



Estadística en salud

Los números no son
solo para los matemáticos

Brayan Esneider Patiño Palma
Diego Fernando Afanador Restrepo



Estadística en salud

Los números no son
solo para los matemáticos





Estadística en salud

Los números no son
solo para los matemáticos



Brayan Esneider Patiño Palma

Diego Fernando Afanador Restrepo

Patiño Palma, Brayan Esneider

Estadística en salud : los números no son solo para los matemáticos / autor Brayan Esneider Patiño Palma , Diego Fernando Afanador-Restrepo -- Bogotá: Fundación Universitaria del Área Andina, 2023

ISBN (Digital) 978-958-5139-95-4

Incluye índice.

1. Estadística médica. - 2. Educación - estadísticas. - 3. Indicadores de condiciones de salud. - 4. Medicina - Casos, historias clínicas, estadísticas.

Catalogación en la publicación Biblioteca Fundación Universitaria del Área Andina (Bogotá)

616 scdd22

ESTADÍSTICA EN SALUD. LOS NÚMEROS NO SON SOLO PARA LOS MATEMÁTICOS

© Fundación Universitaria del Área Andina. Bogotá, diciembre de 2023

© Brayan Esneider Patiño Palma, Diego Fernando Afanador Restrepo

ISBN (digital): 978-958-5139-95-4

FUNDACIÓN UNIVERSITARIA DEL ÁREA ANDINA

Calle 71 No. 13-21, Bogotá, Colombia

Correo electrónico:

publicaciones@areandina.edu.co

Proceso editorial

Dirección editorial:

Omar Eduardo Peña Reina

Coordinación editorial:

Camilo Andrés Cuéllar Mejía

Diseño de carátula:

Proceditor Ltda.

Diagramación:

Proceditor Ltda.

Calle 1C No. 27A-01, Bogotá, Colombia

Tel.: 757 9200

Correo electrónico: proceditor@yahoo.es

Todos los derechos reservados. Queda prohibida la reproducción total o parcial de esta obra y su tratamiento o transmisión por cualquier medio o método sin autorización escrita de la Fundación Universitaria del Área Andina y sus autores.

BANDERA INSTITUCIONAL

Pablo Oliveros Marmolejo †

Gustavo Eastman Vélez

Miembros Fundadores

Carlos Patricio Eastman Barona

Presidente de la Asamblea General y Consejo Superior

José Leonardo Valencia Molano

Rector Nacional y Representante Legal

Martha Patricia Castellanos Saavedra

Vicerrectora Nacional Académica

Karol Milena Pérez Calderón

Vicerrectora Nacional de Crecimiento y Desarrollo

Erika Milena Ramírez Sánchez

Vicerrectora Nacional Administrativa y Financiera

Mauricio Andrés Hernández Anzola

Vicerrector Nacional de Experiencia y Felicidad

Felipe Baena Botero

Rector - Seccional Pereira

Gelca Patricia Gutiérrez Barranco

Rectora - Sede Valledupar

María Angélica Pacheco Chica

Secretaria General



Omar Eduardo Peña Reina

Director Nacional de Investigaciones

Jackeline Marín Tapias

Decana Facultad de Ciencias de la Salud y del Deporte

Seccional Pereira

Leidy Maritza Obando Ante

Directora programa de Fisioterapia

Seccional Pereira

Camilo Andrés Cuéllar Mejía

Subdirector Nacional de Publicaciones



Autores	11
---------	----

Prólogo	13
---------	----

Capítulo 1. Conceptos básicos de estadística
15

Introducción	17
Conceptos previos	17
Introducción al SPSS (Statistical Package for the Social Sciences)	22

Capítulo 2. Estadística descriptiva
25

Tabla de distribución de datos	27
Medidas descriptivas	28
Medidas de tendencia central	29
Medidas de dispersión	32
Medidas de ubicación	36
Medidas de forma	37
Gráficos descriptivos en salud	40

Capítulo 3. Estadística inferencial. Generalidades
43

Técnicas inferenciales. Consideraciones previas	45
-------------------------------------------------	----

Estimación por intervalos	46
Contrastando hipótesis sobre parámetros poblacionales	49
Muestreo e inferencia estadística	50
Muestreo probabilístico o aleatorio	50
Muestreo no probabilístico	52
Determinación de tamaños muestrales	54
Distribuciones muestrales	57

Capítulo 4. Contraste de hipótesis
63

Errores en el contraste de hipótesis	68
Contrastes bilaterales y unilaterales	71
Contraste de hipótesis sobre una proporción	74

Capítulo 5. Pruebas estadísticas
75

Prueba Chi-cuadrado	77
Prueba t para muestras independientes	79
Prueba t para muestras relacionadas	81

Comparativa de tres o más grupos independientes (ANOVA)	83
ANOVA de efectos fijos y aleatorios	84
Coefficiente de determinación	85
Inferencia y tabla ANOVA	87
Pruebas post hoc	91
Coefficiente de correlación de Pearson	92

Capítulo 6. Pruebas no paramétricas

95

Prueba de rangos de signos de Wilcoxon	97
Prueba de McNemar	100
Prueba de Kruskal-Wallis	103

Estadístico Jonckheere-Terpstra	106
Estadístico U de Mann-Whitney	109
Test de Friedman	112
Test de Q de Cochran	115

Capítulo 7. Revisión sistemática y metaanálisis

117

Revisión sistemática	119
Proceso de búsqueda de artículos	119
Extracción de la información	120
Metaanálisis	121

Referencias

127

Lista de tablas

Contenido

Tabla 1.1	<i>Tipos de variables</i>	19	Tabla 6.3	<i>Tabla descriptiva test Wilcoxon</i>	99
Tabla 2.1	<i>Tabla distribución de frecuencias</i>	27	Tabla 6.4	<i>Tabla de asociación o 2 x 2</i>	101
Tabla 2.2	<i>Diferentes tipos de medias</i>	30	Tabla 6.5	<i>Descripción de la muestra</i>	107
Tabla 3.1	<i>Resumen tipos de muestreo probabilístico</i>	52	Tabla 6.6	<i>Descripción de la muestra ordenada</i>	107
Tabla 3.2	<i>Resumen tipos de muestreo no probabilístico</i>	54	Tabla 6.7	<i>Descripción de la muestra con rangos</i>	107
Tabla 4.1	<i>Tipos de errores en la estadística</i>	70	Tabla 6.8	<i>Diferencia de rangos</i>	108
Tabla 5.1	<i>Distribución de frecuencias de respuesta al tratamiento según sexo</i>	77	Tabla 6.10	<i>Descripción de eventos</i>	113
Tabla 5.2	<i>Frecuencias esperadas bajo la hipótesis nula</i>	78	Tabla 6.11	<i>Descripción de eventos ordenado</i>	113
Tabla 5.3	<i>Datos nivel obesidad y nivel de colesterol</i>	86	Tabla 6.12	<i>Descripción de eventos con rangos</i>	114
Tabla 5.4	<i>Tabla ANOVA</i>	89	Tabla 7.1	<i>Extracción de la información para metaanálisis</i>	120
Tabla 5.5	<i>Tabla ANOVA para el ejemplo 5.1</i>	90	Tabla 7.2	<i>Tamaño del efecto de cada estudio y tamaño del efecto medio</i>	125
Tabla 6.1	<i>Notas estudiantes</i>	98			
Tabla 6.2	<i>Diferencia entre las medidas relacionadas</i>	99			

Lista de figuras

Figura 1.2	<i>Pasos y representación gráfica de la estadística</i>	18	Figura 5.1	<i>Diagrama de dispersión de una variable cuantitativa (y) según el grupo de pertenencia (G1, G2 o G31)</i>	84
Figura 1.2	<i>Población y muestra</i>	22	Figura 5.2	<i>Diferencias en las medias en función de la muestra seleccionada</i>	88
Figura 2.1	<i>Conceptualización de la desviación estándar</i>	33	Figura 5.3	<i>Coefficiente de correlación de Pearson</i>	93
Figura 3.1	<i>Diferentes tipos de distribuciones muestrales</i>	59	Figura 7.1	<i>Funnel Plot con heterogeneidad baja</i>	122
Figura 3.2	<i>Gráfico Q-Q</i>	60	Figura 7.2	<i>Funnel Plot con heterogeneidad alta</i>	123
Figura 3.3	<i>Gráficos de densidad e histograma</i>	61	Figura 7.3	<i>Forest Plot resultado del metaanálisis</i>	125
Figura 4.1	<i>Pruebas de hipótesis</i>	67			
Figura 4.2	<i>Aplicación regla de decisión</i>	68			
Figura 4.3	<i>Contraste de hipótesis unilateral</i>	72			
Figura 4.4	<i>Contraste de hipótesis unilateral</i>	73			

Autores

Brayan Esneider Patiño Palma

Fisioterapeuta de la Fundación Universitaria del Área Andina, Especialista en Estadística Aplicada de la Fundación Universitaria los Libertadores, Magíster en Actividad Física y Deporte y estudiante del Doctorado en Ciencias de la Salud de la Universidad Autónoma de Manizales. Docente de tiempo completo e investigador del grupo de investigación ZIPATEFI de la Fundación Universitaria del Área Andina, seccional Pereira.

Diego Fernando Afanador-Restrepo

Estudiante de primer año de Doctorado en Cuidados Integrales y Servicios en Salud, Universidad de Jaén. Miembro del subcomité de programa de la Especialización en Terapia Manual Ortopédica de la Universidad de La Sabana. Docente de tiempo completo e investigador del grupo de investigación ZIPATEFI de la Fundación Universitaria del Área Andina.

Prólogo

Es un placer presentarles este libro escrito por mi hermano, un apasionado de la investigación y la escritura. Desde muy joven, ha demostrado un gran interés por el conocimiento y la exploración de nuevas ideas. Su curiosidad y dedicación lo han llevado a convertirse en un experto en su campo de estudio y a escribir varios artículos y libros sobre temas diversos.

Este libro es una muestra más de su compromiso, en sus páginas encontrarán una obra bien documentada y rigurosa, que aborda temas sobre “Estadística para profesionales de la salud”, tema de gran relevancia para nuestra comunidad. Su estilo claro y conciso, su capacidad para explicar conceptos complejos de manera sencilla y su habilidad para mantener al lector interesado y comprometido, son una muestra de su destreza como escritor.

La presente obra tiene como objetivo brindar una visión general de los métodos cuantitativos y descriptivos utilizados en el campo de la salud, lo que es un derecho fundamental y una condición esencial para el bienestar de las personas y las comunidades. Por lo tanto, es fundamental contar con herramientas y técnicas que nos permitan analizar y comprender los datos relacionados con la salud.

Esta obra está dirigida a estudiantes, profesionales de la salud y cualquier persona interesada en comenzar a dirimir de una manera sencilla y clara el análisis de datos en el campo de la salud. En sus páginas se aborda la gran necesidad de utilizar los datos y estadísticas para mejorar la calidad de la atención y evaluar la efectividad de los tratamientos.

Esperamos que esta obra sea de gran utilidad para todos aquellos que buscan mejorar la salud y el bienestar de las personas a través de la medicina y la fisioterapia.

Agradezco a mi hermano por su dedicación y compromiso con la investigación y la escritura y espero que este libro sea una fuente de inspiración para todos aquellos que buscan explorar nuevos horizontes y expandir su conocimiento.

