals asses both clinical significance and statistical significance. *Ann Intern Med* 1991; 114 (6): 515-517.
8. Gehlbach SH. Interpreting the medical literature: practical epidemiology for clinicians. 5th ed. USA: McGraw-Hill. Medical; 2006; p. 166.
9. Livingston EH. Who was student and why do we care so much about his t-test. *J Surg Research* 2004;118:158.
10. Moreno V, Vallescar R, Martín M. Las pruebas no paramétricas en el análisis estadístico de datos. *Aten Primaria* 1991; 8 (1): 58-60.
11. Moser CJ, Stevens MB, Homogeneity of variance in the two-sample means test, *The American Statistician* 46 (1992) 19-21.
12. Newcombe, RG. Confidence intervals for an effect size measure based on the Mann-Whitney statistic. Part 2: Asymptotic methods and evaluation. *Statistics in Medicine*, (2006) 25, 559-573.
13. Rajiv Gandhi, Holly N. Smith, Nizar N. Mahomed, Randy Rizek, MD, Mohit Bhandari. Incorrect Use of the Student t Test in Randomized Trials of Bilateral Hip and Knee Arthroplasty Patients. *The Journal of Arthroplasty* Vol. 26 N. 5. 2011.
14. Riegelman, RK. Studying a Study and Testing a Test: How to Read the Medical Evidence. Philadelphia: Lippincott Williams & Wilkins, 1999.

U DE MANN-WHITNEY

*José Manuel Pérez Atanasio
Victor Fernando Luján Celis
Maritza Belén Sandoval Rincón*

Para poder comprender la prueba U de Mann Whitney tenemos que recordar que las pruebas de hipótesis tienen como objeto identificar si la diferencia entre los datos de dos grupos es estadísticamente significativa, lo cual lo podemos identificar gráficamente con claridad en la siguiente figura.

Figura 1 – Diferencia entre los grupos A y B. 1 la diferencia entre los grupos es estadísticamente significativa. 2 La diferencia entre los grupos no es estadísticamente significativa



Descripción de la prueba

La prueba de U de Mann-Whitney nos permite identificar si existe diferencia entre dos muestras independientes, la variable de interés debe presentar una distribución no normal. Esta prueba consiste en comparar las medianas de ambos grupos y determinar si existe una diferencia estadísticamente significativa. Es necesario hacer énfasis en que la U de Mann Whitney es específica para comparar solo dos muestras independientes si deseamos comparar más de 2 muestras se usarían otras pruebas.

Formula

$$U_1 = \frac{R_1 - n_1(n_1 + 1)}{2}$$

$$U_2 = \frac{R_2 - n_2(n_2 + 1)}{2}$$

Donde U_1 = Muestra con el menor número de pacientes

U_2 = Tamaño de muestra mayor

R_1 y R_2 = Suma de rangos para cada grupo

Ejemplo en el área de la cirugía de columna

Para entender más fácilmente esta prueba presentamos el siguiente ejemplo:

En un Centro Médico de Alta Especialidad de Columna, las patologías degenerativas de la columna cervical se tratan con dos tipos de abordajes quirúrgicos, el abordaje tipo A defendido por el Dr. Juan como el mejor, y el abordaje tipo B defendido por el Dr. José. Para poder identificar si existe diferencia estadísticamente significativa en los días que tarda en cicatrizar, planteamos la siguiente hipótesis:

- Existe una diferencia estadísticamente significativa entre el tiempo de cicatrización del abordaje A y del abordaje B.

Primero se registran en una hoja de cálculo los valores del tiempo de cicatrización en ambos grupos y se realiza el ordenamiento "rankeado" de todos los datos en orden del menor al mayor como se muestra en la siguiente tabla:

Tabla 1 – Ordenamiento de los datos

Abordaje tipo A n1=6	Orden Global (Rank)	Abordaje tipo B n2=5	Orden Global (Rank)
15	5	11	1
14	4	17	7
12	2	19	9
20	10	13	3
18	8	21	11
16	6		
Suma de los rangos	35		31

Posteriormente calculamos la mediana de cada grupo: La mediana del abordaje tipo A es: 15.5 y la mediana del abordaje tipo B es: 17, ahora sumamos los valores obtenidos en el orden global de cada uno de los grupos (suma de los rangos), Grupo A: 35 Grupo B: 31, ya con estos valores despejamos la fórmula.

$$U_1 = \frac{31 - 5(5+1)}{2} = 16$$

$$U_2 = \frac{35 - 6(6+1)}{2} = 14$$

Finalmente el menor valor de U que en este caso es 14 se coteja en una tabla de valores críticos de alfa (α) de 0.05 para la prueba U de Mann-Whitney para conocer el valor de p que en este caso es de 0.567

Interpretación

Como podemos ver la “ p ” no es significativa esto es, no es <0.05 por lo que se interpreta que no existe una diferencia estadísticamente significativa entre el tiempo de cicatrización con el abordaje A y el abordaje B, por lo que descartamos la hipótesis planteada.

Para realizar esta prueba en los softwares de análisis estadísticos como el programa SPSS es necesario seleccionar el apartado de pruebas no paramétricas y utilizar la comparación de medianas entre muestras independientes.

LECTURA COMPLEMENTARIA

1. Clifford B.R., Taylor R.A., Bioestadística, México: Pearson Educación; 2008.
2. Dawson B., Trapp R.G., Bioestadística médica. 4ª ed., México: Manual Moderno; 2005.
3. Nachar N., The Mann-Whitney U: A test for Assessing Whether two Independent Samples Come from the Same Distribution, *Tutorials in Quantitative Methods for Psychology*, 2008;vol 4(1):13-20.
4. Rivas. R.R., Moreno P.J., Talavera O.J., Diferencias de medianas con la U de Mann-Whitney, *Rev Med Inst Mex Seguro Soc.* 2013; 51(4):414-419.
5. Rivas. R.R., Moreno P.J., Talavera O.J., Pertinencia de la prueba estadística, *Rev Med Inst Mex Seguro Soc.* 2013;51(Supl):S30-34.
6. Wayne W.D. Bioestadística Base para el análisis de las ciencias de la salud. 4ª ed., México: Limusa Wiley. 2011.

PRUEBA DE KRUSKAL-WALLIS

*Esteban Araya Ramírez
Fabián Víquez Monge*

Esta metodología se utiliza en los casos en los cuales se poseen dos o más grupos de datos y se desea averiguar si estos provienen de una misma población; o bien, de poblaciones distintas pero con las mismas características entre sí.

Estadísticamente, se trata de comprobar la hipótesis de que k muestras independientes provienen de una población con la misma mediana.

Se encuentra catalogada como una prueba **no paramétrica**, es decir, no se necesita que los datos que se analizan posean una distribución con parámetros específicos.

Además de esto, tampoco es necesario que se cumpla el supuesto de igualdad de varianzas entre los grupos de datos (homogeneidad de varianzas).

La prueba se define matemáticamente de la forma siguiente:

$$H = \frac{\frac{12}{N(N+1)} \sum \frac{\sum Rc^2}{ni} - 3(N+1)}{L}$$

Donde:

H= Valor estadístico de la prueba de Kruskal-Wallis.

N= tamaño total de la muestra.

Rc²= sumatoria de los rangos elevados al cuadrado.

ni= tamaño de la muestra de cada grupo.

L= ajuste dado por el ajuste de ligas o empates de los rangos.

El ajuste de L se calcula de la manera siguiente:

$$L = 1 - \frac{\sum(Li^3 - Li)}{N^3 - N}$$

Donde:

L_i = valor de número de empates de un rango.

N = tamaño total de la muestra.

Ejemplo

Una enfermera lleva a cabo un estudio transversal multicéntrico para conocer la percepción de malestar de los trabajadores en el ámbito laboral, y quiere determinar si existen diferencias significativas entre las tres centrales eléctricas.

El estudio fue llevado a cabo en tres centrales eléctricas con un total de 22 sujetos, diagnosticados previamente con síndrome de *Burnout*.

La primera fábrica (A) contaba con 7 sujetos, la segunda (B), con 10 y la última (C) con 5 sujetos.

Para conocer la opinión de cada trabajador se les aplicó un cuestionario que incluye una serie de preguntas que fueron valoradas según una escala de 1 a 10; siendo los valores más altos los relacionados con sentimientos negativos.

La variable de estudio tiene un carácter ordinal, por tanto la medida de tendencia central a utilizar es la mediana.

En el estudio la hipótesis nula (H_0), será la de igualdad de medianas en los distintos grupos; siendo la hipótesis alternativa (H_a) por su parte, la existencia de diferencias significativas entre los distintos grupos de estudio.

En primer lugar, se desarrolla la base de datos de estudio y se obtiene un descriptivo del estudio:

ID	Fabrica	Puntuación	VAR-4	VAR-5	VAR-6	VAR-7	VAR-8	VAR-9	VAR-10
1	1	1	6						
2	2	1	5						
3	3	1	7						
4	4	1	5						
5	5	1	7						
6	6	1	9						
7	7	1	8						
8	8	2	5						
9	9	2	4						
10	10	2	6						
11	11	2	4						
12	12	2	3						
13	13	2	3						
14	14	2	2						
15	15	2	6						
16	16	2	5						
17	17	2	3						
18	18	3	8						
19	19	3	9						
20	20	3	7						
21	21	3	9						
22	22	3	6						
23									
24									

Fabrica	Mediana	Puntuación
A	1	6
	2	5
	3	7
	4	5
	5	7
	6	9
	7	8
B	8	5
	9	4
	10	6
	11	4
	12	3
	13	3
	14	2
	15	6
	16	5
	17	3
C	18	8
	19	9
	20	7
	21	9
	22	6

En este estudio se comparan tres grupos independientes además, el tamaño de la muestra es pequeño y la variable resultado se trata de una variable de carácter ordinal. Para el análisis estadístico, la prueba estadística de elección es la prueba de **Kruskal-Wallis**.

Para ello mediante la ayuda de un programa estadístico, en el apartado *Anova*, se selecciona la opción *Kruskal-Wallis*.

Posteriormente, se coloca en la variable respuesta (y) y la variable ordinal (variable puntuación) y la variable explicativa (b) la variable cualitativa (variable fábrica).

A continuación, se selecciona el apartado de *Kruskal-Wallis* para ver los resultados de la prueba y comprobar si existen diferencias, además se pueden ver en el apartado de estadísticos las medianas de los tres grupos.

En el caso de la prueba de *Kruskal-Wallis*, hay que fijarse en el valor de p que es de 0,0017.

The screenshot shows two windows from the G-Stat software. The main window displays the results of the Kruskal-Wallis test for the variable 'Puntuación por Fabrica'. The test statistic is 2.0000 and the p-value is 0.0017. The secondary window provides a detailed breakdown of statistics for each of the three groups.

Grupos	n	Suma de Rangos
1	7	101.0000
2	10	63.0000
3	5	89.0000

Grupos	1	2	3
N	7	10	5
Mediana	7.0000	4.0000	8.0000
Moda	5.0000	3.0000	9.0000
Máximo	5.0000	2.0000	6.0000
Mínimo	9.0000	6.0000	9.0000
Rango	4.0000	4.0000	3.0000
Cuartil Inferior	5.0000	3.0000	7.0000
Cuartil Superior	8.0000	5.0000	9.0000
Rango Intercuartilico	3.0000	2.0000	2.0000

Por último, al ser el valor menor a 0,05 se rechaza la hipótesis nula (H_0) de igualdad y se acepta la hipótesis alternativa (H_a) de diferencia de medianas con un intervalo de confianza del 95%.

A posteriori se lleva a cabo mediante la prueba de Dunn, las comparaciones entre los distintos grupos de estudio. De esta forma, se obtiene que existen diferencia entre la fábrica A y B y entre la fábrica B y C.

Puntuación	N	Rango Medio	Homogéneos
2	10	6.3000	X
1	7	14.4286	X
3	5	17.8000	X

Contraste	Diferencia	+/- Límite
2 VS 1	*8.1286	*7.5870 —
3 VS 1	-3.3714	9.0147
3 VS 2	*-11.5000	*8.4325 —

Por tanto, se puede decir que existen diferencias significativas en las medianas de la percepción de malestar de los trabajadores en el ámbito laboral entre la fábrica A y la fábrica B y la fábrica C con un intervalo de confianza del 95%.

No podemos afirmar que existan diferencias significativas entre la fábrica A y C.

LECTURA COMPLEMENTARIA

1. Altman DG: Practical statistics for medical research. 1a edición. Chapman and Hall, Londres, 1991.
2. Lopez, JP. Contraste de hipótesis. Comparación de más de dos medias independientes mediante pruebas no paramétricas: Prueba de Kruskal-Wallis. Enfermería del Trabajo 2013; III: 116-117.
3. Siegel S, Castellan NJ Jr: Nonparametric statistics for the behavioral sciences. 2ª edición. McGraw-Hill, New York, 1988.