José Diocidez Patiño Palma

Magíster en Gestión de la Tecnología Educativa

Directivo Docente

CAPÍTULO 1. CONCEPTOS BÁSICOS DE ESTADÍSTICA



Introducción

En la introducción de esta sección se resalta la importancia de la estadística como herramienta crucial para el análisis de datos en la investigación, ya que permite la síntesis y descripción de información sobre características específicas de individuos o elementos en una población y a partir de estos poder inferir o llegar a conclusiones sobre toda una población. Además, se explican algunos conceptos previos necesarios para entender las nociones y técnicas de análisis que se presentarán en el resto de capítulos, los cuales son el objetivo principal de este trabajo.

Conceptos previos

Estadística

La estadística es una rama de las matemáticas que se encarga de recopilar, organizar, analizar, interpretar y presentar datos. Su objetivo principal es proporcionar una comprensión cuantitativa de los fenómenos naturales y sociales que nos rodean, mediante el uso de métodos estadísticos.

La estadística se aplica en una amplia variedad de campos, como la ciencia, la medicina, la economía, la ingeniería, la psicología, la educación, la política, entre otros, para estudiar y analizar datos y tomar decisiones informadas basadas en evidencia (Figura 1.1).

En el caso particular de la medicina y en general en las ciencias de la salud, la estadística toma el nombre de bioestadística, la cual es definida como la aplicación de la estadística en el campo de la biología y la medicina. Esta se utiliza para analizar y comprender los datos biológicos y de salud, hacer inferencias y tomar decisiones basadas en la información recolectada. La bioestadística juega un papel importante en la investigación biomédica, en la evaluación de tratamientos y en la planificación de programas de salud pública. Su objetivo principal es la aplicación de métodos estadísticos para extraer información de los datos, y así proporcionar una base sólida para la toma de decisiones en el ámbito de la salud.



Figura 1.2

Pasos y representación gráfica de la estadística

Nota. Elaboración propia.

Variables

En estadística, una variable se refiere a cualquier característica o propiedad que pueda medirse o cuantificarse en los individuos o elementos de una población. En otras palabras, una variable es una característica que puede variar de un individuo a otro y puede tomar distintos valores.

Por ejemplo, la edad de una persona, su género, el nivel de estudios, el ingreso anual, la cantidad de horas de sueño, la presión arterial, son todos ejemplos de variables que pueden ser medidas y cuantificadas. Es importante destacar que las variables pueden clasificarse en diferentes tipos según la naturaleza de los valores que toman. La comprensión de la clasificación de las variables en estadística es fundamental para el análisis y la interpretación de los datos. El conocimiento de los tipos de variables y su naturaleza permite una mejor comprensión de los resultados y una elección más acertada de las herramientas y técnicas estadísticas adecuadas para su análisis. Por ejemplo, para las variables cualitativas

se pueden utilizar técnicas de análisis de frecuencia y proporciones, mientras que para las variables cuantitativas se pueden utilizar medidas de tendencia central y de dispersión, y se pueden realizar análisis de correlación y regresión, análisis que se abordarán más adelante.

Tipos de variables

Las variables cualitativas, no tienen valores numéricos y pueden ser clasificadas en diferentes categorías o estados. Este tipo de variables pueden ser ordinales o no ordinales, dependiendo de si sus categorías pueden ser ordenadas o no. Por ejemplo, el sexo es una variable cualitativa no ordinal, mientras que el nivel de educación (calificado como primaria, secundaria, pregrado, postgrado) es una variable cualitativa ordinal.

Por otra parte, las variables cuantitativas sí tienen valores numéricos, estas pueden ser discretas o continuas, dependiendo de si sus valores son finitos o infinitos y numerables o no numerables. Por ejemplo, el número de hijos o ingresos en un hospital son variables cuantitativas discretas, ya que solo pueden tomar valores enteros específicos. En cambio, el nivel de colesterol, la glicemia o el nivel de ácido úrico son variables continuas, ya que pueden tomar cualquier valor dentro de un intervalo de números infinitos (Tabla 1.1).

Tabla 1.1

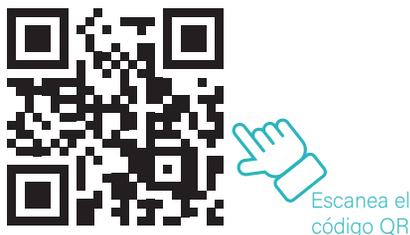
Tipos de variables

Tipo de variable	Descripción	Ejemplos
Cualitativa nominal	Los valores no tienen un orden lógico y no son numéricos.	Género, color de ojos, país de origen.
Cualitativa ordinal	Los valores tienen un orden lógico, pero no se puede determinar la distancia entre ellos.	Nivel educativo (primaria, secundaria, universidad), grado de satisfacción (muy satisfecho, satisfecho, insatisfecho).
Cuantitativa discreta	Los valores son numéricos enteros y finitos.	Número de hijos, número de mascotas, cantidad de empleados en una empresa.
Cuantitativa continua	Los valores son numéricos y pueden tomar cualquier valor en un rango determinado.	Edad, altura, peso, ingresos, temperatura.

Nota. Elaboración propia.

Durante la lectura de esta obra se presentarán numerosos ejemplos y ejercicios que se explican detalladamente por medio de videos que se encuentran en diversas listas de

reproducción disponibles en YouTube. A continuación, se presenta un video en el cual se complementa el tema visto anteriormente.



Estadística descriptiva

La estadística descriptiva es una rama de la estadística que se enfoca en la recolección, organización, análisis y presentación de datos con el objetivo organizar, describir y resumir las características de una población o muestra.

Las técnicas utilizadas en la estadística descriptiva incluyen:

- Medidas de tendencia central: son medidas que indican el valor central de un conjunto de datos. Las más comunes son la media, la mediana y la moda.
- Medidas de dispersión: son medidas que indican la variabilidad de un conjunto de datos. Las más comunes son el rango, la desviación estándar y el coeficiente de variación.
- Tablas de frecuencia: son tablas que muestran la distribución de frecuencias de una variable, es decir, el número de veces que aparece cada valor o rango de valores en un conjunto de datos.
- Gráficos: son representaciones visuales de los datos que permiten visualizar patrones y tendencias en los mismos. Los más comunes son el histograma, el diagrama de barras, el diagrama de sectores y el diagrama de caja y bigotes.

La estadística descriptiva es útil para resumir y presentar información de manera clara y concisa, lo que facilita su comprensión y toma de decisiones basadas en los datos.

Estadística inferencial

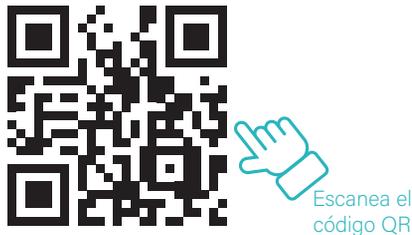
La estadística inferencial es una rama de la estadística que se ocupa de hacer inferencias o conclusiones acerca de una población a partir de los datos obtenidos de una muestra de esa población. En otras palabras, la estadística inferencial se utiliza para hacer generalizaciones sobre una población a partir de los resultados obtenidos en una muestra representativa de la misma.

Para hacer inferencias sobre la población, se utilizan técnicas de estimación y pruebas de hipótesis. La estimación se utiliza para obtener un valor aproximado de un parámetro poblacional desconocido a partir de los datos de la muestra. Por ejemplo, si se quiere

conocer la media de edad de una población, se puede obtener una estimación de esta medida a partir de la media de edad de una muestra representativa.

Por otro lado, las pruebas de hipótesis se utilizan para determinar si una afirmación acerca de la población es verdadera o no. Por ejemplo, si se afirma que la media de edad de una población es mayor a cierto valor, se puede realizar una prueba de hipótesis para determinar si hay suficiente evidencia en la muestra para respaldar dicha afirmación.

A continuación, presentamos un video que complementa el tema abordado previamente.



Población y muestra

En estadística, una población se refiere a un conjunto completo de individuos, objetos, eventos o medidas que comparten una o varias características comunes de interés. La medida usada para esa característica toma el nombre de *parámetro*. Por ejemplo, si estamos interesados en estudiar el peso promedio (Media = Parámetro) de todos los recién nacidos en un país, entonces la población estaría compuesta por todos los recién nacidos de ese país.

Por otro lado, una muestra es un subconjunto seleccionado de la población que se utiliza para estimar las características de la población. La muestra debe ser representativa de la población, es decir, debe ser seleccionada de tal manera que refleje las mismas características que la población de interés. La selección de la muestra se realiza con el objetivo de reducir el costo, el tiempo y la complejidad de la investigación. La medida usada para el análisis de la o las características toma el nombre de *estadístico*.

Por ejemplo, si queremos estudiar el peso promedio (media = estadístico) de todos los recién nacidos en un país, no es posible medir el peso de todos los recién nacidos del país. En su lugar, podemos seleccionar una muestra representativa de recién nacidos y medir su peso para luego usar esta información para hacer una inferencia sobre el peso promedio de la población de recién nacidos del país.

Es importante destacar que la selección de la muestra debe ser realizada de forma cuidadosa, ya que la elección inadecuada de la muestra puede llevar a resultados incorrectos o poco precisos.

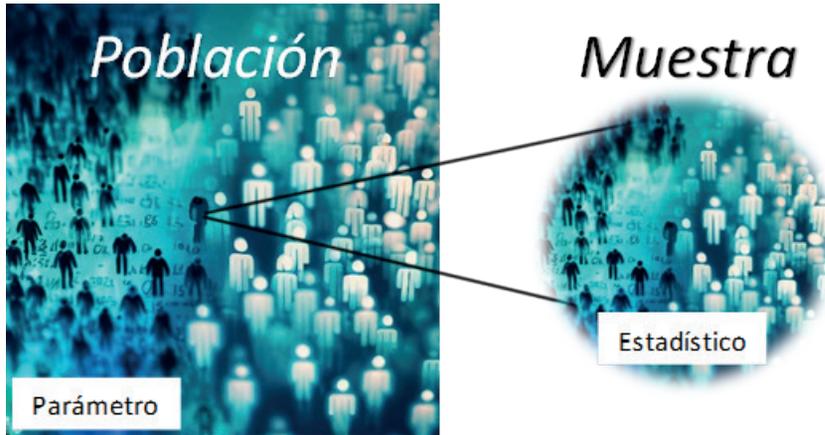


Figura 1.2

Población y muestra

Nota. Elaboración propia.

Introducción al SPSS (Statistical Package for the Social Sciences)

El SPSS (Statistical Package for the Social Sciences) es un software estadístico que ha sido utilizado durante décadas en diversas áreas, desde la investigación y la psicología hasta la medicina y la economía. SPSS es una herramienta muy versátil que permite la manipulación, análisis y presentación de datos de manera eficiente y efectiva.

Una de las principales ventajas de SPSS es su facilidad de uso. A través de una interfaz gráfica intuitiva, los usuarios pueden realizar tareas como la importación de datos, la limpieza y preparación de datos, y la realización de análisis estadísticos complejos sin necesidad de conocimientos avanzados en programación.

Otra gran ventaja de SPSS es su capacidad para manejar grandes conjuntos de datos. SPSS puede importar y analizar datos de una variedad de formatos, incluyendo hojas de cálculo, bases de datos y archivos de texto. También permite el manejo de grandes conjuntos de datos sin afectar la velocidad y el rendimiento del software.

SPSS es un software estadístico muy completo que ofrece una amplia gama de técnicas estadísticas y herramientas de análisis, incluyendo análisis descriptivos, pruebas de hipótesis, análisis de regresión, análisis de correlación, análisis de supervivencia, entre otros. Además, SPSS ofrece la posibilidad de crear gráficos y tablas para presentar los resultados de manera visualmente atractiva.

Durante la lectura de esta obra se presentarán numerosos ejemplos y ejercicios que se explican detalladamente a través de videos en diversas listas de reproducción disponibles en YouTube. En este caso en particular, se dispone de una serie de videos que muestran la interfaz y los primeros pasos del uso del software SPSS.



Escanea el
código QR

CAPÍTULO 2. ESTADÍSTICA DESCRIPTIVA



Tabla de distribución de datos

La técnica de resumen de datos consiste en clasificar las diversas observaciones según sus categorías correspondientes y luego representarlas en forma de tabla. En el caso de variables cualitativas, las categorías de agrupación son determinadas por la propia variable, aunque se podrían realizar agrupaciones diferentes para obtener los resultados deseados. A estas tablas se les llama “tablas de distribución de frecuencias”. Este tipo de técnica es muy útil para sintetizar grandes cantidades de datos y obtener una visión general de las tendencias y patrones presentes en los datos.

Ejemplo 1.1

Supongamos que un profesional del área de la salud está interesado en analizar la presencia de dolor en sus pacientes y ha recopilado información sobre el tipo de dolor que padecen. Los posibles valores son “dolor de espalda”, “dolor de cuello” y “otro tipo de dolor”. Los resultados de las observaciones se muestran en la Tabla 2.1.

Tabla 2.1

Tabla distribución de frecuencias

Tipo de dolor	Frecuencia	Frecuencia relativa	Frecuencia acumulada	Frecuencia relativa acumulada
Dolor de espalda	15	0.30	15	0.30
Dolor de cuello	8	0.16	23	0.46
Otro tipo de dolor	27	0.54	50	1.00
Total	50	1.00		

Nota. Elaboración propia.

Para la construcción de la tabla, el primer paso es identificar los valores posibles de la variable cualitativa. En este caso, los valores posibles son “dolor de espalda”, “dolor de cuello” y “otro tipo de dolor”. Posteriormente, se cuenta cuántos pacientes hay en cada categoría. Para ello, se revisan las respuestas de los pacientes y se asignan a la categoría

correspondiente. Por ejemplo, si un paciente reporta que tiene dolor de espalda, se le asigna a la categoría “dolor de espalda”.

Una vez que se ha asignado cada paciente a su categoría correspondiente, se cuenta cuántos pacientes hay en cada categoría. Por ejemplo, si hay 15 pacientes que reportan dolor de espalda, se anota “15” en la columna de frecuencias correspondiente a esa categoría, tal y como aparece en la tabla.

La frecuencia relativa se calcula dividiendo la frecuencia de cada categoría entre el total de observaciones, en este caso, el total de pacientes. Por ejemplo, para la categoría “dolor de espalda”, la frecuencia relativa se calcula dividiendo $15/50 = 0.294$, esto equivaldría al 29% de la muestra. De esta manera, la frecuencia relativa representa la proporción de observaciones que pertenecen a cada categoría.

La frecuencia acumulada se calcula sumando las frecuencias de cada categoría y acumulándolas hacia abajo. Por ejemplo, la frecuencia acumulada para la categoría “dolor de espalda” es simplemente su frecuencia, 15. La frecuencia acumulada para la categoría “dolor de cuello” es la suma de la frecuencia de esa categoría y la frecuencia acumulada de la categoría anterior, es decir, $8 + 15 = 23$. La frecuencia acumulada para la categoría “otro tipo de dolor” es la suma de su frecuencia y la frecuencia acumulada de las dos categorías anteriores, es decir, $27 + 8 + 15 = 50$; mismo procedimiento que se lleva a cabo para determinar la frecuencia relativa acumulada $0.54 + 0.16 + 0.30 = 1.00$.

Con esta información, el profesional puede entender mejor la distribución de la variable de estudio y posterior a su análisis puede tomar decisiones más informadas sobre sus tratamientos. En este sentido, siguiendo la idea del ejemplo inicial el profesional puede observar que más de la mitad de los pacientes reportan otro tipo de dolor que no sea de espalda o cuello, lo que podría indicar la necesidad de ampliar su conocimiento sobre otros tipos de dolor y sus estrategias de tratamiento.

Medidas descriptivas

Las tablas de distribución de frecuencias son una herramienta útil para resumir la información proporcionada por los datos de una variable, en especial en el caso de variables cualitativas o categóricas. Estas tablas muestran la frecuencia con la que cada valor de la variable aparece en los datos y pueden ser utilizadas para identificar patrones y tendencias.

Sin embargo, para el análisis de variables cuantitativas se debe profundizar en otro tipo de medidas de resumen de datos. Las medidas de tendencia central, como la media, la mediana y la moda, proporcionan información sobre la localización de los datos y ayudan a entender en qué punto se concentran los valores de la variable. Por otro lado, las medidas de dispersión, como la desviación estándar y el rango intercuartil, permiten conocer la variabilidad de los datos y cuán dispersos están alrededor de la medida de tendencia central.

También es importante mencionar que existen medidas de forma, como la asimetría y la curtosis, que pueden ser utilizadas para conocer la forma de la distribución de los datos. Todo lo anterior se describirá a continuación.

Medidas de tendencia central

Las medidas de tendencia central ofrecen datos acerca de la posición o ubicación de los datos observados. Algunas de las medidas más comunes son la media, la mediana y la moda. Para ejemplificar el cálculo de estas medidas, presentamos el siguiente ejemplo:

Ejemplo 2.1

Supongamos que deseamos analizar la edad de los estudiantes de una universidad en particular. Tomamos una muestra aleatoria de 20 estudiantes y registramos sus edades en años:

22, 19, 21, 24, 25, 19, 20, 23, 22, 20, 21, 19, 25, 24, 22, 23, 20, 22, 21, 23

Media

La media aritmética es una de las medidas de tendencia central más empleadas en el análisis de datos. Esta medida representa el promedio de los datos y se obtiene a partir de la inclusión de todos los datos observados en su cálculo, siguiendo la siguiente fórmula:

$$x = \frac{\sum x_i}{n} = \frac{20+19+21+24\dots+23}{20} = 21,9$$

Donde $x_1, x_2, x_3, \dots, x_n$ son las observaciones de la variable y n el número total de observaciones. La media aritmética es una medida de tendencia central que proporciona una ubicación general del conjunto de datos. Sin embargo, hay que tener en cuenta que la media es sensible a valores extremos u observaciones atípicas. Si un valor se encuentra significativamente alejado del resto, afectará considerablemente al valor de la media. En este caso, la media se desplazará hacia el valor atípico, y no representará adecuadamente la distribución de los datos. Para resolver este problema, se pueden utilizar medidas robustas, que otorgan menos peso a las observaciones alejadas, permitiendo una mejor representación del conjunto de datos. Por tanto, es importante evaluar cuidadosamente la presencia de valores atípicos en los datos antes de utilizar la media como medida de tendencia central.

Tabla 2.2
Diferentes tipos de medias

Tipo de medida	Fórmula	Características	Ejemplo
Media aritmética	$x = \frac{\sum x_i}{N}$	Sensible a valores atípicos	La media aritmética de 10, 12, 15, 18 y 22 $x = \frac{\sum x_i}{N} = \frac{10+12+15+18}{5} = 15,4$
Media geométrica	$G = \sqrt[n]{\prod x_i}$	Se utiliza para calcular la tasa de crecimiento promedio, la tasa de retorno promedio o la tasa de variación proporcional en una serie de datos.	Si un activo genera ganancias de 10% el primer año, 25% el segundo año y el tercer año el 15%. ¿Cuál es la ganancia promedio? $G = \sqrt[3]{0,10 \cdot 0,25 \cdot 0,15} = 0,15$ En conclusión, la ganancia del activo es de 15%
Media armónica	$H = \frac{n}{\sum \frac{1}{x_i}}$	Se utiliza para calcular la tasa media aritmética de las tasas individuales, como la velocidad media.	Calcular la velocidad promedio de 3 ciclistas que recorren la misma distancia (10 kilómetros) sabiendo que el ciclista A va a 10 km/h, el ciclista B va a 20 km/h y el ciclista C va a 40 km/h $H = \frac{3}{\frac{1}{10} + \frac{1}{20} + \frac{1}{40}} = 17,14 \text{ km/h}$
Media cuadrática	$MC = \sqrt{\frac{\sum x_i^2}{n}}$	Es útil para describir datos que tienen grandes variaciones entre los valores, como los precios de las acciones en el mercado de valores.	Se evidencian los siguientes errores en la medida del peso de 3 personas. Evaluador 1: 2 kg Evaluador 2: -4 kg Evaluador 3: 3 kg Determinar el error promedio $MC = \sqrt{\frac{2^2 + (-4)^2 + 3^2}{3}} = 3,1$ El promedio de error en la medición es de 3.10 kg.
Media ponderada	$Xp = \frac{\sum x_i \cdot w_i}{\sum w_i}$	Se utiliza cuando los diferentes datos tienen diferentes pesos o importancias en la distribución.	Determinar el promedio ponderado de las calificaciones de un estudiante de acuerdo a sus notas. Actividad Nota Peso Consulta 31/50 10% Taller 45/50 20% Examen 23/50 70% $Xp = \frac{(31 \cdot 0,1) + (45 \cdot 0,2) + (23 \cdot 0,7)}{0,1 + 0,2 + 0,7} = 28/50$

Nota. Elaboración propia.

Mediana

Otra medida de tendencia central que se utiliza como alternativa al cálculo de la media es la mediana, ya que no se ve afectada por observaciones atípicas o extremas. La mediana se define como el valor que divide a un conjunto de datos en dos partes iguales, dejando el mismo número de observaciones a su izquierda y a su derecha.

Es importante destacar que, aunque podría haber infinitos valores que cumplan este requisito para un conjunto de observaciones, se utiliza una fórmula estándar para calcular la mediana, lo que garantiza que esta medida será única para el conjunto de datos. En primer lugar, es necesario ordenar los datos de menor a mayor.

19, 19, 19, 20, 20, 20, 21, 21, 21, 22, 22, 22, 22, 23, 23, 23, 24, 24, 25, 25

Dado que el conjunto de números es par, debe calcularse el valor promedio de los datos centrales. En general, se calculará en primer lugar el rango de la mediana, que informará sobre la posición que debe ocupar esta, una vez ordenados los datos de menor a mayor, de la siguiente forma:

$$m = \frac{n+1}{2} = \frac{20+1}{2} = 10.5$$

La mediana sería un valor entre el dato que ocupa la posición 10 y el dato que ocupa la posición 11, que en este caso corresponde a los valores 22 y 22.

La mediana se obtendrá entonces:

$$m = \frac{22+22}{2} = 22$$

En caso de incluir una observación adicional (la número 26) al conjunto original de $n=21$ observaciones, entonces se obtendría lo siguiente:

19, 19, 19, 20, 20, 20, 21, 21, 21, 22, 22, 22, 22, 23, 23, 23, 24, 24, 25, 25, 26

Donde

$$m = \frac{n+1}{2} = \frac{21+1}{2} = 11$$

En esta situación, como la cantidad de datos es impar, existe únicamente un valor ubicado en el centro, lo que genera una distribución simétrica de la misma cantidad de datos hacia la izquierda y la derecha del valor central, que corresponde a la posición 11 y que cuenta con diez valores a su izquierda y diez a su derecha.

Cabe destacar que, si se añadiera un valor extremo como 88 años, el valor de la mediana permanecería inalterable, lo cual no sucede con la media. Además, en el cálculo de la mediana solo intervienen uno o dos datos directamente con su valor, mientras que todos

los demás datos lo hacen indirectamente a través de su orden. Por consiguiente, es posible inferir que, al emplear la mediana, se pierde parte de la información que los datos proporcionan en comparación con la media.