CAPÍTULO 22

**COEFICIENTE DE CORRELACIÓN
DE SPEARMAN**

*Fernando Alvarado
Miguel Farfán*

El conjunto de técnicas estadísticas que permiten medir la intensidad de asociación o relación que puede existir entre dos o más variables se conoce como correlación. Esta correlación puede existir entre al menos dos variables (una dependiente y otra independiente) o entre más de dos variables (una dependiente y dos o más independientes). En el último caso se habla de correlación múltiple.

Los tipos de asociación (correlación) entre dos variables pueden ser lineales, positivos, negativos o nulos.

El coeficiente de correlación de Spearman (Rho de Spearman) es una prueba de naturaleza no paramétrica útil para determinar (cuantificar) si existe una correlación entre dos variables de naturaleza numérica (ordinales) cuya distribución sea así mismo no normal o no paramétrica.

Esta medición de correlación tiene tres objetivos fundamentales; el primero, es determinar si dicha correlación realmente existe (los valores de una variable tienden a ser más altos o más bajos para los valores más altos o más bajos de la otra variable), el segundo, es poder predecir el valor de una variable conociendo el valor determinado de la otra variable y por último, valorar el nivel de concordancia o relación entre los valores de las variables.

El coeficiente de correlación de Spearman tiene gran utilidad en variables que, dado su comportamiento no normal, presentan valores extremos en sus datos que, afectarían estadísticos como el coeficiente de correlación de Pearson (Ver capítulo respectivo).

Cabe resaltar que la Rho de Spearman mide la existencia y la intensidad de asociación de dos cantidades y no detecta el nivel de acuerdo o concordancia de estas; por lo anterior, no debe ser usado para comparar valores procedentes de instrumentos que midan el mismo evento.

Hipótesis del coeficiente de correlación de Spearman

H_0 : No existe asociación entre las dos variables

H_a : Existe una relación o asociación entre las dos variables (Positiva o negativa).

Ecuación

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

En donde $d_i = r_{xi} - r_{yi}$ es la diferencia entre los rangos de X e Y.

Interpretación

Para la interpretación, los autores, antes de correr la prueba, deben definir cuál es el objetivo de la investigación, así como determinar la importancia de las relaciones de las variables que se encuentran involucradas en la prueba, para, en caso de presencia de casualidad, no dejar las conclusiones de la correlación exclusivamente a un estadístico matemático.

La correlación de Spearman arroja un valor numérico de la intensidad de asociación (positiva o negativa) de una tendencia de dos variables a crecer o decrecer juntas y no indica causalidad entre ellas.

Según la literatura, existen al menos cuatro maneras de interpretar la Rho de Spearman en escalas numéricas, pero en términos generales, todas las escalas determinan rangos de la prueba entre -1 y +1. Aquellos valores que se acercan a 1, indican una correlación fuerte y positiva; aquellos valores que se acercan a -1, indican una correlación fuerte y negativa. Los valores cercanos a 0 indican la no existencia de correlación **lineal**. El signo positivo o negativo previo al valor numérico, indica el aumento o la disminución de una variable con relación a la otra. Es decir, un signo negativo refleja el aumento de una variable a medida que la otra disminuye o viceversa y uno positivo, el aumento de una variable a medida que la otra lo hace o la disminución de una a medida que la otra lo hace en simultáneo.

Es poco frecuente encontrar en estudios clínicos, variables con valores aproximados a +0,95. Se sugiere la revisión de datos y cálculos en búsqueda de errores o inconsistencias, con valores mayores o menores a 1.

Martínez Ortega del instituto superior de Ciencias Médicas de la Habana (ver referencia en bibliografía) realiza, en conjunto con sus colaboradores, una búsqueda de la bibliografía con relación al estadístico Rho de Spearman y resume 4 diferentes escalas para interpretar la prueba, a saber:

Escala 1:

El coeficiente de correlación oscila entre -1 y $+1$, el valor 0 que indica que no existe asociación lineal entre las dos variables en estudio.

Escala 2:

Correlación negativa perfecta	-1
Correlación negativa moderada débil	-0,5
Ninguna correlación	0
Correlación positiva moderada Fuerte	+0,5
Correlación positiva perfecta	+1

Escala 3:

- 1) Perfecta $R = 1$
- 2) Excelente $R = 0.9 \leq R < 1$
- 3) Buena $R = 0.8 \leq R < 0.9$
- 4) Regular $R = 0.5 \leq R < 0.8$
- 5) Mala $R < 0.5$

Escala 4:

Rango	Relación
0-0,25	Escasa o nula
0,26-0,50	Débil
0,51-0,75	Entre moderada y fuerte
0,76-1,00	Entre fuerte y perfecta

Ejemplo de la aplicación del estadístico Rho de Spearman

Feise et al. (ver referencia en bibliografía) en su estudio de construcción y validación de una *escala de calidad de vida para escoliosis idiopática* (SQLI por sus siglas en inglés) de 22 items, acuden al estadístico de correlación Rho de Spearman, al comparar su nueva escala, con el ya conocido *Perfil de calidad de vida para deformidades de la columna* (QLPSD por sus siglas en inglés).

En un grupo de 84 pacientes, 74 pacientes adolescentes intervenidos por escoliosis idiopática (manejo quirúrgico, con brace o sin brace), y 14 controles (hermanos), se realizó la aplicación de las dos escalas (SQLI y QLPSD). Los resultados numéricos globales de las escalas fueron comparados con el fin de determinar si existía o no una correlación positiva entre ellas.

Feise describe una clasificación adicional para la interpretación del estadístico Rho de Spearman como se muestra en la tabla 2.

Al comparar los valores numéricos globales de las dos escalas, obtienen un estadístico de correlación Rho de Spearman de +0.79, asumiendo una correlación muy buena entre ellas. (Figura 1).

Tabla 2 – Interpretación de la Rho de Spearman de acuerdo a Feise et al.

Rango	Relación
0-0,2	Pobre
0,2-0,4	Débil
0,41-0,6	Buena
0,61-0,8	Muy buena
0,81-1	Excelente

Figura 1 – Tabla de correlación de escalas SQLI y QLPSD en escoliosis idiopática del adolescente en el estudio de Feise et al.

	SQLI Self-esteem	SQLI Back Pain	SQLI Physical Activity	SQLI Moods & Feelings	SQLI Global	QLPSD Psychosocial Functioning	QLPSD Sleep Disturbance	QLPSD Back Pain	QLPSD Body Image	QLPSD Back Flexibility
SQLI back pain	0.200									
SQLI physical activity	0.264	0.283								
SQLI moods & feelings	0.377	0.498	0.208							
SQLI global	0.642	0.659	0.552	0.782						
QLPSD psychosocial functioning	0.345	0.410	0.364	0.332	0.477					
QLPSD sleep disturbance	0.301	0.476	0.107	0.576	0.525	0.550				
QLPSD back pain	0.293	0.813	0.354	0.523	0.709	0.636	0.564			
QLPSD body image	0.545	0.462	0.220	0.465	0.624	0.415	0.517	0.538		
QLPSD back flexibility	0.343	0.466	0.457	0.351	0.549	0.545	0.313	0.562	0.334	
QLPSD global	0.478	0.697	0.381	0.611	0.788	0.723	0.730	0.861	0.753	0.706

LECTURAS COMPLEMENTARIAS

1. Abraira. V, Pérez de Vargas A. Métodos Multivariantes en Bioestadística. ED. Centro de Estudios Ramón Areces. España, 1996.
2. Feise RJ, Donaldson S, Crowther ER, Menke JM, Wright JG. Construction and validation of the scoliosis quality of life index in adolescent idiopathic scoliosis. *Spine (Phila Pa 1976)*. 2005 Jun 1;30(11):1310-5. PubMed PMID:15928558.
3. Martínez Ortega, R. M. (2009). El coeficiente de correlación de los rangos de Spearman caracterización. *Rev. Haban Cienc Méd La Habana, VIII(2)*, 1–19. Retrieved from <<http://scielo.sld.cu/pdf/rhcm/v8n2/rhcm17209.pdf>>.
4. Pita Fernández S, Pértega Díaz S. Unidad de Epidemiología Clínica y Bioestadística. Utilización e Interpretación de las Técnicas de Correlación disponible. Complejo Hospitalario Juan Canalejo. A Coruña (España): Cad Aten Primaria; 4: 141.
5. Santander Montes AJ, Ruiz Vaquero R. Relación entre variables cuantitativas. Cuba. *Informática Médica II*. Editorial ECIMED; 2004, p.1-30.

CAPÍTULO 23

PRUEBA DE MCNEMAR

*Esteban Araya Ramírez
Fabián Víquez Monge*

En ciertas ocasiones se realizan varias mediciones de la misma característica para cada uno de los individuos que se incluyen en una investigación. En estos casos, el interés se centra en comparar si hay o no cambios significativos entre las mediciones efectuadas en dos momentos diferentes, normalmente, antes y después de la aplicación de un tratamiento o de la realización de alguna intervención.

La prueba de Mc Nemar se utiliza para decidir si puede o no, aceptarse que determinado “*tratamiento*” induce un cambio en la respuesta de los elementos sometidos al mismo, es aplicable a los diseños “*antes – después*” en los que cada elemento actúa como su propio control.

Consiste en una prueba no paramétrica para 2 variables binarias relacionadas que contrasta los cambios en las respuestas utilizando la distribución de χ^2 .

No solamente es útil para detectar los cambios en las respuestas en los estudios experimentales del antes y después, sino también para comparar dos tipos de tratamiento.

Este procedimiento es útil cuando las muestras son 2 y resultan dependientes, el tipo de escala es nominal.

Los resultados correspondientes a una muestra de N elementos se disponen en una tabla de frecuencias de 2×2 para recoger el conjunto de las respuestas de los mismos elementos antes y después.

La escala de medición para X e Y es nominal con 2 categorías tales como: positivo o negativo, hembra o macho, presencia o ausencia, que se pueden denominar 0 y 1.

		Clasificación Y	
		(+) $Y_i = 0$	(-) $Y_i = 1$
Clasificación X	(+) $X_i = 0$	A (0,0)	B (0,1)
	(-) $X_i = 1$	C (1,0)	D (1,1)

Para ejemplificar aún mejor la tabla anterior, se puede simplificar a:

Antes / Después		-	+
-	a	b	
+	c	d	

En las celdas de la tabla, **a** es el número de elementos cuya respuesta es la misma, (-); **b** es el número de elementos cuya respuesta es (-) antes del *tratamiento* y (+) después de éste; **c** es el número de elementos que han cambiado de (+) a (-); y **d** es el número de elementos que mantienen la respuesta (+).

Por tanto, $b + c$ es el número total de elementos cuyas respuestas han cambiado y son los únicos que intervienen en el contraste.

La hipótesis nula (H_0) es que el *tratamiento* no induce cambios significativos en las respuestas, es decir, los cambios observados en la muestra se deben al azar, de forma que es igualmente probable un cambio de + a - que un cambio de - a +.

Así pues, si H_0 es cierta, de los $b+c$ elementos cuya respuesta ha cambiado es de esperar que $(b+c)/2$ hayan pasado de + a -, y $(b+c)/2$ hayan pasado de - a +.

En otras palabras, si H_0 es cierta, la frecuencia esperada en las correspondientes celdas es $(a+b) / 2$.

La hipótesis alternativa (H_a) puede ser no direccional, cuando postula que la probabilidad de un cambio de + a - tiene distinta probabilidad que un cambio de - a +, o direccional, cuando predice que un cambio de - a + es más (o menos) probable que un cambio de + a -.

El estadístico de prueba que permite contrastar si existen diferencias significativas entre las frecuencias esperadas y las observadas es:

$$X^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

O_i = frecuencia observada en la i -ésima celda

E_i = frecuencia esperada en la i -ésima celda si H_0 es cierta

K = número de celdas

Para contrastar la significación de los cambios interesan sólo las celdas que recogen cambios, por tanto el estadístico puede expresarse como:

$$X^2 = \frac{\left[b - \frac{b+c}{2}\right]^2}{\frac{b+c}{2}} + \frac{\left[c - \frac{b+c}{2}\right]^2}{\frac{b+c}{2}} = \frac{(b-c)^2}{b+c}$$

Si H_0 es cierta, el estadístico X^2 tiene distribución aproximadamente χ^2 con 1 grado de libertad.