Moda

La moda es una medida de tendencia central que consiste en el valor más común en un conjunto de datos, es decir, aquel que se repite con mayor frecuencia. Al aplicarla al ejemplo anterior, se observa que varios valores tienen una alta frecuencia de aparición en el conjunto de datos. Por lo tanto, en este caso se tendrían diferentes modas. Cabe destacar que la moda es una medida adecuada para datos discretos o agrupados en intervalos, pero puede no existir o no ser única en datos continuos.

Este resultado demuestra que la moda no necesariamente es única para un conjunto de datos. Además, la moda puede ser cualquier valor que se repita más que cualquier otro valor en el conjunto de datos, aunque no sea una medida resumen confiable y se considera la medida más débil hasta el momento. En el caso de variables cuantitativas continuas, es común encontrar frecuencias bajas en la mayoría de los valores observados, por lo que una alternativa para calcular la moda es agrupar los datos en intervalos y encontrar el intervalo o intervalos con mayor frecuencia absoluta, que se llama intervalo modal.

Medidas de dispersión

Las medidas de tendencia central permiten conocer la ubicación de los datos, pero no proporcionan información sobre su variabilidad o dispersión. En la Figura 2.1 se muestran dos conjuntos de datos con la misma media y mediana, pero con distribuciones diferentes. En el caso (1), los datos están más dispersos alrededor de la media o mediana, mientras que en el caso (2) están más concentrados. Por lo tanto, es necesario contar con medidas que indiquen la dispersión de los datos y permitan distinguir entre diferentes situaciones.

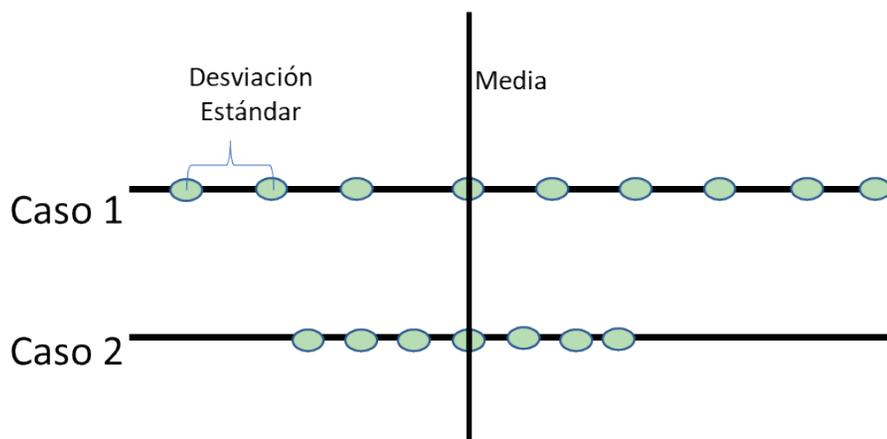


Figura 2.1

Conceptualización de la desviación estándar

Nota. Elaboración propia.

Rango o recorrido

El rango es una medida simple y fácil de entender para medir la dispersión de los datos. Se calcula encontrando la diferencia entre el valor más alto y el más bajo observado en el conjunto de datos. En el ejemplo 2-1, el rango se puede obtener de la siguiente manera:

$$r = X_{max} - X_{min} = 5 - 19 = 6$$

El rango muestra que la distancia entre el dato mayor y menor es de 6 años; sin embargo, es importante destacar que en la construcción del rango solo intervienen 2 de los datos observados, los más extremos. Esto hace que el rango sea extremadamente sensible a observaciones extremas y no tenga en cuenta gran parte de la información disponible. Por lo tanto, es recomendable utilizar medidas de dispersión más robustas, que proporcionen una visión más completa de la variabilidad de los datos.

Varianza y desviación típica o estándar

La desviación estándar y la varianza son medidas estadísticas que se utilizan para describir la dispersión o variabilidad de un conjunto de datos.

La varianza es una medida de la dispersión que indica cuánto se alejan los datos de su valor medio o promedio. Matemáticamente, se calcula como la media de las desviaciones

al cuadrado de cada valor respecto a su valor medio. La fórmula para calcular la varianza se expresa como:

$$S^2 = \frac{\sum (x_i - \bar{X})^2}{n-1}$$

Donde x_i corresponde a cada valor del conjunto de datos, \bar{X} al promedio del conjunto de datos y n es el número de datos.

Con los datos del ejemplo anterior quedaría:

$$S^2 = \frac{\sum (x_i - \bar{X})^2}{n-1} = \frac{(19-21,4)^2 + (20-21,4)^2 \dots (25-21,4)^2}{20-1} = 9,26 \text{ años}^2$$

La varianza se expresa en unidades al cuadrado de la variable, por lo que puede resultar difícil de interpretar. Para obtener una medida en las mismas unidades que la variable original, se extrae la raíz cuadrada de la varianza, lo que da lugar a la denominada desviación típica o estándar. Esta medida representa la dispersión de los datos en torno a la media y se expresa en las mismas unidades que la variable original, lo que facilita su interpretación y comparación con otros datos.

$$S = \sqrt{S^2} = \sqrt{9,26} = 3,04 \text{ años}$$

Coeficiente de variación

La desviación típica es una medida resumen de las distancias de cada dato a la media en las mismas unidades que la variable original, lo que significa que depende de dichas unidades de medida. Por esta razón, no es posible comparar directamente las desviaciones típicas de dos conjuntos de datos que tienen unidades de medida diferentes.

Además, la relación entre la desviación típica y la dispersión de los datos no es siempre directa. Aunque en general se puede afirmar que, a mayor desviación típica, mayor es la dispersión de los datos, esto solo es posible si los conjuntos de datos que se pretenden comparar tienen la misma media. En caso contrario, dos conjuntos de datos con diferentes medias y la misma desviación típica pueden tener diferentes niveles de dispersión.

Ejemplo 2.2

Consideremos dos conjuntos de datos, el primero con los valores 1, 2, 3, 4 y 5, y el segundo con los valores 6, 7, 8, 9 y 10. Ambos conjuntos tienen una desviación típica de 1.4142, pero el primero tiene una media de 3 y el segundo una media de 8. Es evidente que estos 2 conjuntos de datos no tienen el mismo nivel de dispersión, ya que el segundo está más alejado de su media que el primero.

Por lo tanto, la desviación típica es una medida útil para resumir la dispersión de un conjunto de datos en relación con su media y en las mismas unidades que la variable original. Sin embargo, no se puede comparar directamente la desviación típica de dos conjuntos de datos con unidades de medida diferentes; además, es necesario tener en cuenta la media de los conjuntos de datos para interpretar correctamente su nivel de dispersión.

Para sortear lo planteado anteriormente existe el coeficiente de variación (CV), la cual es una medida de la variabilidad relativa de un conjunto de datos, que se expresa en términos porcentuales. El CV se calcula dividiendo la desviación estándar de los datos por la media y multiplicando el resultado por 100.

El CV se utiliza principalmente para comparar la variabilidad de 2 o más conjuntos de datos que tienen diferentes unidades de medida o diferentes escalas. Al expresar la variabilidad en términos relativos, el CV permite comparar conjuntos de datos que pueden tener medias y desviaciones estándar muy diferentes.

Un CV bajo indica que la variabilidad de los datos es relativamente pequeña en comparación con la media, mientras que un CV alto indica que la variabilidad de los datos es relativamente grande en comparación con la media. Por lo tanto, el CV se utiliza para evaluar la estabilidad o consistencia de los datos en un conjunto de mediciones o experimentos.

Para ilustrar esta cuestión se propone el siguiente ejemplo:

Ejemplo 2.3

Supongamos que se midió la edad y el nivel de actividad física de una muestra de 50 pacientes y se obtuvieron los siguientes datos:

Edad: 42, 44, 40, 39, 41, 37, 43, 38, 36, 45, 46, 47, 43, 44, 39, 38, 36, 41, 42, 43, 44, 45, 46, 47, 39, 38, 41, 42, 43, 44, 45, 46, 47, 39, 38, 41, 42, 43, 44, 45, 46, 47, 39, 38, 41, 42, 43, 44, 45, 46, 47

Nivel de actividad física: 120, 90, 180, 75, 110, 240, 135, 200, 90, 120, 180, 300, 60, 90, 150, 210, 120, 130, 180, 100, 120, 270, 80, 90, 150, 190, 150, 90, 100, 180, 240, 170, 70, 90, 200, 120, 90, 180, 210, 210, 90, 90, 240, 120, 90, 270, 250, 240, 200, 190

A partir de estos datos se determina que la media para la edad es de 42 años y la desviación estándar es 3.11 años. Por otra parte, la media para la variable actividad física es de 152.6 min/sem y la desviación estándar es de 60.09 min/sem. La cuestión es: ¿qué variable presenta una mayor dispersión o variabilidad?

Para determinar esto se hace necesario construir una medida de dispersión relativa que no dependa de las unidades de medida (adimensional). Por lo tanto, trabajando con los datos del ejemplo se tendrá que:

$$CV_{edad} = \frac{3,11}{42} = 0.07 = (7\%)$$

$$CV_{AF} = \frac{60,29}{152,6} = 0.39 = (39\%)$$

Lo anterior pone en manifiesto que la variable edad tiene una variabilidad observada mucho menor que la correspondiente a la variable nivel de actividad física.

Es importante tener en cuenta que el CV no es una medida adecuada cuando la media de los datos es muy cercana a 0, ya que el coeficiente de variación puede volverse infinito en este caso. Además, el CV puede verse afectado por valores extremos o valores atípicos en el conjunto de datos, ya que estos pueden aumentar la desviación estándar y, por lo tanto, el valor del CV. En estos casos, puede ser necesario utilizar otras medidas de variabilidad o realizar análisis más detallados para evaluar la consistencia de los datos.

Medidas de ubicación

Percentiles o cuartiles

Los percentiles o cuartiles son medidas estadísticas que dividen un conjunto de datos ordenados en partes iguales. El cálculo de los percentiles y cuartiles es útil para entender la distribución de los datos y para identificar valores extremos o atípicos.

Los cuartiles son los valores que dividen el conjunto de datos en cuatro partes iguales. El primer cuartil (Q1) es el valor que separa el 25% inferior de los datos del 75% superior, el segundo cuartil (Q2) es el valor que separa el 50% inferior del 50% superior (es decir, es equivalente a la mediana), y el tercer cuartil (Q3) es el valor que separa el 75% inferior del 25% superior.

Los percentiles por su parte son medidas estadísticas que dividen un conjunto de datos ordenados en 100 partes iguales, representando la posición relativa de un valor con respecto a los demás valores del conjunto. Es decir, el percentil P indica que P por ciento de los valores del conjunto son iguales o menores que el valor correspondiente al percentil.

En este sentido y para ejemplificar, si se ordenan las edades de un grupo de 100 personas y se encuentra que la edad del percentil 25 es 30 años, significa que el 25% de las personas en el grupo tienen 30 años o menos.

Ejemplo 2.4

Imaginemos que tenemos el siguiente conjunto de datos de edades de un grupo de personas:

$$\{10, 12, 15, 16, 18, 20, 22, 25, 28, 30, 32, 35, 37, 39, 40\}$$

Para calcular los percentiles de este conjunto de datos, primero debemos ordenarlo de menor a mayor:

{10, 12, 15, 16, 18, 20, 22, 25, 28, 30, 32, 35, 37, 39, 40}

A continuación, se calculará el percentil 30 y 60 de la siguiente manera:

$$r_{pk} = \frac{k}{100}(n+1) = r_{30} = \frac{30}{100}(15+1) = 4,8$$

$$r_{pk} = \frac{k}{100}(n+1) = r_{60} = \frac{60}{100}(15+1) = 9,6$$

El percentil p 30 se obtendrá a partir del valor que se encuentra en la posición 4,8 del conjunto de datos, lo que implica que su valor se situará entre el valor que ocupa la posición 4 y el valor que ocupa la posición 5. De manera similar, el percentil p 60 se calculará a partir del valor que se encuentra en la posición 9,6, lo que significa que su valor estará entre el valor que ocupa la posición 9 y el que ocupa la posición 10. Por lo tanto, para identificar el dato exacto será necesario calcular una media ponderada para obtener el valor final de la siguiente forma:

$$p_{30} = (1-f)x_i + fx_{i+1} = (1-0,8)16 + 0,8(18) = 17,6$$

$$p_{60} = (1-f)x_i + fx_{i+1} = (1-0,6)28 + 0,6(30) = 29,2$$

Donde f es la parte fraccionaria del rango del percentil correspondiente.

Una forma de analizar los resultados obtenidos a partir del cálculo del percentil 60 es observar que el 60% de la muestra presenta edades iguales o inferiores a 29,2 años, mientras que el 40% restante supera esta cifra. En otras palabras, el valor del percentil indica la proporción que separa al conjunto de datos ya sea por debajo o encima de dicho valor.

Medidas de forma

Las medidas de forma son una herramienta útil para analizar la distribución de los datos de una variable. Estas medidas nos permiten comprender cómo se comportan los datos en términos de simetría o apuntamiento.

Coefficiente de asimetría

El coeficiente de asimetría es una medida estadística que nos permite analizar la simetría o asimetría de una distribución de datos. Esta medida se calcula comparando la frecuencia y la magnitud de los valores por encima y por debajo de la media.

Cuando una distribución de datos es simétrica, los valores que están por encima de la media son equivalentes en magnitud y frecuencia a los valores que están por debajo de la

media, lo que indica una distribución equilibrada. El coeficiente de asimetría se expresa en un valor numérico que puede ser positivo o negativo, dependiendo de la dirección de la asimetría. Un valor positivo indica que la cola de la distribución está más sesgada hacia la derecha, mientras que un valor negativo indica que la cola está más sesgada hacia la izquierda.

Ahora bien, para obtener una comprensión inicial de la simetría de una distribución de datos, se puede comparar la media y la mediana. Si la diferencia entre ambas es pequeña, entonces la distribución puede considerarse simétrica. Por otro lado, si la media es significativamente mayor o menor que la mediana, entonces la distribución será asimétrica hacia la derecha o hacia la izquierda, según corresponda. Esta aproximación sencilla es útil para realizar una evaluación preliminar de la simetría de una distribución y puede proporcionar una indicación general sobre la forma en que se agrupan los datos. El coeficiente de asimetría es la medida más sofisticada para el estudio de la simetría y se calcula de la siguiente forma:

$$As = \frac{3(\bar{x} - Me)}{S}$$

Donde \bar{x} corresponde al promedio, Me a la mediana y S a la desviación estándar

Ejemplo 2.5

Supongamos que se tienen los siguientes 15 valores de una variable X :

$$X = \{12, 16, 14, 18, 20, 21, 22, 24, 26, 30, 32, 36, 40, 44, 50\}$$

Entonces:

$$As = \frac{3(\bar{x} - Me)}{S} = \frac{3(26,13 - 24)}{12,34} = 0,59$$

Si el valor del coeficiente As está cerca de 0, la distribución de los datos será aproximadamente simétrica. Si el valor del coeficiente As es superior a 0, la distribución será asimétrica por la derecha, mientras que si es inferior a 0 será asimétrica por la izquierda. En el ejemplo el coeficiente de asimetría resultante es positivo, lo que indica que la distribución es ligeramente asimétrica hacia la derecha. Esto significa que la cola derecha de la distribución es más larga que la cola izquierda y que hay valores extremos más altos.

Curtosis

La curtosis es una medida estadística que describe la forma de una distribución de datos y su concentración alrededor de la media. En otras palabras, indica si una distribución es más o menos “apuntada” que la distribución normal estándar.

La curtosis se calcula comparando la distribución de datos con la distribución normal estándar, que tiene una curtosis de 0. Si la distribución tiene una mayor concentración de datos alrededor de la media que la distribución normal, se dice que tiene una curtosis positiva o “leptocúrtica”. Si la distribución tiene una concentración de datos menor alrededor de la media que la distribución normal, se dice que tiene una curtosis negativa o “platicúrtica”. Si la distribución es similar a la distribución normal, se dice que tiene una curtosis de 0 o “mesocúrtica”.

El cálculo de la curtosis se realiza mediante la siguiente fórmula:

$$Cu = \frac{\sum (X_i - \bar{x})^4}{n s^4} - 3 = \frac{(12 - 26,13)^4 + (26 - 26,13)^4 \dots (50 - 26,13)^4}{15 \cdot 12,34^4} - 3 = -0,259$$

Donde: n es el tamaño de la muestra, xi es el valor de cada observación de la muestra, \bar{x} es la media de la muestra y s es la desviación estándar de la muestra

Por lo tanto:

Si $Cu > 0$ ® La distribución es más apuntada que la normal (leptocúrtica).

Si $Cu = 0$ ® La distribución es igual de apuntada que la normal (mesocúrtica).

Si $Cu < 0$ ® La distribución es menos apuntada que la normal que más se le aproxima (platicúrtica).

En el caso del ejemplo, el valor de la curtosis es ligeramente menor a 0, por lo que el apuntamiento de la distribución es similar al de la normal o ligeramente menos apuntada.

Para poner en práctica los conceptos abordados en este capítulo, recomendamos revisar la siguiente lista de reproducción de YouTube, donde se muestran de manera resumida los conceptos vistos anteriormente y un ejemplo paso a paso de cómo calcular todos los estadísticos descriptivos para una muestra de datos en SPSS. Este recurso puede ser de gran ayuda para consolidar los conocimientos adquiridos en el capítulo y para aplicarlos en el análisis de datos en diversos ámbitos, desde las ciencias sociales hasta las ciencias de la salud.



Escanea el código QR

Gráficos descriptivos en salud

En el ámbito de la salud, la estadística descriptiva se convierte en una herramienta indispensable para comprender y comunicar de manera efectiva la información contenida en conjuntos de datos. Los gráficos descriptivos desempeñan un papel crucial en este proceso, ya que permiten visualizar patrones, distribuciones y comparaciones clave. En este apartado, exploraremos cuatro tipos fundamentales de gráficos descriptivos utilizados comúnmente en el campo de la salud: boxplots, histogramas, gráficos de segmentos y gráficos de barras.

- **Boxplot:** El boxplot, también conocido como diagrama de caja y bigotes, se destaca como una herramienta excepcional para representar la distribución de datos. En el contexto de la salud, un boxplot puede proporcionar una visión rápida de la variabilidad y la presencia de valores atípicos en medidas cruciales, como niveles de glucosa o presión arterial.
- La caja del boxplot representa el rango intercuartílico, ofreciendo una idea clara de la dispersión de los datos. Los “bigotes” indican la variabilidad más allá de este rango, y los puntos fuera de los bigotes pueden señalar valores atípicos. Este tipo de visualización es valioso para identificar patrones y comparar distribuciones entre diferentes grupos, como pacientes con y sin una condición específica.
- **Histogramas:** Los histogramas son esenciales para comprender la frecuencia y la forma de la distribución de datos continuos. En el ámbito de la salud, a manera de ejemplo, un histograma podría representar la distribución de índices de masa corporal en una población específica.
- La construcción del histograma implica dividir el rango de datos en intervalos (o “bins”) y contar la frecuencia de observaciones en cada intervalo. Este tipo de gráfico proporciona una visión clara de la forma de la distribución, identificando modas y asimetrías que pueden tener implicaciones clínicas.
- **Gráfico de barras:** Los gráficos de barras son efectivos para comparar categorías o grupos. En el contexto de la salud, podrían utilizarse para representar la frecuencia de diferentes tipos de enfermedades en una población.
- Estos gráficos proporcionan una representación visual intuitiva de las diferencias entre categorías, permitiendo una fácil interpretación de las disparidades y la prevalencia relativa de diferentes condiciones de salud.
- **Gráficos de tortas o segmentos:** Los gráficos de tortas o segmentos son útiles para representar la proporción de una categoría en relación con el todo. En el ámbito de la salud, podrían utilizarse para mostrar la distribución de diferentes grupos de pacientes en un estudio.

Este tipo de gráfico facilita la visualización de proporciones relativas y destaca la composición porcentual de categorías, proporcionando una representación clara de la distribución de variables categóricas en la muestra analizada.

La selección adecuada y la interpretación precisa de estos gráficos son cruciales para comunicar de manera efectiva los hallazgos estadísticos en el campo de la salud. Estas herramientas visuales no solo simplifican la comprensión de los datos, sino que también capacitan a los profesionales de la salud para tomar decisiones informadas respaldadas por evidencia. Para una exploración más detallada, se ha creado una lista de reproducción en YouTube. En estos videos, cada concepto previamente abordado se explica de manera minuciosa y se ilustra con ejemplos en el software SPSS, ofreciendo una guía completa para la aplicación práctica de estos gráficos en el análisis estadístico en el contexto de la salud.



Escanea el
código QR

CAPÍTULO 3. ESTADÍSTICA INFERENCIAL. GENERALIDADES



En la investigación de fenómenos aleatorios, es común el objetivo de extraer conclusiones sobre características de interés de la población objeto de estudio a partir de una muestra de dicha población. En este proceso de generalización de resultados, conocido como inferencia estadística, la forma en que se obtienen los datos de la muestra es crucial.

Este capítulo se enfoca en los conceptos fundamentales de la inferencia estadística, como las técnicas inferenciales de intervalos de confianza y contrastes de hipótesis, de los cuales el muestreo aleatorio es la base para la ejecución.

En algunos casos, la información disponible solo se limita a ciertos individuos o elementos de la población en cuestión, y esto puede deberse a diversas razones como la destrucción del elemento durante la recopilación de información, la imposibilidad real de observar a todos los individuos de una población infinita o el costo económico exorbitante de entrevistar a toda la población.

Por ejemplo, en el caso de un técnico que desea estimar el tiempo promedio de duración de un lote de prótesis, no tendría sentido observar el tiempo de duración de todas las prótesis del lote, ya que esto requeriría su destrucción. En el caso de un estudio que busca estimar el tiempo promedio de espera en la sala de urgencias de un centro hospitalario, la población estaría conformada por todos los posibles usuarios del servicio de urgencias, lo cual es una población infinita y, por lo tanto, imposible de observar a todos los individuos. En el tercer caso, en el cual se busca comprobar si la proporción de fumadores en España es superior al 45%, la población estaría compuesta por todos los habitantes del país. Aunque teóricamente es posible observar a toda la población, podría resultar inviable económicamente entrevistar a todos los habitantes.

Así pues, en todos estos escenarios, es imprescindible elegir un subgrupo de elementos de la población que se está estudiando y, a continuación, medir la variable relacionada con la característica que se desea investigar. En este mismo sentido, la selección de la muestra debe ser cuidadosa para evitar sesgos y asegurarse de que la muestra represente de manera adecuada la población de interés.

Técnicas inferenciales. Consideraciones previas

Como se mencionó previamente, es común intentar sacar conclusiones acerca de ciertas características o parámetros de la población de estudio con base en la información obtenida de una muestra de la misma. La elección de la técnica estadística adecuada dependerá del tipo de afirmaciones que se desee realizar sobre dichos parámetros desconocidos de la población. Para ejemplificar esta situación, se presentan los siguientes casos:

Ejemplo 3.1

Se dispone de datos de 450 pacientes afectados por una patología específica, provenientes de diferentes hospitales de un municipio colombiano. A partir de esta muestra, se ha obtenido una media de tiempo de estancia de 8 días con una desviación estándar de 1,5 días. Por lo tanto, el objetivo es utilizar esta información para estimar el tiempo medio de estancia en el hospital para toda la población de pacientes afectados por la misma patología.