La aproximación es más precisa si se realiza la corrección de continuidad de Yates, quedando el estadístico:

$$X^2 = \frac{(|b-c| - 1)^2}{b+c}$$

Grados libertad	Probabilidad de un valor superior - Alfa (α)				
	0,1	0,05	0,025	0,01	0,005
1	2,71	3,84	5,02	6,63	7,88
2	4,61	5,99	7,38	9,21	10,60
3	6,25	7,81	9,35	11,34	12,84
4	7,78	9,49	11,14	13,28	14,86
5	9,24	11,07	12,83	15,09	16,75
6	10,64	12,59	14,45	16,81	18,55

Ejemplo

Un investigador en medicina preventiva observa que los empleados en una fábrica padecen frecuentemente un cuadro diarreico, motivo de gran ausencia.

Todos los empleados comen en el comedor de la fábrica como goce de una prestación laboral.

El investigador supone que el común denominador de la causa de la diarrea es el sitio de ingestión de alimentos, es decir, existe una higiene inadecuada en la preparación de la comida; sin embargo, la higiene personal de los empleados no es suficiente para atribuir toda la culpa al personal de la cocina.

Por lo tanto, elige una muestra al azar de 50 individuos, de los cuales resulta que 34 de ellos presentan un cuadro diarreico frecuente y 16 no lo padecen, así, sugiere que, bajo vigilancia, se apliquen medidas de higiene personal, consistentes en exhaustivo lavado de manos antes de ingerir alimentos, en un período de dos semanas.

Al finalizar el tratamiento, obtiene los resultados siguientes: de los 34 sujetos con un cuadro diarreico frecuente, después del tratamiento de lavado de manos, 16 lograron hacer desaparecer el proceso intestinal y 18 persistieron con evacuaciones diarreicas; a su vez el grupo de 16 personas asintomáticas, 4 de ellas presentaron diarrea a pesar del lavado de manos y 12 se mantuvieron en las mismas condiciones.

Elección de la prueba estadística

El modelo experimental tiene dos muestras dependientes.

Planteamiento de la hipótesis

- Hipótesis alterna (H_a) El lavado de manos, como medida preventiva y factor de higiene personal, presenta cambios significativos de enfermedad diarreica en los empleados que asisten al comedor de la fábrica en estudio.
- Hipótesis nula (H_o) Las diferencias que se observan en las frecuencias de cambio por el lavado de manos se deben al azar.

Nivel de significancia

Para todo valor de probabilidad igual o menor que 0.05, se acepta H_a y se rechaza H_o .

Zona de rechazo

Para todo valor de probabilidad $>$ que 0.05, se acepta H_o y se rechaza H_a .

Resultado de la contingencia 2×2 .

		Después del lavado de manos	
		-	+
Antes del lavado de manos	-	16	18
	+	12	4

Aplicación de la prueba estadística

$$X^2 = \frac{((A - D) - 1)^2}{A + B} = \frac{(16 - 4 - 1)^2}{16 + 4} = \frac{121}{20} = 6.05$$

Cálculo de los grados de libertad (gl)

$$Gl = 1$$

El estadístico X^2 de McNemar se compara con los valores críticos de la ji cuadrada, con el grado de libertad, se observa que 3.84 corresponde a una probabilidad de 0.05, mientras que el cálculo corresponde a una probabilidad $<$ que 0.05.

Decisión

En razón de que el valor de X^2 calculado tiene probabilidad $<$ que 0.05, cae en el nivel de significancia, por lo tanto, se acepta H_a y se rechaza H_0 .

Interpretación

El lavado de manos ejerció significativos cambios en la enfermedad diarreica de los empleados de la fábrica, lo cual significa que la higiene personal contribuye de manera notoria en la frecuencia del proceso intestinal, causa del ausentismo.

La aplicación de la prueba estadística no cancela la participación de la higiene defectuosa en la preparación de los alimentos por parte del personal de la cocina, pues 4/16 empleados, que habían estado asintomáticos y bajo exhaustivo lavado de manos antes de comer, manifestaron diarrea; sin embargo, la prueba estadística revela que la higiene personal participa de manera más significativa que el defecto en la preparación de la comida.

LITERATURA COMPLEMENTARIA

1. Daniel WW: Estadística con aplicaciones a las Ciencias Sociales y a la educación. 1ra, edición en español. McGraw-Hill, México, 1988.
2. Manuel G. M., Danglot B. C., Vega F. L. Sinopsis de pruebas estadísticas no paramétricas. Cuándo usarlas. Rev Mex Pediatr 2003; 70(2); 91-99.
3. Martínez, M. A., Sánchez, A. V, Faulin F.J. Bioestadística Amigable. Segunda edición. España: Editorial Diaz de Santos, 2009.
4. Siegel. S, Castellan N. J. Estadística no paramétrica. 2. edición. Méjico: Editorial Trillas; 1995.
5. Sprent, P. Applied nonparametric statistical methods. 3. edición. Boca Ratón, 2001.

REGRESIÓN LINEAL SIMPLE

*Alisson R. Teles
Asdrubal Falavigna*

Descripción de la prueba

La regresión lineal simple es un modelo matemático utilizado para describir la relación entre dos variables con el fin de utilizar una de ellas para prever el valor de la otra. En otras palabras, es una técnica estadística que se utiliza cuando no hay razón para suponer una relación de causa-efecto entre dos variables cuantitativas y se desea expresar matemáticamente esta relación.

Los objetivos de una prueba de regresión lineal simple son, por consiguiente: (1) evaluar una posible dependencia entre dos variables cuantitativas (x = variable predictora o independiente / y = variable dependiente o resultado), y (2) expresar matemáticamente esta respuesta por medio de una ecuación.

Por ejemplo, se puede calcular la dependencia lineal entre la lordosis lumbar y la incidencia pélvica de una determinada población. Otro ejemplo ampliamente utilizado en la investigación médica, es cuando se desea saber la relación dosis-respuesta para un medicamento en particular para que se pueda prever la reacción fisiológica de un individuo debido a la utilización de una determinada cantidad de medicamento.

Los requisitos para el estudio de la regresión lineal son: la variable dependiente (y) debe ser aleatoria, es decir, obtenida por muestreo; las variables independientes (x) y dependiente (y) deben estar asociadas linealmente (la gráfica entre las dos variables debe presentar una línea recta); y las variancias de la variable dependiente (y), dados diferentes valores de la variable independiente (x) son todas iguales.

Fórmula e interpretación

La ecuación presenta las variables dependientes (y) la independiente (x) y un coeficiente que refleja la intensidad de la relación entre las variables x e y .

La ecuación del análisis de regresión se define por: $y = a + bx$ donde y es la variable dependiente (resultado), x la variable independiente (predictora), a es el coeficiente lineal (valor de y cuando $x = 0$) y b es el coeficiente angular (pendiente de la recta, aumento o disminución en y para cada aumento de una unidad en x). Gráficamente, la regresión lineal simple está representada por el diagrama de dispersión, también llamado diagrama de puntos. La ecuación se representa por la línea recta que expresa la relación lineal entre el valor medio de y para los distintos valores correspondientes de x .

Para determinar la recta de regresión lineal simple, se debe identificar el valor de a (intercepto vertical) y el valor b (coeficiente angular). La estimación de estos parámetros se lleva a cabo mediante una técnica denominada método de mínimos cuadrados. Se calculan los valores de a y b de modo que la suma de los cuadrados de las desviaciones de los puntos observados y , y la propia recta sea un mínimo. La distancia vertical entre el punto y la recta se denomina error o residual de la recta con respecto al punto. Los valores de a y b se calculan utilizando la siguiente fórmula:

$$b = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sum x^2 - \frac{\sum x^2}{n}} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

Una vez calculado el valor de b (coeficiente angular), el valor de a (intercepto vertical) es dado por:

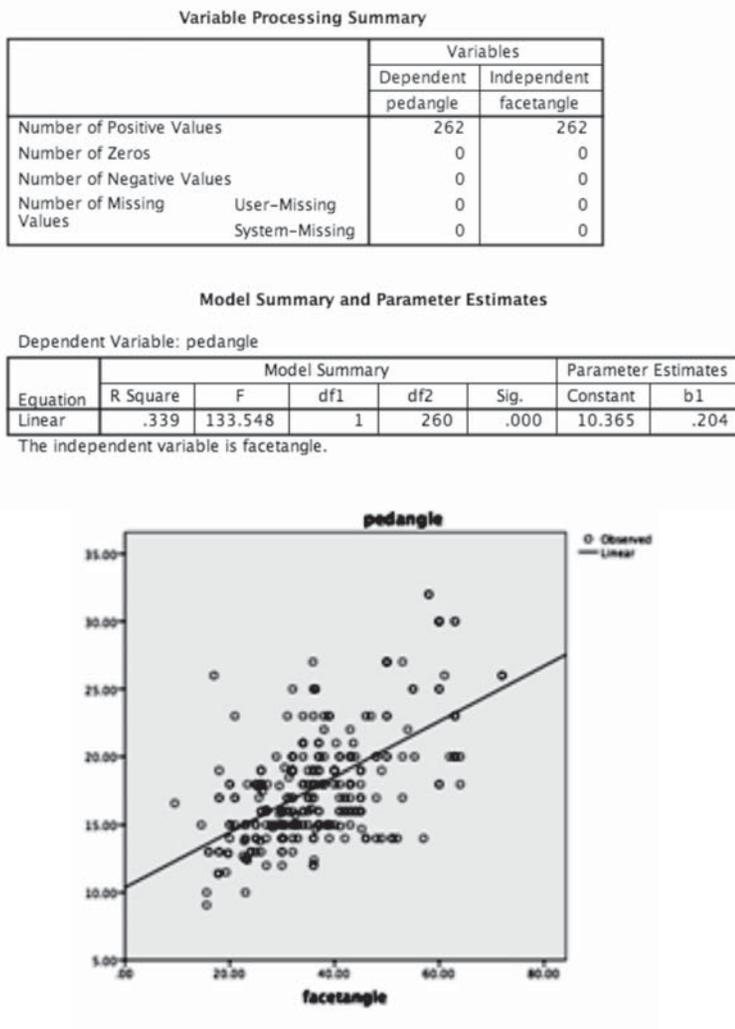
$$a = \bar{y} - b\bar{x}$$

Donde \bar{x} y \bar{y} , son, respectivamente, las medias de los datos relativos a las variables x y y . La recta estimada de regresión lineal es: $\hat{y} = a + bx$, donde \hat{y} es la estimativa de la observación y para un valor específico de la variable x .

En la práctica, la ecuación de regresión lineal simple puede generarse mediante diferentes softwares de análisis de datos.

El análisis genera primero datos que identifican la existencia o no de correlación significativa entre las variables (prerrequisito para obtener la ecuación de regresión), proporciona la gráfica de dispersión con la recta trazada e identifica los valores de a y b . En la figura se presenta un ejemplo de regresión lineal simple utilizando el software SPSS®.

Figura 1 – Ejemplo de análisis de regresión lineal simple utilizando el SPSS®



Nota: En el ejemplo anterior se demuestra la relación entre el ángulo del pedículo (pedangle) y el ángulo de la articulación ínterapofisaria o facetaria (facetangle) de vértebras lumbares examinadas a través de tomografía computada. Se observa (1) que la correlación entre las dos variables es significativa (Sig. .000 representa el valor $P < 0,0001$), (2) se identifica la gráfica de dispersión con la recta de regresión trazada (3) se identifican los valores de a (constant) y b (b_1). Por lo tanto, se puede decir que en la muestra estudiada, es posible definir la relación entre el ángulo del pedículo a través de la siguiente ecuación: ángulo del pedículo = $10.365 + (0.204 \times \text{ángulo de la faceta})$.

Ejemplo en el área de cirugía de columna

El análisis de regresión lineal simple se utiliza cuando se desea prever una variable en relación con la variación de otra. En la práctica, primero se identifica la existencia de una correlación significativa entre dos variables continuas, luego se realiza el análisis de regresión lineal para generar la ecuación que determina la variable de resultado (y) en relación a los valores de la variable predictora (x).

Son varias las aplicaciones en la región de la columna vertebral. Por ejemplo, Hasegawa y colaboradores identificaron una relación entre la discapacidad funcional medida por el índice de discapacidad Oswestry (ODI: *Oswestry Disability Index*) y (1) edad del voluntario del estudio y (2) *índice de desajuste (mismatch)* entre el índice pélvico y la lordosis lumbar. Se estudiaron 136 voluntarios sin enfermedad de la columna vertebral a través de radiografía y evaluación clínica. En el estudio, se identificó la siguiente ecuación entre ODI y edad: $ODI = 0.36 + 0.12 \times \text{edad}$. En este análisis, se obtuvo un valor r (coeficiente de correlación) de 0,2143 (Capítulo de Correlación de Pearson) y un valor de $p = 0,01664$. La ecuación de ODI y *mismatch* incidencia pélvica (IP) y la lordosis lumbar (LL) identificada fue: $ODI = 5,62 + 0,12 (IP-LL)$ con un valor $r = 0,1969$ y $P = 0,0277$. Con esta ecuación, se puede prever el valor de ODI teniendo en cuenta la edad y también el *mismatch* IP-LL.

LECTURAS COMPLEMENTARIAS

1. Hasegawa K., Okamoto M., Hatsushikano S., Shimoda H., Ono M., Watanabe K. Normative values of spino-pelvic sagittal alignment, balance, age, and health-related quality of life in a cohort of healthy adult subjects. *Eur Spine J* 2016. Doi: 10.1007/s00586-016-4702-2.
2. Peat J., Barton B. *Medical Statistics: a guide to data analysis and critical appraisal*. BMJ Books, 2008. ISBN: 978-0-7279-1812-3.
3. Peacock J. L., Peacock P. L. *Oxford Handbook of Medical Statistics*. Oxford University Express, 2011. ISBN: 978-0-19-9555128-6.

DISTRIBUCIÓN DE POISSON

*José Acuña Pinilla
Margarita Gómez Chantraine
Elizabeth Heins*

Descripción de la prueba

Llamada así en honor al físico y matemático de origen francés Simeón-Denis Poisson, (1781-1840) quien la dio a conocer en 1837 por medio de su trabajo: *Recherches sur la probabilité des jugements en matières criminelles et matière civile* (Investigación sobre la probabilidad de los juicios en materias criminales y civiles).

La distribución de Poisson es una de las funciones de probabilidad más importantes que hace parte del grupo de funciones para la distribución de probabilidad para variables aleatorias discretas, es decir, los eventos sólo puede tomar valores enteros (ej: 0,1,2,3,4...n) y cuya aparición de eventos aleatorios son independientes entre sí, presentándose durante un intervalo de tiempo o espacio.

Puntualmente se especializa en la frecuencia de sucesos con probabilidades muy bajas con relación a la frecuencia de la no ocurrencia por tal razón, se consideran como “sucesos raros”. Su aproximación a la binomial se dará cuando la probabilidad de presentarse el evento (**p**) es muy cercano a cero, o el tamaño muestral (**n**) sea superior a 30 ($p < 0,1$, $n > 30$).

Esta función parte de una frecuencia de ocurrencia media (λ) conocida también como parámetro positivo para la esperanza de ocurrencia y que identifica al número de ocurrencias o eventos promedio que se presentan dentro un intervalo de tiempo o espacio. Además, involucra el número de ocurrencias del evento o fenómeno (**x**) que se espera identificar (la función nos da la probabilidad que el evento suceda precisamente x veces). Por ejemplo, si el suceso estudiado tiene lugar en promedio 5 veces por minuto y estamos interesados en la probabilidad de que ocurra x veces dentro de un intervalo de 10 minutos, usaremos un modelo de distribución de Poisson con $\lambda = 10 \times 5 = 50$.

Características de la distribución de Poisson

La distribución resulta de un conjunto de suposiciones acerca de un proceso implícito para formar un conjunto de observaciones numéricas. Las siguientes afirmaciones describen lo que se conoce como proceso de Poisson.

- a) Las ocurrencias de los eventos son independientes. La aparición de un evento (x) en un intervalo de tiempo o espacio no tienen efecto sobre la probabilidad de presentarse una siguiente aparición del evento estudiado en el mismo intervalo, o en algún otro.
- b) Teóricamente, debe ser posible la aparición del evento estudiado en un número infinito de veces dentro del intervalo de tiempo o espacio.
- c) La probabilidad de una sola aparición del evento estudio en un intervalo dado es proporcional a la dimensión del intervalo de tiempo o espacio (o algo de volumen de la materia).
- d) La probabilidad de más de una ocurrencia del evento estudio en cualquier fracción infinitesimal del intervalo de tiempo o espacio, es insignificante.

Una característica interesante de la distribución de Poisson es que la media y la varianza son iguales.

Aplicación

La distribución de Poisson se emplea para describir varios procesos, entre otros:

- a) El número de lesiones vasculares de vasos principales en cirugía de columna vía anterior al año.
- b) El número de lesiones radicales en discectomías mínimamente invasivas del total de pacientes operados por esta técnica en un periodo determinado.
- c) El número de tornillos pediculares que presentaron pull-out en la reducción del total de niveles funcionales intervenidos en cirugía de escoliosis.
- d) El número de tornillos pediculares mal posicionados en cirugía fusión lumbar posterior al año.
- e) El número de casos con pseudoartrosis posterior a una cirugía de fusión cervical anterior al mes.

Cada una de estas variables aleatorias representa el número total de ocurrencias de un fenómeno durante un periodo de tiempo fijo o en una región fija del espacio. Expresa la probabilidad de un número (x) de apariciones del evento en un tiempo determinado, si estos eventos ocurren con una frecuencia media conocida y son independientes del tiempo transcurrido desde la última ocurrencia o suceso.

La probabilidad de la aparición del número de eventos se puede calcular de manera puntual o acumulada, es decir, una cantidad agrupada de posibles eventos en el intervalo de tiempo o espacio.

Función Densidad de Probabilidad:

$$f(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad \text{sí } x = 0, 1, 2, 3, \dots, n$$

Se puede demostrar que $f(x) \geq 0$ para cada (x) y que $\sum_x f(x) = 1$ de modo que la distribución cumple los requisitos para una distribución de probabilidad.

Dónde:

λ = Parámetro esperanza de ocurrencia (# ocurrencias / Unidad de tiempo o espacio)

x = # de eventos del fenómeno esperados (0, 1, 2, 3, 4, ... n)

e = Base de los logaritmos naturales ($e = 2,71828\dots$)

Función de Probabilidad Acumulada:

$$f(X \leq x) = \sum_0^x \frac{e^{-\lambda} \lambda^x}{x!}$$

Varianza $\sigma^2 =$ Media $\mu =$ Parámetro esperanza de ocurrencia λ

$$\text{Desviación } \sigma = \sqrt{\sigma^2} = \sqrt{\mu}$$

Graficas

Figura 1 – La función densidad de probabilidad (FDP) para toda ocurrencia en valores enteros (x) con una expectativa de ocurrencia (λ). La línea continua es solo una guía visual y no representa continuidad

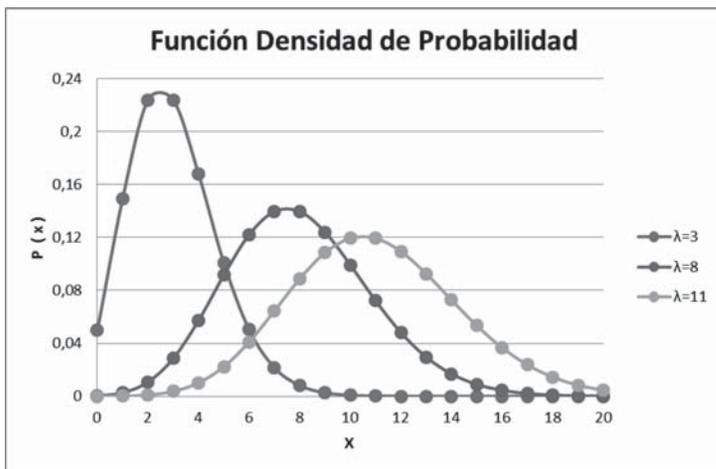
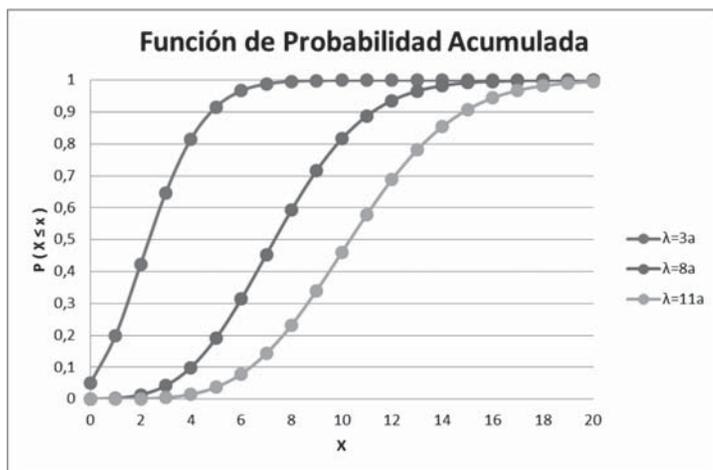


Figura 2 – La función de distribución acumulada (FDA). La FDA es discontinua en los enteros de x y plana en todas partes debido a una variable que es la distribución de Poisson toma solamente valores enteros



Ejemplos en el área de la cirugía de columna

1. En un meta-análisis sobre la presencia de pseudoartrosis en discectomía cervical anterior y fusión con el uso de injertos durante el periodo de seguimiento, Shriver MF. et al. (2015), en los 17 estudios finalmente incluidos con revisión hasta Noviembre 10, 2014 encontraron una ocurrencia de pseudoartrosis en 35 casos durante postquirúrgico de pacientes a los que se les trato con aloinjertos (n:736). ¿Cuál es la probabilidad que durante el siguiente año, entre los pacientes que se les realizará discectomía cervical vía anterior y fusión con aloinjerto, exactamente se presenten dos casos con pseudoartrosis en el seguimiento post-operatorio?

Solución:

$$p(x = 2)$$

$$\lambda = \frac{35(\text{Casos de pseudoartrosis})}{736(\text{Casos con Aloinjertos})} = 0.048$$

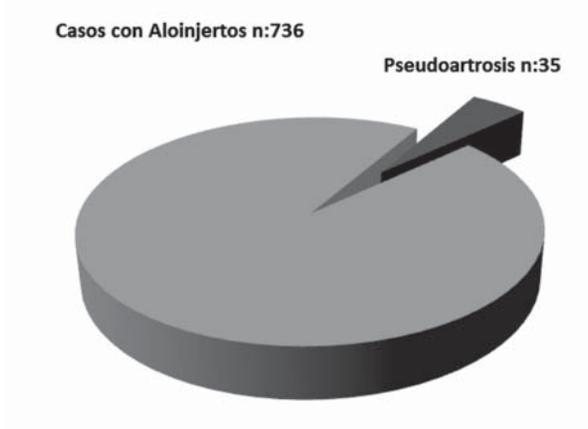
$$x = 2 \text{ (casos esperados con pseudoartrosis)}$$

$$f(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

$$f(x = 2) = \frac{e^{-0.048} * 0.048^2}{2!} = \frac{0.9531 * 0.0023}{2} = 0,0011$$

La probabilidad que se presente exactamente 2 casos con pseudoartrosis en el seguimiento de discectomía cervical vía anterior y fusión con aloinjerto es de 0.11%.

Figura 3 – Frecuencia casos de pseudoartrosis. $p(x = 2)$



2. En un meta-análisis sobre el correcto posicionamiento de los tornillos pediculares con el uso de navegación comparado con el posicionamiento sin navegación, Kosmopoulos V, et al. (2007), encontraron en los estudios incluidos con revisión entre 1966 hasta Marzo 14, 2006 una mediana de mal posicionamiento de tornillos del 9,7% para el grupo sin el uso de navegación en casos In Vivo, equivalentes a una ocurrencia mediana mensual de 8 tornillos mal posicionados.⁵ ¿Cuál es la probabilidad que en un mes seleccionado aleatoriamente sea uno en el cual se presenten 3 o 4 casos con tornillos pediculares mal posicionados en cirugía?

Solución:

$$p(x = 3 \text{ o } 4) = p(x = 3) + p(x = 4)$$

$$\lambda = \frac{8 \text{ (Casos de tornillos mal posicionados)}}{1 \text{ (Unidad de tiempo en meses)}} = 8$$

$$x_1 = 3 \text{ (Casos esperados de tornillos mal posicionados)}$$

$$x_2 = 4 \text{ (Casos esperados de tornillos mal posicionados)}$$

$$f(X = x_1) = \frac{e^{-\lambda} \lambda^x}{x!} \rightarrow f(x_1 = 3) = \frac{e^{-8} * 8^3}{3!} = \frac{0.00034 * 512}{6} = 0,029$$

$$f(X = x_2) = \frac{e^{-\lambda} \lambda^x}{x!} \rightarrow f(x_2 = 4) = \frac{e^{-8} * 8^4}{4!} = \frac{0.00034 * 4096}{24} = 0,057$$

$$f(x_1 = 3 \text{ o } x_2 = 4) = f(x_1 = 3) + f(x_2 = 4)$$

$$f(x_1 = 3) + f(x_2 = 4) = 0,029 + 0,057 = 0,086$$

La probabilidad que se presenten 3 o 4 casos con tornillos pediculares mal posicionados en cirugía es de 8,6%.