Ejemplo 3.2

En un estudio sobre la relación entre la altura y el peso en una población de jóvenes adultos, se recolectó información de 100 individuos seleccionados aleatoriamente. La altura promedio fue de 175 cm, con una desviación estándar de 7 cm, mientras que el peso promedio fue de 70 kg, con una desviación estándar de 10 kg. A partir de estos datos, se busca estimar la talla y peso de la población y si existe una relación lineal significativa entre la altura y el peso.

En el párrafo anterior se proponen dos ejemplos en los que se aplican técnicas inferenciales para obtener información sobre una población. En el primer ejemplo, el objetivo es estimar o calcular el valor de un parámetro específico de la población, en este caso, la media del tiempo de estancia en el hospital. En cambio, en el segundo ejemplo, el objetivo es determinar la relación entre dos variables, la talla y el peso, en la población.

Ambas situaciones son puntos de partida para la aplicación de dos técnicas inferenciales distintas. La primera técnica es la estimación, que se utiliza para obtener una estimación del valor del parámetro desconocido de la población. La segunda técnica es el contraste de hipótesis, que se utiliza para determinar si hay una relación significativa entre dos variables en la población.

Estimación por intervalos

Si se quiere medir un parámetro desconocido de una población a partir de una muestra, existen dos posibilidades: 1) proporcionar un solo valor que represente el parámetro, o 2) proporcionar un intervalo que incluya el verdadero valor del parámetro con un cierto nivel de confianza. La diferencia clave entre estos dos enfoques es que, mientras que el enfoque de estimación puntual solo proporciona una aproximación del valor del parámetro sin información sobre su margen de error, el enfoque de intervalos de confianza proporciona información sobre la magnitud probable del parámetro y su margen de error.

Por ejemplo, en un estudio sobre litiasis biliar en un municipio se obtuvo información sobre el nivel promedio de colesterol en 500 individuos, con una media de 226,58 mg/100 ml y una desviación típica de 49,39 mg/100 ml. Si se quisiera conocer si el nivel promedio de colesterol de todos los individuos del municipio es de 210 mg/100 ml, se podría utilizar la media muestral como estimador puntual del parámetro poblacional. Sin embargo,

también se podría construir un intervalo de confianza al 95% para el valor del parámetro poblacional, lo que permitiría conocer la magnitud probable del parámetro y su margen de error. Lo anterior se realizaría de la siguiente forma:

$$IC_{95\%} = \left[\bar{x} - 1,96 \frac{S}{\sqrt{n}} ; \bar{x} + 1,96 \frac{S}{\sqrt{n}} \right] = [222,25; 230,90]$$

Por lo tanto, se puede afirmar con un nivel de confianza del 95%, es decir, con una posibilidad de error del 5% que el nivel promedio de colesterol en la población se situará entre 222,23 mg/100 ml y 230,93 mg/100 ml.

Es importante destacar que, aunque el intervalo de confianza proporciona más información que una estimación puntual, su construcción requiere de un estimador puntual del parámetro poblacional. En general, para construir intervalos de confianza para cualquier parámetro poblacional, es necesario tener un estimador puntual del mismo. La cuestión entonces es determinar cuáles son los estimadores puntuales adecuados para los diferentes parámetros poblacionales de interés, como las medias, proporciones, varianzas, diferencias de medias, etc. Este problema, junto con el método de construcción de intervalos de confianza, se discutirá más adelante.

Intervalo de confianza para una media

Cuando deseamos estimar la media de una población a partir de una muestra, utilizamos un intervalo de confianza para tener en cuenta la variabilidad inherente de los datos muestrales. Un intervalo de confianza es un rango de valores dentro del cual es probable que se encuentre el verdadero valor del parámetro poblacional con un cierto nivel de confianza.

La construcción de un intervalo de confianza para la media se basa en la distribución muestral de la media, que sigue una distribución normal si la muestra es lo suficientemente grande (según el teorema del límite central). El intervalo de confianza se calcula utilizando la siguiente fórmula:

$$\text{Media de la muestra} \pm (\text{Valor crítico (z)} * \text{Error estándar})$$

Donde:

Media de la muestra: es el promedio de los datos en la muestra.

Valor crítico: Es un valor obtenido de una tabla de valores críticos de la distribución normal estándar, que depende del nivel de confianza deseado. Por ejemplo, para un nivel de confianza del 95%, el valor crítico es aproximadamente 1.96.

Error estándar: Es una medida de la variabilidad de los datos muestrales y se calcula dividiendo la desviación estándar de la muestra entre la raíz cuadrada del tamaño de la muestra.

El intervalo de confianza resultante nos brinda una estimación de la media poblacional con un nivel de confianza determinado. Es decir, si construimos un intervalo de confianza del 95% para la media de las alturas de los estudiantes universitarios y obtenemos un rango de 165 cm a 170 cm, podemos afirmar con un 95% de confianza que la verdadera media de las alturas poblacionales se encuentra dentro de este rango.

Intervalos de confianza para la proporción

Los intervalos de confianza también se utilizan para estimar la proporción de una población a partir de una muestra. La construcción de un intervalo de confianza para la proporción se basa en la distribución muestral de la proporción, que sigue una distribución binomial si la muestra es aleatoria y suficientemente grande.

La fórmula para construir un intervalo de confianza para la proporción es la siguiente:

$$\text{Proporción de la muestra} \pm (\text{Valor crítico} * \text{Error estándar})$$

Donde:

Proporción de la muestra: es la proporción de éxitos (por ejemplo, la proporción de individuos que poseen una característica de interés) en la muestra.

Valor crítico: es obtenido de una tabla de valores críticos de la distribución normal estándar, que depende del nivel de confianza deseado.

Error estándar: es una medida de la variabilidad de los datos muestrales y se calcula utilizando la fórmula:

$$\text{Error Estándar} = \sqrt{\frac{p(1-p)}{n}}$$

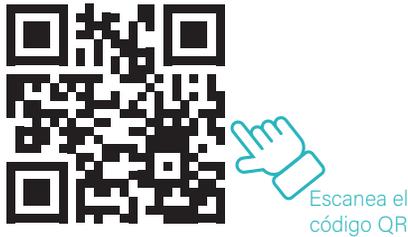
El intervalo de confianza resultante nos brinda una estimación de la proporción poblacional con un nivel de confianza determinado. Es decir, si construimos un intervalo de confianza del 90% para la proporción de personas que prefieren el color azul en una población y obtenemos un rango de 0.35 a 0.45, podemos afirmar con un 90% de confianza que la verdadera proporción poblacional de personas que prefieren el color azul se encuentra entre el 35% y 45% de la población.

Es importante tener en cuenta que, a mayor tamaño de muestra y mayor nivel de confianza deseado, el intervalo de confianza será más estrecho, lo que indica una estimación más precisa del parámetro poblacional. Sin embargo, construir intervalos de confianza requiere de ciertas suposiciones y condiciones, como que la muestra sea aleatoria y representativa de la población, que los datos sean independientes y que la distribución de los datos sea aproximadamente normal.

Para profundizar en el tema invitamos a ver el siguiente video para comprender mejor el concepto teórico.



Y el siguiente video ejemplo para el cálculo de los intervalos de confianza con el software SPSS.



Contrastando hipótesis sobre parámetros poblacionales

Cuando se busca comprobar alguna hipótesis sobre uno o más parámetros poblacionales desconocidos, como en el caso de determinar si hay diferencias en los niveles medios de ácido úrico entre hombres y mujeres, se aplican técnicas de contrastes de hipótesis. Esta técnica se basa en hacer una afirmación sobre el o los parámetros poblacionales y estudiar la coherencia entre esta afirmación y la información proporcionada por la muestra observada.

En la estadística inferencial, se compara la hipótesis planteada sobre la población con lo observado en la muestra para determinar si la hipótesis es verdadera o no. Para medir la discrepancia entre la hipótesis y los datos de la muestra, se utiliza una medida de discrepancia, como por ejemplo la diferencia entre el estimador puntual de la muestra y el valor hipotético del parámetro poblacional.

Por ejemplo, si se quiere determinar si el nivel medio de colesterol en un municipio es de 210 mg/100 ml y se tiene información de una muestra de 500 individuos de dicha población, con un nivel medio de colesterol observado de 226,58 mg/100 ml, se puede utilizar la diferencia entre el nivel medio observado y el valor hipotético como medida de discrepancia. Si la discrepancia es grande, entonces es menos probable que la hipótesis sea verdadera.

$$\text{Discrepancia} = \bar{x} - \mu = 226,58 - 210 = 16,58$$

Es importante destacar que existen diversas medidas de discrepancia que pueden ser utilizadas en la estadística inferencial, como los errores estándar, los intervalos de

confianza y las pruebas de hipótesis, entre otros. El uso de una medida de discrepancia específica dependerá del objetivo de la investigación y del tipo de datos obtenidos en la muestra.

Muestreo e inferencia estadística

Las técnicas estadísticas de intervalos de confianza y pruebas de hipótesis son utilizadas para hacer afirmaciones acerca de parámetros desconocidos de la población basándose en la información proporcionada por una muestra. Sin embargo, debido a que la muestra sólo representa una parte de la población, cualquier afirmación hecha acerca de los parámetros poblacionales estará sujeta a un error muestral, también conocido como error aleatorio, que es inherente al proceso de muestreo.

Es crucial contar con estimadores puntuales precisos de los parámetros poblacionales en el desarrollo de estas técnicas, y estos se construyen utilizando los datos de la muestra. La forma como se seleccionan los individuos o elementos de la muestra puede influir significativamente en los resultados obtenidos y, por lo tanto, en el valor del estimador puntual.

Teniendo en cuenta lo mencionado anteriormente, la muestra seleccionada debe ser representativa de la población objetivo, ya que las estimaciones se desviarán de los verdaderos valores de los parámetros poblacionales si la muestra no es representativa, lo que resultará en un error sistemático o sesgo. Para reducir o eliminar este error sistemático, se deben utilizar métodos de muestreo que contribuyan a conseguir muestras representativas de la población objetivo, y es necesario cuantificar el error muestral o aleatorio asociado a las estimaciones obtenidas a partir de ellas.

Muestreo probabilístico o aleatorio

El muestreo probabilístico es un proceso de selección de una muestra aleatoria de una población con el objetivo de hacer inferencias sobre la población en su totalidad. En la investigación de mercados, la selección de una muestra aleatoria es crucial para garantizar que las conclusiones obtenidas sean representativas de la población objetivo.

Existen varios tipos de muestreo probabilístico, cada uno con sus propias fortalezas y debilidades. A continuación, analizaremos los cuatro tipos principales de muestreo probabilístico: muestreo aleatorio simple, muestreo estratificado, muestreo por conglomerados y muestreo sistemático.

Muestreo aleatorio simple:

El muestreo aleatorio simple (MAS) es el método más común de muestreo probabilístico utilizado en la investigación de mercados. Este método de selección de muestras se

utiliza para garantizar que cada elemento de la población tenga una probabilidad igual de ser seleccionado en la muestra. En otras palabras, cada elemento de la población tiene la misma oportunidad de ser seleccionado.