3. En un estudio prospectivo multicéntrico sobre el correcto posicionamiento de los tornillos pediculares con el uso de sistema de imágenes con brazo en O y navegación Stealth Station en los niveles torácico, lumbar y sacro, Van de Kelft E, et al. (2012), en los 3 centros de neurocirugía de 2 países entre Noviembre 2009 y Abril 2011 encontraron una ocurrencia de mal posicionamiento de tornillos en un 2,5%, de los cuales reportan una ocurrencia anual de 5 tornillos con compromiso de la cortical medial del pedículo.⁹ ¿Cuál es la

probabilidad que durante el siguiente año entre los pacientes que se les realizará fijación vía posterior con tornillos pediculares, se presenten en cirugía al menos 4 tornillos pediculares mal posicionados y comprometida la cortical medial?

Solución:

La posibilidad de aparición de eventos es infinita y la sumatoria de todos los posibles eventos es igual a 1. Entonces, podemos usar el concepto de complementariedad al restar al 1 la probabilidad de 3 o menos tornillos mal posicionados. Es decir:

$$p(x \geq 4) = 1 - p(x \leq 3)$$

Por lo anterior se debe despejar primero la probabilidad acumulada de $p(x \leq 3)$

$$p(X \leq x) = \sum_0^x \frac{e^{-\lambda} \lambda^x}{x!} \rightarrow p(X \leq 3) = \sum_0^3 \frac{e^{-5} 5^3}{3!}$$

$$= p(x=0) + p(x=1) + p(x=2) + p(x=3)$$

$$p(x \geq 4) = 1 - [p(x=0) + p(x=1) + p(x=2) + p(x=3)]$$

Entonces:

$$\lambda = \frac{5 \text{ (Casos de tornillos mal posicionados)}}{1 \text{ (Unidad de tiempo en años)}} = 5$$

$$x = 0, 1, 2 \text{ y } 3 \text{ (Casos esperados de tornillos mal posicionados)}$$

$$f(X \geq x) = 1 - \sum_0^x \frac{e^{-\lambda} \lambda^x}{x!}$$

$$f(X \geq 4) = 1 - \sum_0^3 \frac{e^{-5} 5^3}{3!}$$

$$p(x \geq 4) = 1 - [p(x=0) + p(x=1) + p(x=2) + p(x=3)]$$

$$p(x \geq 4) = 1 - \left[\frac{e^{-5} * 5^0}{0!} + \frac{e^{-5} * 5^1}{1!} + \frac{e^{-5} * 5^2}{2!} + \frac{e^{-5} * 5^3}{3!} \right]$$

$$p(x \geq 4) = 1 - [0,0067 + 0,0337 + 0,0842 + 0,1404]$$

$$p(x \geq 4) = 1 - 0,2650 = 0,7350$$

Tabla 1 – Calculo distribución de Poisson para densidad y acumulada para x= 0 hasta x=3 con un $\lambda =5$

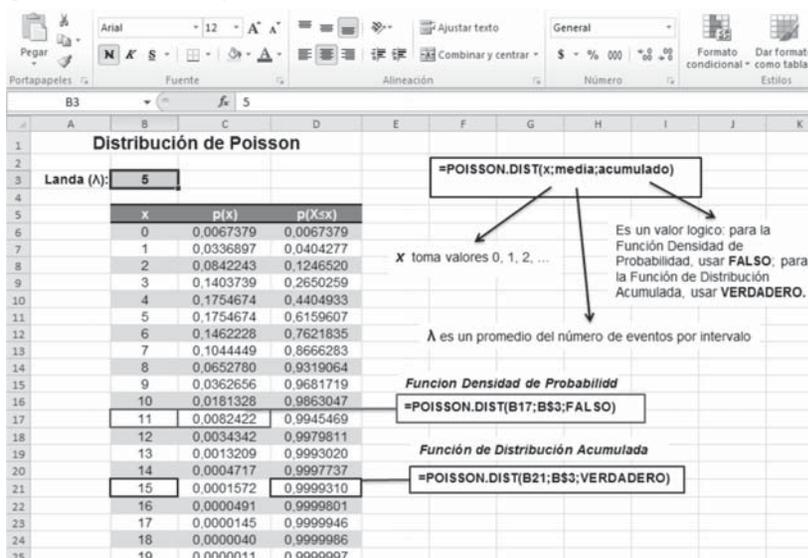
X	P(X=x) FDP	P(X≤x) FDA
0	0,0067	0,0067
1	0,0337	0,0404
2	0,0842	0,1247
3	0,1404	0,2650

La probabilidad que se presenten 4 o más casos al año con tornillos pediculares mal posicionados y compromiso de la cortical medial durante cirugía es de 73.50%.

Tabla 2 –Transformación de la fórmula de Poisson para el cálculo en Excel
Como calcular distribución de probabilidad de Poisson en Excel.

	Función	Fórmula en Excel
Probabilidad densidad	$f(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$	→ = POISSON.DIST(x; λ; 0)
Probabilidad acumulada	$f(X \leq x) = \sum_0^x \frac{e^{-\lambda} \lambda^x}{x!}$	→ = POISSON.DIST(x; λ; 1)

Figura 4 – Los componentes de la formula son:



- (x) número de eventos del fenómeno esperados.
- (λ) Parámetro esperanza de ocurrencia.

(acumulado) es un valor lógico, para la función densidad de probabilidad, usar "FALSO" o "0"; para la función de distribución acumulada, usar "VERDADERO" o "1".

Tabla 3 – Transformación de la fórmula de Poisson para el cálculo en SPSS.

Como calcular distribución de probabilidad de Poisson en SPSS

	Función	Fórmula en SPSS
Probabilidad densidad	$f(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$	Transformar → Calcular variable Expresión numérica = PDF.POISSON(x, λ)
Probabilidad acumulada	$f(X \leq x) = \sum_0^x \frac{e^{-\lambda} \lambda^x}{x!}$	Transformar → Calcular variable Expresión numérica = CDF.POISSON(x, λ)

Figura 5 – Ingresar los valores del número de eventos del fenómeno esperados (X), Seleccionar en transformar la función calcular variable (1). Nombrar la variable para la función densidad de probabilidad (FDP) (2). Seleccionar en el grupo de funciones "FDP & FDP no centrada" (3). En funciones y variables especiales seleccionar con doble clic "Pdf. Poisson" (4). Completar la fórmula sustituyendo en "(?)" primero por la variable X seguido por el valor del parámetro esperanza de ocurrencia (λ) (5). Finalizar con "Aceptar" para ejecutar el cálculo.

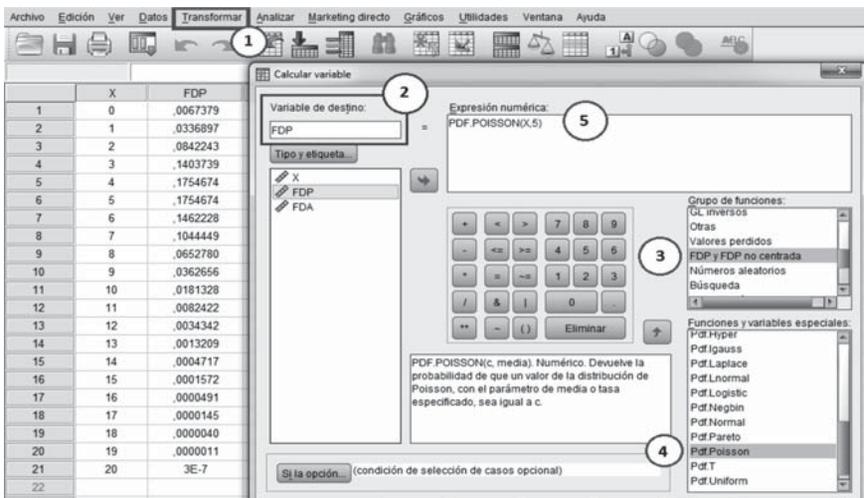


Figura 6 – Ingresar los valores del número de eventos del fenómeno esperados (X), Seleccionar en transformar la función calcular variable (1). Nombrar la variable para la función de distribución acumulada (FDA) (2). Seleccionar en el grupo de funciones “FDA & FDA no centrada” (3). En funciones y variables especiales seleccionar con doble clic “Cdf. Poisson” (4). Completar la formula sustituyendo en “(?,?)” primero por la variable X seguido por el valor del parámetro esperanza de ocurrencia (λ) (5). Finalizar con “Aceptar” para ejecutar el cálculo

	X	FDP	FDA
1	0	,0067379	,0067379
2	1	,0336897	,0404277
3	2	,0842243	,1246520
4	3	,1403739	,2650259
5	4	,1754674	,4404933
6	5	,1754674	,6159607
7	6	,1462228	,7621835
8	7	,1044449	,8666283
9	8	,0652780	,9319064
10	9	,0362656	,9681719
11	10	,0181328	,9863047
12	11	,0082422	,9945469
13	12	,0034342	,9979811
14	13	,0013209	,9993020
15	14	,0004717	,9997737
16	15	,0001572	,9999310
17	16	,0000491	,9999801
18	17	,0000145	,9999946
19	18	,0000040	,9999986
20	19	,0000011	,9999997
21	20	3E-7	,9999999
22			
23			

Interpretación

La distribución de Poisson es muy útil cuando se desea conocer las posibilidades de aparición de un número de eventos poco frecuentes y/o su impacto en el desarrollo de una cirugía de columna. Podemos conocer la posibilidad de aparición de una cantidad específica de eventos en el que estamos interesados (ejemplo 1), como también en un conjunto de posibles apariciones en un periodo de tiempo determinado (ejemplos 2, 3).

Por otro lado, la distribución de Poisson permite hacer evaluaciones sobre estrategias para minimizar la aparición de eventos no deseados en cirugía de columna como en los casos de mal posicionamiento de tornillos pediculares (Ej 2 y3). Al hacer estos cálculos podríamos predecir cambios en las posibilidades de aparición de eventos no deseados a futuro, en periodos de tiempo definidos, y así permitir el control del número de eventos a presentar en este tipo de cirugía. El método es aplicable a cualquier evento poco frecuente.

LECTURA COMPLEMENTARIA

1. Ahlbom A. Biostatistic for Epidemiologists. Florida: Lewis. Publishers. 2010.
2. Everitt B.S., Skrondal A. The Cambridge Dictionary of Statistics. 4th Ed. Cambridge, 2010.
3. Glover T, Mitchell K. An Introduction to Biostatistics. 3th Ed. Illinois, Waveland Press Ind, 2016.
4. Haight F. Handbook of the Poisson Distribution, Wiley, New York, 1967.
5. Kosmopoulos V, Schizas C . Pedicle screw placement accuracy: a meta-analysis. Spine (Phila Pa 1976) 2007; 32:E111-20.
6. Levine D.M., Stephan D.F. Even You Can Learn Statistics. A Guide for Everyone Who Has Ever Been Afraid of Statistics. Second Edition. New Jersey, Pearson, 2010.
7. Rosner B. Fundamentals of Biostatistics. 7th Ed. Boston, Books/Cole, 2011.
8. Shriver MF, Lewis DJ, Kshetry VR, Rosenbaum BP, Benzel EC, Mroz TE. Pseudoarthrosis rates in anterior cervical discectomy and fusion: a meta-analysis. Spine J. 2015; 15(9):2016-27.
9. Van de Kelft E, Costa F, Van der Planken D, Schils F. A prospective multicenter registry on the accuracy of pedicle screw placement in the thoracic, lumbar, and sacral levels with the use of the O-arm imaging system and StealthStation Navigation. Spine (Phila Pa 1976). 2012; 37(25):E1580-7.
10. Walpole R. E., Myers R. H., Myers S. L. Probability & Statistics for Engineers & Scientists. 9th Ed. Boston, Pearson, 2012.
11. Wayne WD. BIOSTATISTICS A Foundation for Analysis in the Health Sciences. 10th Ed. New York, Willey, 2013.

ODDS RATIO

*José Manuel Pérez Atanasio
Víctor Fernando Luján Celis
Maritza Belén Sandoval Rincón*

Descripción de la prueba

El *odds ratio* o razón de momios es una prueba estadística que corresponde a la razón entre la probabilidad de que un evento ocurra y la probabilidad de que no ocurra, expresa el riesgo de presentar una condición o enfermedad si se tiene algún factor de riesgo.

Formula

Figura 1– Formula del odds ratio

$$OR = \frac{(a \times d)}{(b \times c)}$$

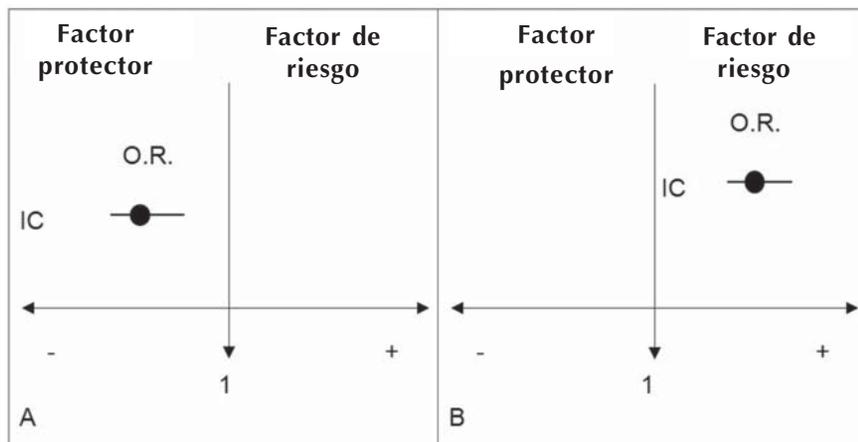
Para obtener el odds ratio primero se tiene que elaborar una tabla de contingencia o tabla de 2x2, que contiene los espacios a, b, c, d.

Figura 2 – Tabla de Contingencia o de 2 x 2

	Pacientes que desarrollaron la enfermedad	Pacientes que no desarrollaron la enfermedad
Pacientes con el factor de riesgo	a	b
Pacientes sin el factor de riesgo	c	d

El valor obtenido del odds ratio nos indica si el factor estudiado es un factor protector (<1) o un factor de riesgo (>1) para desarrollar la enfermedad, esto se puede ejemplificar en la siguiente imagen.

Figura 3 – Representación gráfica del odds ratio como factor protector (A) y como factor de riesgo (B).



Ejemplo en el área de la cirugía de columna

El Dr. Juan cirujano de columna ha laborado por más de 20 años en el Hospital X, en fechas recientes ha visto un incremento de las infecciones del sitio quirúrgico en los pacientes post operados de instrumentación en la columna lumbar, así mismo ha observado un incremento de la obesidad de los pacientes, por lo que se plantea la siguiente pregunta de investigación:

¿La obesidad es un factor de riesgo para presentar infección del sitio quirúrgico en pacientes post operados de instrumentación en la columna lumbar?

Para responder esta pregunta elaboró un protocolo de investigación con diseño metodológico de casos y controles, revisó los expedientes de 269 pacientes e identificó a los que presentaron infección del sitio quirúrgico (casos) y los pacientes que no presentaron (controles) así mismo identificó los que tenían obesidad considerándolo como el factor de riesgo. Una vez que obtuvo los datos elaboró la siguiente tabla.

Tabla 1– Tabla de contingencia para calcular el odds ratio

	Pacientes que desarrollaron infección (casos)	Pacientes que no desarrollaron infección (controles)	
Pacientes con obesidad	a 42	b 55	97
Pacientes sin obesidad	c 28	d 144	172
	70	199	269

a = pacientes con obesidad que desarrollaron infección

b = pacientes con obesidad que no desarrollaron infección

c = pacientes sin obesidad que desarrollaron infección

d = pacientes sin obesidad que no desarrollaron infección

Con estos datos despejó la fórmula mencionada anteriormente.

$$\text{Odds ratio} = \frac{(\mathbf{a}) 42 \times (\mathbf{d}) 144}{(\mathbf{b}) 55 \times (\mathbf{c}) 28}$$

Para complementar los datos utilizó un software de análisis estadístico para obtener el intervalo de confianza a 95% y el valor de p obteniendo los siguientes resultados:

$$\text{Odds ratio} = 3.9$$

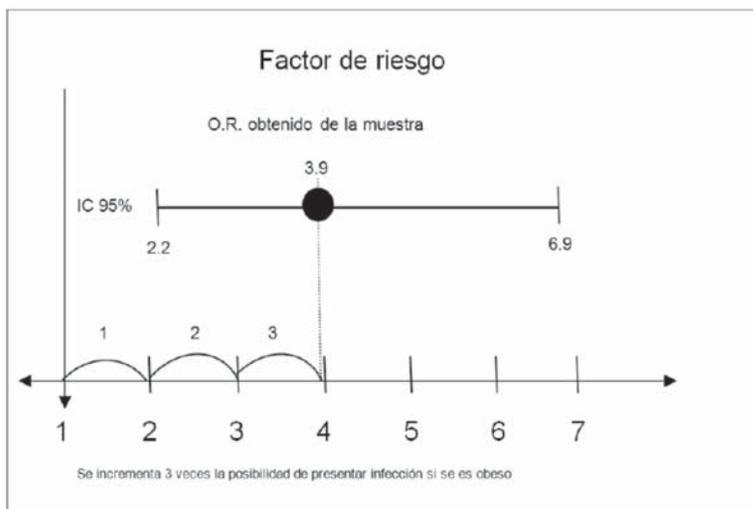
Intervalo de confianza (95%): 2.2 a 6.9

$$p: <0.0001$$

Interpretación

El odds ratio y su intervalo de confianza obtenidos se pueden ubicar gráficamente sobre una línea horizontal que tiene como punto central el número 1 como valor de referencia. (Figura 4).

Figura 4 – Representación gráfica del *odds ratio* obtenido.



Con estos resultados se interpreta que los pacientes obesos tienen 3 veces más de probabilidades de presentar infección en el sitio quirúrgico que los que no presentan obesidad.

El *odds ratio* siempre se expresa en conjunto con su intervalo de confianza, normalmente del 95%. Este intervalo de confianza se entiende entonces que si tomaran al azar 100 pacientes con obesidad el 95% de las veces tendrían un *odds ratio* de entre 2.2 y 6.9. Intervalos de confianza con valores cercanos al *odds ratio* nos refleja que se analizó un número suficiente de sujetos, por el contrario si el resultado tiene valores alejados, nos indica que el número de sujetos estudiados fue muy pequeño.

El valor de *p* nos orienta sobre la probabilidad de que los datos obtenidos se deban al azar, en este caso interpretamos que existe una probabilidad muy baja (0.001) de que estos datos sean debidos al azar.

LECTURA COMPLEMENTARIA

1. Aedo.S.M., Pavlov. S.D, Clavero.F.Ch., Riesgo Relativo y odds ratio ¿Qué son y cómo se interpretan? Rev. Obstétrica Ginecológica: 2010; v. 5: 51-54.
2. Cerda J., Vera C., Rada G., Odds ratio: aspectos teóricos y prácticos, Rev Med Chile 2013; 141: 1329-1335.
3. Clifford B.R., Taylor R.A., Bioestadística, México: Pearson Educación; 2008.
4. Dawson B., Trapp R.G., Bioestadística médica. 4. ed. México: Manual Moderno; 2005.
5. Pérez A. J.M., Principios de bioestadística para comprender los resultados reportados en los artículos científicos, Medigraphic, vol.11 no 2; 2015.
6. Talavera J.O., Rivas.R.R., Pérez M.R., Relevancia Clínica. Rev Med Inst Mex Seguro Soc. 2013; 51 (Supl):S42-S46.
7. Wayne W.D. Bioestadística Base para el análisis de las ciencias de la salud. 4. ed. México: Limusa Wiley; 2011.

Unidad 4

CAPÍTULO 27

Lo que da valor y fuerza a las evidencias

Mohsen Sadeghi Naini

Vafa Rahimi Movaghar

CAPÍTULO 27

LO QUE DA VALOR Y FUERZA A LAS EVIDENCIAS

*Mohsen Sadeghi Naini
Vafa Rahimi Movaghar*

La fuerza tarea canadiense utilizó por primera vez el Nivel de Evidencia (NDE o LoE – *Level of Evidence*) en el examen periódico de salud de 1979, al realizar un relevamiento en la literatura relevante buscando identificar 138 condiciones médicas potencialmente evitables. En aquel informe, la eficacia de las intervenciones médicas se clasificó en tres categorías principales con base en la calidad de las evidencias. (Tabla 1).

Tabla 1 – Nivel de evidencias según la fuerza tarea canadiense.

Eficacia de la intervención	
I	Evidencias obtenidas de por lo menos un estudio controlado aleatorizado.
II-a	Evidencias obtenidas de una cohorte bien delineada o de estudios analíticos de casos y controles, preferencialmente provenientes de más de un centro o grupo de investigaciones.
II-b	Evidencias obtenidas de comparaciones entre tiempos o lugares con o sin intervención. Resultados considerables de experimentos no-controlados (tales como los resultados de la introducción de la penicilina en la década de 1940) también podrían considerarse como ese tipo de evidencias.
III	Opiniones de autoridades respetadas, basadas en experiencia clínica, estudios descriptivos, o informes de comités de especialistas.

Aquí, la calidad de las evidencias se definió con relación a tres conceptos principales:

1. Estudios controlados aleatorizados (ECAs), o revisiones sistemáticas de muchos ECAs, generalmente brindan más apoyo de evidencias que los estudios observacionales.
2. Estudios clínicos comparativos ofrecen más apoyo de evidencias que el razonamiento “mecanístico” (“fundamentación fisiopatológica”) de ciencias más básicas.
3. Estudios clínicos comparativos ofrecen apoyo más fuerte de evidencias que una opinión especializada.

Las evidencias se tomaron en cuenta para puntuar las recomendaciones de 1 a 5. (Tabla 2).

Tabla 2 – Puntaje de recomendaciones de la fuerza tarea canadiense

Clasificación de recomendaciones	
A	Hay buenas evidencias para apoyar la recomendación de que la condición debe ser específicamente considerada en un examen de salud periódico.
B	Hay evidencias regulares para apoyar la recomendación de que la condición debe ser específicamente considerada en un examen de salud periódico.
C	Hay escasas evidencias sobre la inclusión de la condición en un examen de salud periódico, y se pueden hacer recomendaciones con base en otros motivos.
D	Hay evidencias regulares para apoyar la recomendación de que la condición debe ser excluida de consideración en un examen de salud periódico.
E	Hay buenas evidencias para apoyar la recomendación de que la condición sea excluida de consideración en un examen de salud periódico.

Este informe se actualizó con pequeñas modificaciones durante los años siguientes. Desde entonces, el NDE ha sido ampliamente utilizado y criticado por diferentes grupos de trabajo para construir su propio esquema. Cada esquema se delineo para responder diferentes preguntas para diferentes interesados, incluso médicos clínicos, investigadores y legisladores. Entre los procesos de evolución del NDE, una importante medida fue la que implementó el centro de medicina basada en evidencias (CEBM – *Center for Evidence Based Medicine*), en septiembre de 2000, que aceleró el proceso de encuentro de evidencias por el médico clínico.

En aquella época, un problema común a muchos de los esquemas de ordenamiento de evidencias era que el puntaje de las evidencias se refería solamente a terapia y prevención, sin incluir pruebas diagnósticas, marcadores pronósticos o daños.

El equipo del CEBM actualizó el NDE proponiendo la inclusión de diferentes aspectos de problemas de salud, incluso terapia, prevención, etiología, pronóstico, diagnóstico, diagnóstico diferencial y análisis económico y de decisión. (Tabla 3).

Tabla 3 – Nivel de evidencias propuesto por el CEBM 2009.

Nivel	Terapia / Prevención, Etiología / Daños	Pronóstico	Diagnóstico	Diagnóstico diferencial/estudio de prevalencia de síntomas	Análisis económico y de decisión
1a	Revisión sistemática (con homogeneidad) de ECAs	Revisión sistemática (con homogeneidad) de estudios de cohorte de incepción; "CDR (<i>critical design review</i>)" validada en diferentes poblaciones	Revisión sistemática (con homogeneidad) de estudios diagnósticos de Nivel 1; CDR" con estudios 1b de diferentes centros clínicos	Revisión sistemática (con homogeneidad) de estudios de cohorte prospectivos	Revisión sistemática (con homogeneidad) de estudios económicos de Nivel 1
1b	ECA individual (con Intervalo de Confianza estrecho)	Estudios de cohorte de incepción individuales con un > 80% de seguimiento; CDR validado en una única población	Validación de estudios de cohorte con buenos estándares de referencia; o CDR probado en un centro clínico	Estudios de cohorte prospectivos con buen seguimiento	Análisis basado en costos clínicamente sensibles o alternativas; revisión(es) sistemática(s) de la evidencia; incluso análisis de sensibilidad multi-direccionales
1c	Todos o ninguno	Serie de casos con todos o ninguno	SpRims y SnNouts absolutos	Serie de casos con todos o ninguno	Análisis absolutos de mejor valor o peor valor
2a	RS (con homogeneidad) de estudios de cohorte	RS (con homogeneidad) de o estudios de cohorte retrospectivos o grupos de control no tratados en ECAs	RS (con homogeneidad) de estudios diagnósticos de Nivel >2	RS (con homogeneidad) de estudios de 2b y mejores	RS (con homogeneidad) de estudios económicos de Nivel >2

Continuación...