El muestreo aleatorio simple se utiliza a menudo cuando se dispone de una lista completa de elementos de la población, y se desea seleccionar una muestra representativa de manera eficiente. Este método puede ser muy efectivo en la selección de muestras para estudios de opinión pública y en la investigación de mercados en general. Sin embargo, el muestreo aleatorio simple puede no ser adecuado para poblaciones heterogéneas, ya que puede ser difícil garantizar que cada subgrupo esté representado adecuadamente.

Muestreo estratificado:

El muestreo estratificado es un método de selección de muestras en el que la población se divide en subgrupos o estratos y se selecciona una muestra de cada estrato. Este método se utiliza para garantizar que cada subgrupo de la población esté representado en la muestra.

El muestreo estratificado es especialmente útil cuando la población es heterogénea y se desea asegurar que cada subgrupo esté representado adecuadamente. Este método también puede ser más eficiente que el muestreo aleatorio simple en términos de composición de la muestra, ya que los estratos con mayor variabilidad pueden requerir un tamaño de muestra mayor, mientras que los estratos con menor variabilidad pueden requerir un tamaño de muestra menor.

Muestreo por conglomerados:

El muestreo por conglomerados es un método de selección de muestras en el que la población se divide en grupos o conglomerados, y se selecciona una muestra de cada conglomerado. Este método se utiliza cuando es impracticable seleccionar una muestra aleatoria simple de la población, y en su lugar, es más fácil seleccionar una muestra aleatoria simple de cada conglomerado.

El muestreo por conglomerados puede ser más eficiente que el muestreo aleatorio simple cuando la población es muy grande o está dispersa geográficamente. Sin embargo, este método puede ser menos preciso que otros métodos de muestreo, ya que la variabilidad dentro de cada conglomerado puede ser mayor que la variabilidad entre conglomerados.

Muestreo sistemático:

El muestreo sistemático es un método de selección de muestras en el que la población se ordena, y se selecciona una muestra con cada cierto número de elementos. Este método implica seleccionar un elemento aleatorio de la población y luego tomar cada k -ésimo elemento a partir de ese punto. La constante k se determina dividiendo el tamaño de la población (N) por el tamaño de la muestra deseada (n). Por ejemplo, si queremos seleccionar una muestra de 100 elementos de una población de 1000, k sería 10. Luego seleccionamos un elemento aleatorio de los primeros 10 elementos de la población y tomamos cada décimo elemento a partir de ese punto.

El muestreo sistemático puede ser menos preciso que otros métodos si hay algún patrón en la forma como se organizan los elementos de la población, como una periodicidad o un ordenamiento sistemático. En estos casos, el muestreo sistemático puede llevar a una subrepresentación o sobrerrepresentación de ciertos subgrupos de la población.

A continuación, se muestra una tabla que resume los diferentes tipos de muestreo junto con sus ventajas y desventajas

Tabla 3.1

Resumen tipos de muestreo probabilístico

Tipo de muestreo	Descripción	Ventajas	Desventajas
Muestreo aleatorio simple	Cada elemento de la población tiene una probabilidad conocida de ser seleccionado	Permite estimar parámetros de la población con precisión	Requiere una lista completa y actualizada de la población
Muestreo estratificado	La población se divide en estratos y se selecciona una muestra aleatoria simple de cada estrato	Asegura una representación adecuada de subgrupos en la muestra	Requiere información precisa sobre los estratos y su tamaño
Muestreo por conglomerados	La población se divide en conglomerados y se selecciona un subconjunto de conglomerados de forma aleatoria	Eficiente en términos de costos y tiempo	Puede introducir sesgos si los conglomerados no son representativos
Muestreo sistemático	Se selecciona un elemento de la población de forma aleatoria y luego se seleccionan los elementos siguientes a intervalos fijos	Fácil de implementar y menos costoso que el muestreo aleatorio simple	Puede introducir sesgos si hay patrones en la población

Nota. Elaboración propia.

Muestreo no probabilístico

El muestreo no probabilístico es un enfoque de selección de muestras en el que los elementos de la población no son seleccionados con una probabilidad conocida. A diferencia del muestreo probabilístico, donde cada elemento tiene una probabilidad clara de ser seleccionado, el muestreo no probabilístico se basa en la elección subjetiva del investigador o en la disponibilidad de los elementos en la población.

A continuación, explicaremos en detalle el concepto de muestreo no probabilístico y los diferentes tipos de enfoques que se utilizan en la práctica.

El muestreo no probabilístico se utiliza en situaciones en las que no es posible o práctico utilizar un enfoque probabilístico debido a limitaciones de recursos, tiempo o acceso a la población. Aunque el muestreo no probabilístico puede ser menos preciso y generalizable que el muestreo probabilístico, sigue siendo valioso en muchas situaciones y puede proporcionar resultados útiles para la investigación y la toma de decisiones.

Uno de los tipos más comunes de muestreo no probabilístico es el muestreo por conveniencia. En este enfoque, los elementos de la muestra se seleccionan en función de su disponibilidad y accesibilidad para el investigador. Por ejemplo, si un investigador está interesado en estudiar la satisfacción del cliente en una tienda en particular, puede seleccionar a los clientes que visitan la tienda durante un período de tiempo específico como su muestra. Sin embargo, este enfoque puede introducir sesgos, ya que solo se incluyen en la muestra a aquellos que están fácilmente disponibles o dispuestos a participar.

Otro tipo de muestreo no probabilístico es el muestreo por juicio. En este caso, el investigador utiliza su propio criterio para seleccionar los elementos de la muestra. Por ejemplo, un investigador que estudia la adicción a las drogas podría seleccionar a individuos para su muestra en función de su conocimiento y experiencia en el tema. Sin embargo, este enfoque también puede ser sesgado, ya que la selección está sujeta a la subjetividad y la opinión del investigador.

El muestreo intencional es otro tipo de muestreo no probabilístico en el que los elementos de la muestra se seleccionan con un propósito específico en mente. Por ejemplo, un investigador interesado en estudiar los efectos de un programa de capacitación laboral puede seleccionar como muestra a los trabajadores de una empresa particular debido a la accesibilidad de esta. Aunque este enfoque puede ser útil para estudiar el grupo específico de interés, puede no ser representativo para la población en general.

El muestreo por cuotas es otro enfoque de muestreo no probabilístico que se basa en la selección de elementos en función de ciertas características predefinidas para asegurar que la muestra tenga una composición similar a la de la población. Por ejemplo, si se desea estudiar la opinión de los votantes en una elección presidencial, se puede seleccionar a un cierto número de votantes de diferentes edades, géneros y ubicaciones geográficas para asegurar que la muestra represente adecuadamente la diversidad de la población votante. Sin embargo, este enfoque también puede tener sesgos, ya que la selección de los elementos se basa en criterios subjetivos y no en una probabilidad de selección conocida.

Por último, el muestreo en bola de nieve es otro enfoque de muestreo no probabilístico que se basa en la selección de elementos a través de referencias o contactos

previos. En este caso, los primeros elementos de la muestra son seleccionados por el investigador y luego se les pide que refieran a otros posibles participantes que cumplan con los criterios de interés del estudio. Este enfoque es comúnmente utilizado en investigaciones cualitativas o en situaciones donde la población de interés es difícil de alcanzar o identificar de antemano. Sin embargo, este enfoque también puede tener sesgos, ya que la muestra puede estar influenciada por la red social o contactos del primer participante seleccionado.

A continuación, se muestra una tabla que resume los diferentes tipos de muestreo no probabilísticos junto con sus ventajas y desventajas

Tabla 3.2

Resumen tipos de muestreo no probabilístico

Tipo de muestreo	Descripción	Ventajas	Desventajas
Muestreo por cuotas	Selección de elementos en función de características predefinidas	Asegura una composición similar a la de la población	Puede tener sesgos
Muestreo por conveniencia	Selección de elementos basada en disponibilidad y conveniencia	Fácil y rápido de llevar a cabo	No sigue una probabilidad conocida de selección, puede tener sesgos
Muestreo intencional	Selección de elementos de forma deliberada por el investigador	Útil para estudiar subgrupos específicos	Puede tener sesgos
Muestreo bola de nieve	Selección de elementos a través de referencias o contactos previos	Común en investigaciones cualitativas o poblaciones difíciles de alcanzar	Puede tener sesgos

Nota. Elaboración propia.

Determinación de tamaños muestrales

Determinar el tamaño muestral adecuado es un paso crucial en el diseño de un estudio de investigación, ya que afecta la precisión y confiabilidad de los resultados obtenidos. El tamaño muestral se refiere al número de elementos que se seleccionarán de la población para formar parte de la muestra. Existen diferentes métodos para determinar el tamaño

muestral, y estos varían según el tipo de estimación que se desee realizar, ya sea para la media o para proporciones, así como si la población es finita o infinita.

Para población finita con estimación de la media, uno de los métodos más comunes es la fórmula de tamaño muestral para población finita. Esta fórmula se basa en el uso de un factor de corrección que tiene en cuenta el tamaño de la población (N), la desviación estándar poblacional (σ), y el nivel de confianza deseado (Z) y el margen de error aceptable (e). La fórmula para determinar el tamaño muestral (n) se expresa de la siguiente manera:

$$n = \frac{\left(z_{\frac{\alpha}{2}}^2 * \theta^2 * N \right)}{\left(z_{\frac{\alpha}{2}}^2 * \theta^2 \right) + (e^2 * (N - 1))}$$

Por otro lado, si se desea estimar una proporción en una población finita, se puede utilizar la fórmula de tamaño muestral para población finita para proporciones. Esta fórmula también incorpora el tamaño de la población (N), el nivel de confianza deseado (Z), el margen de error aceptable (E), y la proporción poblacional estimada (p). La fórmula se expresa de la siguiente manera:

$$n = \frac{\left(N * z_{\frac{\alpha}{2}}^2 * p * (1 - p) \right)}{(e^2 * (N - 1)) + (z_{\frac{\alpha}{2}}^2 * p * (1 - p))}$$

Es importante tener en cuenta que, en ambos casos, si el tamaño de la población es grande ($N \geq 10,000$ aproximadamente) o simplemente se desconoce, se puede utilizar la fórmula de tamaño muestral para población infinita, que es una aproximación válida. La fórmula de tamaño muestral para población infinita para estimación de la media se expresa de la siguiente manera:

$$n = \frac{z_{\frac{\alpha}{2}}^2 * S^2}{e}$$

Y la fórmula de tamaño muestral para población infinita para estimación de proporciones se expresa de la siguiente manera:

$$n = \frac{z_{\frac{\alpha}{2}}^2 * p * (1 - p)}{e^2}$$

Ejemplo 3.2

Supongamos que deseamos calcular el tamaño de muestra requerido para estimar la edad promedio de las personas que realizan deporte regularmente en una población específica, con un nivel de confianza del 0,95 o del 95%, y una precisión de 0,5 años. Es importante tener en cuenta que para determinar el tamaño de muestra es necesario conocer el valor de S^2 , lo cual no es posible sin haber observado previamente una muestra. Por lo tanto, se debe tomar en una primera etapa una muestra relativamente pequeña de la población (n_1) para obtener una primera estimación o buscar información en la literatura sobre S^2 . De esta manera, se puede calcular el tamaño adecuado de la muestra para garantizar una estimación precisa de la edad media de la población deportista con el nivel de confianza deseado.

En el ejemplo que se está considerando, se supone que la desviación estándar es de 11. Al sustituir este valor en la expresión correspondiente, se podrá obtener el tamaño de muestra necesario.

$$n = \frac{1,96^2 * 11^2}{0,5^2} = 169.03 \sim 170$$

Por lo tanto, el tamaño muestral necesario para estimar la edad media de los individuos que practican deporte de forma regular en la población con una confianza del 95% y una precisión de 0,5 años, y conocida la desviación estándar de 11, es aproximadamente 170 sujetos.