2b	Estudio de cohorte individual (incluso ECA de baja calidad ; p.ej. ; <80% de seguimiento)	Estudio de cohorte retrospectivo o seguimiento de pacientes de control no tratados en un ECA ; Derivación de CDR" o validado en la muestra dividida solamente	Estudio de cohorte exploratorio con buenos estándares de referencia; CDR" tras derivación o validado apenas con muestra dividida o bancos de datos	Estudio de cohorte retrospectivo, o seguimiento pobre	Análisis basada en costos clínicamente sensibles o alternativas; revisiones) limitada(s) de las evidencias o estudios únicos; e inclusión de análisis de sensibilidad multi-direccional
2c	Investigación de "desenlaces"; Estudios ecológicos	Investigación de "desenlaces"		Estudios ecológicos	Auditoría o desenlaces de investigación
3a	RS (con homogeneidad) de estudios de casos y controles		RS (con homogeneidad) de estudios 3b y mejor	RS (con homogeneidad) con estudios 3b y mejor	RS (con homogeneidad) de estudios 3b e mejor
3b	Estudio individual de casos e controles		Estudio no consecutivo; o sin estándares de referencia aplicados consistentemente	Estudio de cohorte no consecutivo o población muy limitada	Análisis basado en alternativas limitadas o costos, estimativas de baja calidad datos, pero incluyendo análisis de sensibilidad incorporando variaciones clínicamente sensibles
4	Series de casos (y estudios y cohorte y casos-control de baja calidad)	Series de casos (y estudios de cohorte de baja calidad pronóstica)	Estudio de casos y controles, estándar de referencia baja o no-dependiente	Series de casos o estándares de referencia suplantados	Análisis sin análisis de sensibilidad
5	Opinión especializada sin evaluación crítica explícita, o basada en fisiología, investigación de laboratorio o "primeros principios",	Opinión especializada sin evaluación crítica explícita, o basada en fisiología, investigación de laboratorio o "primeros principios",	Opinión especializada sin evaluación crítica explícita, o basada en fisiología, investigación de laboratorio o "primeros principios"	Opinión especializada sin evaluación crítica explícita, o basada en teoría económica o "primeros principios"	Opinión especializada sin evaluación crítica explícita o basada en teoría económica o "primeros principios"

Las primeras jerarquías de evidencias (incluso el informe de la CEBM en 2000), se introdujeron principalmente para evaluar la calidad de las evidencias para terapéutica. Esas jerarquías colocaban a los ECAs categóricamente por encima de estudios observacionales y juicio especializado por debajo de observaciones no controladas que hacían la clasificación de evidencias sencilla y fácil de usar.

En años recientes, este enfoque ha sido criticado por su sencillez. Por ejemplo, algunas reacciones adversas a drogas son tan convincentes que un informe subjetivo bien documentado puede ofrecer evidencias definitivas de una asociación causal y no se necesita ninguna otra confirmación. Además, la capacidad de mostrar que pacientes que sufrían de una enfermedad considerada fatal antiguamente, pueden sobrevivir tras un nuevo tratamiento, establece evidencias suficientes acerca de su eficacia.

Algunos ejemplos históricos de tratamientos con efectos espectaculares originados de estudios observacionales, incluyen:

- Insulina para diabetes.
- Transfusión de sangre para choque hemorrágico grave.
- Desfibrilación para fibrilación ventricular.
- Reducción cerrada y colocación de talas para fractura de huesos largos con desplazamiento.
- Neostigmina para miastenia grave.
- Traqueostomía para obstrucción de la tráquea.
- Suturas para la reparación de grandes heridas.
- Estreptomina para meningitis tuberculosa.
- Drenaje para dolor asociado a abscesos.
- Presión o sutura para interrumpir una hemorragia.
- Éter para anestesia.
- Válvula unidireccional o drenaje para pneumotórax y hemotórax.

Por otro lado, nadie puede saltar de un avión con y sin paracaídas para alcanzar evidencias de nivel 1 acerca del papel que ejerce el paracaídas.

Aunque la actual opinión sea a favor de considerar todos los tipos de evidencias disponibles (incluso estudios observacionales) al realizar recomendaciones, también se debe recordar que, ya que el pronóstico y el efecto del tratamiento interactúan con ruido y señal, no todos los efectos del tratamiento son tan fáciles de demostrar.

Considerando esos puntos débiles, esfuerzos más recientes evitan la abyección común de elevar la categoría de evidencias simplemente por el diseño del estudio (ensayo clínico versus estudios observacionales) y pasan a permitir la elevación del NDE de estudios observacionales con efectos espectaculares. Esta alteración en la interpretación de la calidad de las evidencias ha sido otra evolución en el NDE.

En 2006, el Grupo de Trabajo de Puntaje de Recomendaciones, Evaluación y Desarrollo (GRADE-Grading of Recommendations, Assessment, Development and Evaluations) propuso un enfoque claro y sistemático a la realización de juicios sobre la calidad de evidencias y fuerza de recomendaciones. GRADE clasifica la calidad de un cuerpo de evidencias en cuatro niveles: alta calidad, calidad moderada, baja calidad y calidad muy baja.

La calidad de evidencias en el enfoque GRADE puede cambiar en lo que se refiere a la delineación del estudio y riesgo de sesgo, inconsistencia de los resultados, hecho de ser indirecto (no generalizable), imprecisión (datos escasos), y otros factores (por ejemplo, sesgo de informe). Este enfoque se vuelve popular rápidamente gracias a una revisión sistemática, y ahora es opinión general que se trata del método más efectivo de vincular la evaluación de la calidad de la evidencia a recomendaciones clínicas.

El enfoque GRADE evalúa las evidencias, no estudio por estudio, sino transversalmente a los estudios para desenlaces clínicos específicos. Este enfoque evalúa específicamente cuatro dominios en los estudios incluidos de las revisiones sistemáticas, evaluaciones y lineamientos de tecnología de la salud, incluso:

- Fallas metodológicas dentro de los estudios componentes.
- Consistencia de resultados transversalmente a diferentes estudios.
- Posibilidad de generalizar más ampliamente resultados de investigación a base de pacientes.
- Eficacia presentada por los tratamientos.

Sin embargo, lo que GRADE ha agregado en términos de precisión puede haberse perdido en sencillez. Por lo tanto, su utilización podrá volverse imposible para clínicos muy atareados que buscan las mejores evidencias disponibles para un problema clínico específico. Un conflicto importante entre GRADE y OCEBM es que el enfoque GRADE evalúa la fuerza de las evidencias en lugar de estudios aislados, mientras que

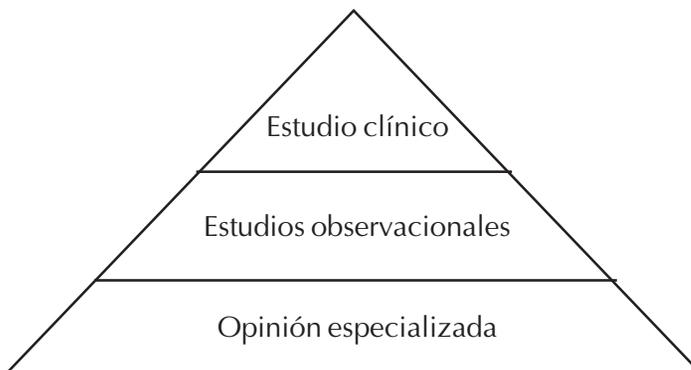
los Niveles de OCEBM enfocan a su público específico (clínicos o pacientes), y adecuan su herramienta para evaluación de estudios individuales cuando no se dispone de una revisión sistemática.

NDE

Los Niveles de Evidencia explican la idea de que hay grados o niveles variables de evidencias para una relación específica de “causa y efecto” en una era científica determinada en la que se pueden ofrecer diferentes tipos **diseños de estudio**. En otras palabras, cuando hablamos de NDE, enfocamos la clasificación del diseño metodológico de estudios que evalúan los efectos de las intervenciones en la atención a la salud. En el mundo real es raro que un único estudio demuestre una hipótesis de manera tan irrefutable que sea imposible cuestionarla. Generalmente, diferentes estudios y diferentes niveles y /o tipos de evidencias contribuyen a la construcción final del cuerpo de evidencias.

Tradicionalmente, el NDE se presenta como una jerarquía de peso que se refiere a la fuerza de la asociación entre la causa y el efecto incorporado específicamente por aquel estudio. (Figura 1).

Figura 1 – Jerarquía del nivel de evidencias



Los estudios con animales constituyen el primer paso para proporcionar evidencias médicas que llevarán a la generación y comprobación de hipótesis. Eso también proporcionará la infraestructura para desarrollar un razonamiento mecanístico para las reglas comúnmente aceptadas.

Las opiniones especializadas sin evaluación crítica o basadas en razonamiento clínico, investigaciones de laboratorio o “primeros principios” brindan el segundo nivel de evidencias.

Los estudios no controlados se dividen en observacionales (por ejemplo, un informe de caso y series de casos) y grupos experimentales (por ejemplo, estudios de antes/después y estudio aleatorizado N-de-1) según el tipo de intervención.

Los estudios observacionales con grupos de control incluyen estudios de cohorte, estudios con casos y controles y estudios transversales, donde el investigador describe lo que ocurre naturalmente, sin perturbar el escenario natural. Este tipo de estudio es más comúnmente usado para anotar asociaciones entre eventos y generar hipótesis.

Los estudios experimentales con grupos de control más comúnmente conocidos como estudios (o ensayos) clínicos (incluso aleatorizados o no aleatorizados) intervienen activamente al asignar a algunos participantes para recibir un tratamiento y otros para recibir una intervención diferente (o ninguna intervención). Obviamente, los estudios experimentales han agregado más valor al proceso de construir el cuerpo de evidencias.

Meta-análisis y revisiones sistemáticas, por otro lado, resumen grupos de estudios en un tema específico de investigación, para evaluar el peso de las evidencias sobre ese tema determinado. (Tabla 4).

Tabla 4 – Taxonomía de Delineación de Estudios que evalúan los Efectos de Intervenciones de Atención a la Salud

Revisión sistemática	Revisión sistemática con meta-análisis Revisión sistemática sin meta-análisis
Estudios experimentales con grupo de control (“ensayos clínicos” o “ensayos”): El investigador controla la decisión referente a la asignación de participantes a los diferentes grupos de intervención	Estudio Controlado Aleatorizado (ECA) Estudio controlado casi aleatorizado (ECA-Q) Estudio clínico controlado (ECC)
Estudios observacionales con grupo de control. La intención del investigador es observar y no interferir en los cuidados rutinarios	Estudio de cohorte Caso de estudio de cohorte de supervivencia Estudio de casos y controles Estudio transversal
Estudios no controlados (sin un grupo de control separado): pueden ser de naturaleza experimental u observacional	Estudios de antes/después Estudio aleatorizado N-de-1 Serie de casos Informes de caso
Opinión especializada	
Estudios con animales	

Además de varios puntos positivos en el aspecto jerárquico de OCEBM-NDE, se plantean algunas suposiciones importantes que generalmente no son válidas, pero que raramente se debaten (por ej. en intervenciones complejas):

- **Equilibrio (Equipoise):** Paciente y clínico no tienen preferencia por un tratamiento.
- **Falta de conocimientos:** realmente se desconoce la “mejor” de dos posibilidades y son inadecuadas las evidencias sobre efectos de tratamiento provenientes de otras fuentes.
- **Preferencia por especificidad:** Son terapéuticamente válidos solamente algunos efectos específicos atribuibles a la intervención.

- **Independencia de contexto:** Hay una extensión “verdadera” de eficacia, o un efecto de tamaño constante, independiente del contexto.
- **Validad ecológica y externa:** Los conocimientos sobre el efecto terapéutico extraído de un ECA son prontamente atribuidos a la práctica clínica, si los criterios de elegibilidad del estudio están de acuerdo con las características de determinado paciente.

La alternativa a la visión jerárquica del nivel de las evidencias es circular. Deriva de la experiencia de que todo método de investigación tiene puntos fuertes y limitaciones que no se pueden fijar por aquel método en sí. Por lo tanto, triangular el resultado alcanzado con un método repitiéndolo con otros métodos podrá llevar a una conclusión más potente.

Tipos de sesgos y herramientas para evaluar el sesgo

La fuerza de las evidencias puede ser amenazada por sesgos (el error sistemático) y por factores de confusión.

Los sesgos pueden desviar el resultado en cualquiera de dos direcciones (subestimación y sobreestimación del verdadero efecto de la intervención), con diferencias de grandeza. Generalmente es imposible medir el impacto de sesgos sobre los resultados de determinado estudio.

El sesgo se puede confundir con imprecisión. El sesgo es un error sistemático que no cambia por medio de replicaciones múltiples de la misma metodología. La imprecisión se eleva a un error aleatorio, lo que significa una variación del muestreo y puede llevar a oscilaciones de los resultados en múltiples replicaciones de la misma metodología, aunque todas esas replicaciones brinden la respuesta correcta en promedio. Podemos tratar la imprecisión aumentando el tamaño de la muestra, pero los sesgos no tienen ese efecto. La imprecisión se refleja en el intervalo de confianza alrededor del resultado analizado, por lo tanto, resultados con más precisión llevan a un intervalo de confianza más estrecho y viceversa.

Los sesgos se clasifican en cinco clases principales. (Tabla 5).

Tabla 5 – El esquema de clasificación de Cochrane para sesgo

Tipo de sesgo	Descripción	Dominios relevantes en el instrumento Cochrane ‘Riesgo de sesgo’
Sesgo de selección	Diferencias sistemáticas entre características de los grupos que se comparan.	<ul style="list-style-type: none"> • Generación de secuencias. • Ocultación de asignación.
Sesgo de performance	Diferencias sistemáticas entre grupos cuanto a los cuidados proporcionados o a la exposición a factores que no son los de las intervenciones de interés.	<ul style="list-style-type: none"> • Cegamiento de los participantes y del personal. • Otras amenazas potenciales a la validez.
Sesgo de detección	Diferencias sistemáticas entre grupos por la manera como se determinan los desenlaces.	<ul style="list-style-type: none"> • Cegamiento de evaluación de desenlace. • Otras amenazas potenciales a la validez.
Sesgo de atrición	Diferencias sistemáticas entre grupos en retiradas de un estudio	<ul style="list-style-type: none"> • Datos de desenlace incompletos.
Sesgo de informe	Diferencias sistemáticas entre hallazgos relatados y no relatados.	<ul style="list-style-type: none"> • Informe selectivo de desenlaces.

Hasta ahora, se han propuesto tres tipos de herramientas para evaluar el sesgo en la investigación clínica, específicamente en estudios clínicos: escalas, listas de verificación y evaluaciones basadas en dominios. Moher y colegas identificaron 34 herramientas que se habían usado para evaluar sesgos en estudios aleatorizados. La herramienta recomendada por la Colaboración Cochrane para evaluar el riesgo de sesgo es una evaluación basada en dominios, en la cual las evaluaciones críticas son separadas para diferentes dominios. Es difícil demostrar la extensión del sesgo en un estudio determinado, por lo tanto, es limitada la posibilidad de validar cualquier herramienta propuesta. (Tabla 6).

Tabla 6 – La Herramienta de la Colaboración Cochrane para evaluar el riesgo de sesgo

Dominio	Soporte para juicio	Revisión del juicio de los autores
<i>Sesgo de selección</i>		
Generación de secuencia aleatoria	Describa el método usado para generar la secuencia de asignación con detalles suficientes para permitir una evaluación mostrando si se deben producir grupos comparables.	Sesgo de selección (asignación sesgada a las intervenciones) debido a la generación inadecuada de una secuencia aleatorizada.
Ocultación de asignación	Describa el método utilizado para ocultar la secuencia de asignación con detalles suficientes para determinar si asignaciones de intervención podrían haberse previsto antes o durante el reclutamiento.	Sesgo de selección (asignación sesgada a las intervenciones) debido a la ocultación inadecuada de asignaciones antes de asignar a las personas.
<i>Sesgo de performance</i>		
Cegamiento de los participantes y de personal. Deben hacerse evaluaciones para cada desenlace principal (o clase de desenlaces)	Describa todas las medidas utilizadas, si las hay, para cegar a los participantes y personal involucrado en el estudio al respecto del conocimiento de cuál intervención recibe el participante. Brinde toda la información que reporta la eficacia o no del cegamiento objetivado.	Sesgo de performance debido al conocimiento de intervenciones asignadas por participantes y personal durante el estudio.
<i>Sesgo de detección</i>		
Cegamiento de la evaluación de desenlace. <i>Deben hacerse evaluaciones para cada desenlace principal (o clase de desenlaces)</i>	Describa todas las medidas usadas, si las hay, para cegar a los evaluadores del desenlace al conocimiento de qué intervención recibió cada participante. Brinde toda la información que reporta la eficacia o no del cegamiento objetivado.	Sesgo de detección debido al conocimiento de las intervenciones alocadas por evaluadores de desenlace.
<i>Sesgo de atrición</i>		
Datos incompletos sobre desenlaces <i>Deben hacerse evaluaciones para cada desenlace principal (o clases de desenlaces).</i>	Describa cuán completos son los datos de desenlaces para cada desenlace principal, incluso atrición y exclusiones del análisis. Mencione si la atrición y las exclusiones se han relatado, números en cada grupo de intervención (comparado a número total de participantes aleatorizados), motivos por los cuáles se reportaron atriciones/exclusiones, y cualquier reinclusión en los análisis realizados por los autores de revisiones.	Sesgo de atrición debido a la cantidad, naturaleza o manipulación de datos incompletos sobre desenlaces.
<i>Sesgo de informe</i>		
Informe selectivo	Mencione como se ha analizado la posibilidad de informe de desenlace selectivo por los autores de revisiones y qué se ha encontrado.	Sesgo de informe debido a informe selectivo de desenlace.
<i>Otros sesgos</i>		
Otras fuentes de sesgo	Mencione cualquier preocupación importante sobre sesgo no discutida en los demás dominios de la herramienta. Si determinadas preguntas/ anotaciones se han pre-especificado en el protocolo de la revisión, deben responderse cada pregunta/anotación.	Sesgo debido a problemas no tratados en otra parte de la tabla.

¿Qué puede violar la fuerza de las evidencias?

Hay factores en potencial que pueden reducir la calidad del cuerpo de evidencias derivadas del análisis de datos reunidos de un ECA, como propuesto en el enfoque GRADE. Esos puntos evalúan diferentes tipos de sesgo en este diseño de estudio específico para luego juzgar con base en la fuerza global de evidencias derivadas de los estudios clínicos incluidos. Por lo tanto, para un ECA este enfoque se describe como el bloque fundamental del proceso de producción de evidencias.

1. Limitaciones en el diseño e implementación: El diseño de cada estudio tiene algunas limitaciones potenciales que se deberían considerar al incorporar su resultado a un ambiente clínico o a una pregunta de investigación. Para estudios aleatorizados, las limitaciones potenciales serían:

- Ocultación de asignación.
- Cegamiento (especialmente cuando los desenlaces se evalúan subjetivamente).
- Pérdida del seguimiento.
- Interrupción temprana para beneficio o relatos selectivos de los desenlaces.

2. Evidencias indirectas: Al usar los resultados del estudio, los criterios de elegibilidad seguidos en aquel estudio deben evaluarse y compararse con el actual escenario clínico. Por ejemplo, una versión restricta de pregunta principal en términos de población, intervención, comparador o desenlaces puede producir evidencias indirectas. Otras fuentes de información indirecta pueden surgir de intervenciones estudiadas (por ejemplo, un estudio relevante muestra una intervención quirúrgica implementada por un cirujano sénior en un centro especializado, lo que podrá traer efectos indirectos de los efectos de la intervención cuando realizada fuera de centros especializados), de comparadores usados (por ejemplo, si la intervención en los grupos de control es menos eficaz que el tratamiento estándar), y de desenlaces evaluados (por ejemplo, indirectos debido a desenlaces alternativos).

3. Heterogeneidad inexplicada o inconsistencia de resultados: Estimativas con grandes diferencias sobre el efecto (heterogeneidad,

en otras palabras), sin explicación razonable, podrían reducir la calidad de las evidencias.

4. Imprecisión de resultados: Intervalos de confianza amplios, independiente de la heterogeneidad, pueden llevar a la reducción de la calidad de las evidencias. Eso se puede tratar con el aumento del tamaño de la muestra.

5. Alta probabilidad de sesgo de publicación: La calidad del nivel de evidencias se puede reducir si los investigadores no relatan estudios (clásicamente aquellos que no presentan algún efecto: sesgo de publicación) o desenlaces (clásicamente aquellos que pueden ser perjudiciales o para los cuales no se ha relatado ningún efecto: sesgo de informe de desenlace selectivo), con base en resultados. La gran influencia de auspiciadores en un estudio hace surgir la pregunta sobre la existencia de estudios no publicados que sugieran la ausencia de beneficios.

LECTURA COMPLEMENTARIA

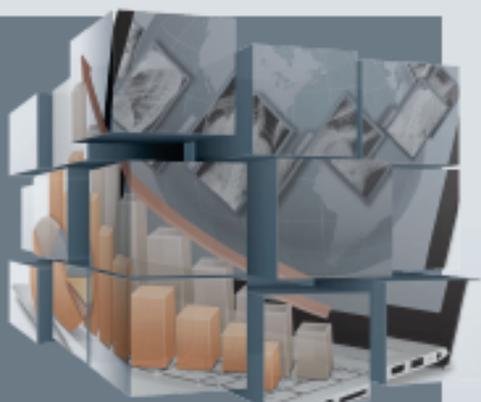
1. Hill N, Frappier-Davignon L, Morrison B. The periodic health examination. *Can Med Assoc J.* 1979;121:1193-254.
2. Force CT, Examination PH. The periodic health examination 2. 1984 update. 1984;130.
3. Examination CTF on the PH. The periodic health examination: 2. 1987 update. *TT - . C Can Med Assoc J.* 1988;138(7):618-26.
4. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. Vol. 9, *Chinese Journal of Evidence-Based Medicine.* 2009. p. 8-11.

Agradecimiento

Agradecemos el trabajo de edición de la lengua inglesa realizado por la Sra. Bitá Pourmand, afiliada al Research Development Center, Sina Hospital, Tehran University of Medical Sciences.

5. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest*. American College of Chest Physicians; 1986; 89(2): 25-35.
6. Judith Welsh, BSN M. Levels of evidence and analyzing the literature. National Institutes of Health Library.
7. Howick J, Chalmers I, Glasziou P, Greenhalgh T, Heneghan C, Liberati A, et al. Explanation of the 2011 Oxford Center for Evidence-Based Medicine (OCEBM) Levels of Evidence (Background Document). *Oxford Cent Evidence-Based Med*. 2011;(12):1-8.
8. Force USPST. Guide to clinical preventive services. DIANE Publishing; 1989.
9. Howick J. *The Philosophy of Evidence-Based Medicine. The Philosophy of Evidence-Based Medicine*. 2011.
10. Aronson JK, Hauben M. Anecdotes that provide definitive evidence. *BMJ Br Med J*. 2006;333(7581):1267-1269.
11. How to read clinical journals: V: To distinguish useful from useless or even harmful therapy. *Can Med Assoc J*. 1981;124(9):1156-1162.
12. Lee MJ, Norvell DC, Dettori JR. *SMART Approach to Spine Clinical Research*. Thieme; 2013.
13. Glasziou P. When are randomised trials unnecessary/? Picking signal from noise Paul Glasziou and colleagues When are randomised trials unnecessary/? Picking signal from noise. 2016; (May).
14. Schünemann HJ, Oxman AD, Vist GE, Higgins JPT, Deeks JJ, Glasziou P GG. Interpreting results and drawing conclusions. In: Higgins JPT GS, editor. *Cochrane Handbook for Systematic Reviews of Interventions Version 510 (updated March 2011)*. The Cochrane Collaboration; 2011.
15. Walach H, Falkenberg T, Fønnebø V, Lewith G, Jonas WB. Circular instead of hierarchical/? methodological principles for the evaluation of complex interventions. 2006;9:1-9.
16. Higgins JPT, Altman DG SJ (editors). Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT GS, editor. *Cochrane Handbook for Systematic Reviews of Interventions Version 510 (updated March 2011)* [Internet]. The Cochrane Collaboration; 2011. Available from: <www.cochrane-handbook.org>.
17. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials*. Elsevier; 1995;16(1):62-73.

Cuando tenemos clara nuestra pregunta, la cual nos lleva a realizar nuestra investigación, contaremos con los resultados, mismos que nos permitirán fortalecer nuestras conclusiones mediante la Inferencia y para ello debemos de conocer los aspectos teóricos relacionados con la Bioestadística Clínica, de donde se derivará que obtengamos conceptos que nos permita tomar decisiones y aplicar el conocimiento obtenido a nuestros pacientes y cumplir con excelencia el papel de prevenir y tratar las enfermedades, incorporando los conceptos a la práctica clínica.



De la Medición a la Interpretación, un proceso esencial que vincula la investigación al mundo cotidiano y a la realidad, en donde nuestras hipótesis rodeadas de probabilidades buscan encontrar el valor que establezca, si existen diferencias o no, entre las opciones.

No olvidar que los números son crudos y no entienden de razones, es ahí donde nuestro conocimiento y la teoría le dará el valor real a los datos, ya que estos no nos da certeza de ningún evento, simplemente nos indica la probabilidad que algo pueda suceder.

José María Jiménez Avila
Regional Research Officer-AOSpine Latin America

Asdrubal Falavigna
Chairperson-AOSpine Latin America

ISBN 978-85-7061-911-2



9 788570 619112