Ejemplo 3.3

Supongamos que deseas estimar la proporción de estudiantes universitarios que tienen un trabajo a tiempo parcial en una universidad concreta, y sabes que la población total de estudiantes en esa universidad es de 10.000. Quieres calcular el tamaño de muestra necesario para obtener un nivel de confianza del 95% y un margen de error del 3%. En este caso, asumimos una estimación de proporción del 50% para obtener el tamaño de muestra máximo, ya que es la estimación más conservadora. Como la población es conocida utilizaremos la siguiente fórmula:

$$n = \frac{\left(N * z_{\frac{\alpha}{2}}^2 * p * (1-p) \right)}{\left(e^2 * (N-1) \right) + \left(z_{\frac{\alpha}{2}}^2 * p * (1-p) \right)} = \frac{(10000 * 1,96^2 * 0.5 * 1 - 0.5)}{(0,03^2 * (10000 - 1)) + (1,96^2 * 0.5 * (1 - 0.5))} = 107,6 \sim 108$$

Por lo tanto, necesitarías una muestra de aproximadamente 107 estudiantes universitarios para estimar la proporción de estudiantes que tienen un trabajo a tiempo parcial en la universidad.

Es importante recordar que el tamaño muestral adecuado puede variar según la naturaleza del estudio, los objetivos de investigación, el diseño del estudio y las características

de la población en estudio. Un tamaño muestral insuficiente puede resultar en estimaciones poco precisas y poco confiables, mientras que un tamaño muestral excesivo puede ser costoso y redundante en términos de recursos y tiempo.

Es fundamental tener en cuenta estas consideraciones al determinar el tamaño muestral en un estudio de investigación. Un tamaño muestral adecuado contribuirá a obtener resultados válidos y confiables que puedan respaldar las conclusiones y recomendaciones del estudio.

Distribuciones muestrales

Una parte fundamental de la estadística es el estudio de las distribuciones de probabilidad, que son modelos matemáticos que describen cómo se distribuyen los valores posibles de una variable aleatoria. En este apartado, realizaremos una revisión detallada de los diferentes tipos de distribuciones en estadística, incluyendo distribuciones discretas y continuas, distribuciones simétricas y asimétricas, así como distribuciones univariadas y multivariadas.

Distribuciones discretas:

Las distribuciones discretas son aquellas en las que la variable aleatoria solo puede tomar valores enteros. Algunos ejemplos de distribuciones discretas son:

Distribución uniforme discreta: En esta distribución, todos los valores posibles tienen la misma probabilidad de ocurrir. Es decir, cada valor tiene la misma probabilidad de ser seleccionado.

Distribución de Bernoulli: Es una distribución binaria que representa un experimento con dos resultados posibles, generalmente etiquetados como éxito o fracaso. Se utiliza para modelar situaciones en las que solo hay dos posibles resultados.

Distribución binomial: Es una distribución discreta que representa el número de éxitos en una serie de n ensayos independientes y de igual probabilidad. Es ampliamente utilizada en estadística para modelar eventos con resultados binarios y con un número fijo de ensayos.

Distribución de Poisson: Es una distribución discreta que se utiliza para modelar la ocurrencia de eventos raros o infrecuentes, como el número de llamadas telefónicas que llegan a un centro de atención al cliente en un período de tiempo determinado.

Distribuciones continuas:

Las distribuciones continuas son aquellas en las que la variable aleatoria puede tomar cualquier valor en un intervalo continuo. Algunos ejemplos de distribuciones continuas son:

Distribución uniforme continua: En esta distribución, todos los valores posibles en un intervalo tienen la misma probabilidad de ocurrir. Es decir, la probabilidad de que la variable

aleatoria caiga en cualquier subintervalo del intervalo dado es proporcional a la longitud de ese subintervalo.

Distribución normal o gaussiana: Es una de las distribuciones más utilizadas en estadística y se caracteriza por tener una forma de campana simétrica alrededor de su media. Es ampliamente utilizada en la modelización de datos en diversas áreas debido a su propiedad de ser una distribución de probabilidad que se encuentra con frecuencia en la naturaleza.

Distribución exponencial: Es una distribución continua que se utiliza para modelar eventos que ocurren de manera aleatoria, pero con una tasa constante, como el tiempo entre llegadas de clientes a un servicio.

Distribución gamma: Es una distribución continua que generaliza la distribución exponencial y se utiliza para modelar eventos que tienen una tasa de ocurrencia variable.

Distribuciones simétricas y asimétricas:

Las distribuciones pueden ser simétricas o asimétricas en función de su forma y la posición de su centro de gravedad. Algunos ejemplos de distribuciones simétricas son la distribución normal y la distribución uniforme. En estas distribuciones, la forma de la curva es simétrica respecto a su media, lo que significa que la probabilidad de que la variable aleatoria tome valores a la derecha o a la izquierda de la media es igual.

Por otro lado, las distribuciones asimétricas son aquellas en las que la forma de la curva no es simétrica y hay una inclinación hacia un lado. Las distribuciones asimétricas pueden ser positivamente asimétricas, lo que significa que la cola de la distribución se extiende hacia la derecha, o negativamente asimétricas, lo que significa que la cola de la distribución se extiende hacia la izquierda. Algunos ejemplos de distribuciones asimétricas son la distribución exponencial y la distribución de Poisson.

Distribuciones univariadas y multivariadas:

Las distribuciones pueden ser univariadas o multivariadas en función del número de variables aleatorias involucradas. Una distribución univariada se refiere a una sola variable aleatoria, mientras que una distribución multivariada se refiere a dos o más variables aleatorias.

Las distribuciones univariadas son las más comunes y ampliamente utilizadas en estadística. Algunos ejemplos de distribuciones univariadas son la distribución normal, la distribución exponencial y la distribución de Poisson, que hemos mencionado anteriormente.

Por otro lado, las distribuciones multivariadas son más complejas y se utilizan para modelar la relación entre dos o más variables aleatorias. Algunos ejemplos de distribuciones multivariadas son la distribución normal multivariada, que se utiliza para modelar múltiples variables aleatorias con una estructura de covarianza, y la distribución de Wishart, que se utiliza para modelar la matriz de covarianza de varias variables aleatorias.

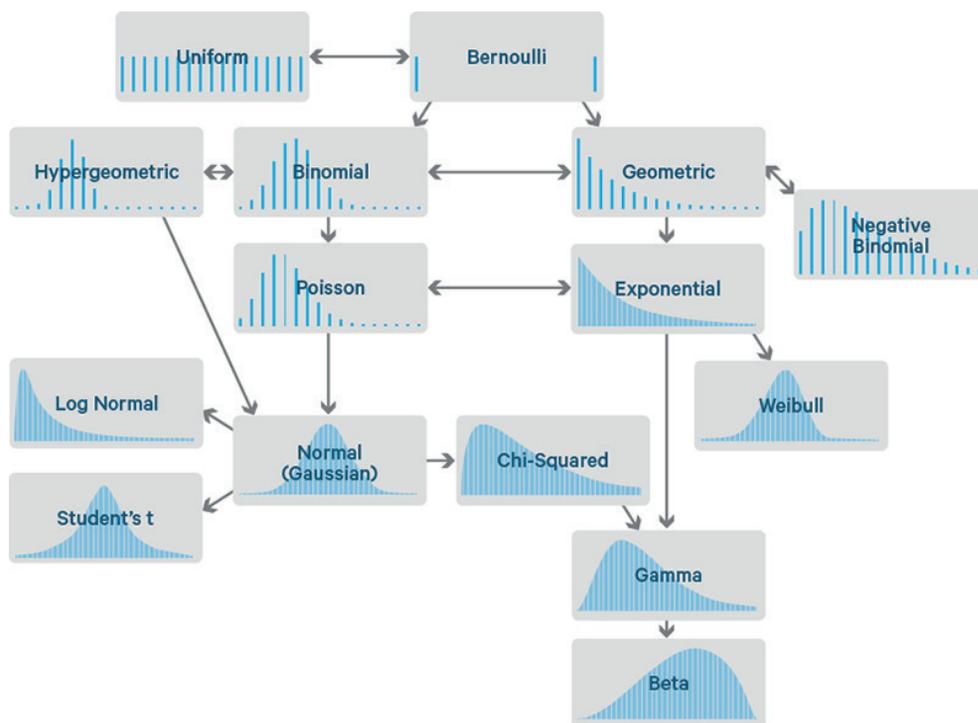


Figura 3.1

Diferentes tipos de distribuciones muestrales

Nota. Tomado de Briega (2016).

Determinar la normalidad de una variable

La normalidad es un concepto importante en estadística, como habíamos visto anteriormente, esta se refiere a la distribución de una variable en una población. Una distribución normal, también conocida como distribución de Gauss o distribución en forma de campana, es una distribución simétrica y unimodal en la cual la mayoría de los datos se concentra cerca de la media y se distribuye simétricamente en ambos lados de la media. Determinar si una variable sigue una distribución normal es esencial en muchos análisis estadísticos, ya que muchos métodos estadísticos se basan en la suposición de normalidad de los datos.

Existen varios métodos para determinar la normalidad de las variables. Algunos de los métodos comunes son los siguientes:

Gráfico de Q-Q (Quantile-Quantile): El gráfico Q-Q es una herramienta gráfica que compara los cuantiles de los datos con los cuantiles esperados de una distribución normal. Si

los puntos en el gráfico siguen una línea diagonal, es indicativo de que los datos se ajustan a una distribución normal.

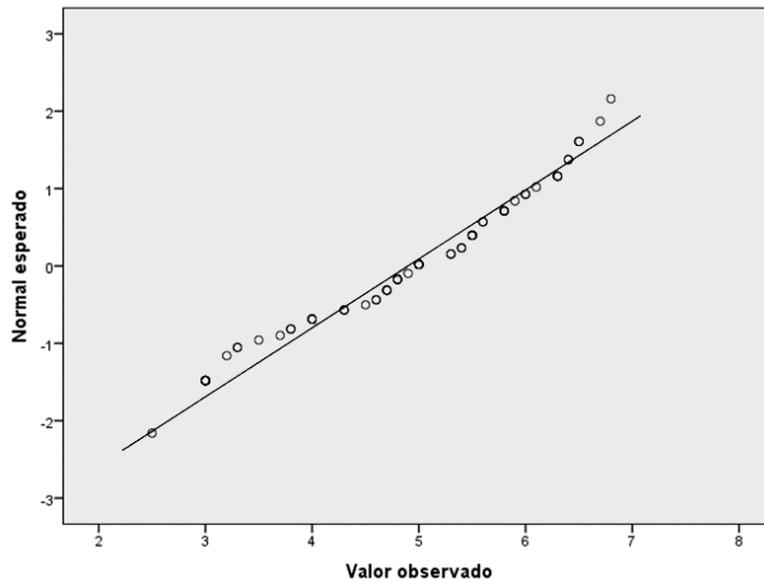


Figura 3.2

Gráfico Q-Q

Nota. Elaboración propia. Construido con SPSS.

Prueba de normalidad de Anderson-Darling: Esta prueba es una mejora de la prueba de bondad de ajuste de Kolmogorov-Smirnov que da más peso a las colas de la distribución. La prueba de Anderson-Darling compara los cuantiles empíricos de los datos con los cuantiles esperados de una distribución normal y calcula una estadística de prueba. Si el valor p obtenido de esta prueba es mayor que un nivel de significancia predefinido, entonces se acepta la hipótesis nula de que los datos siguen una distribución normal.

Prueba de normalidad de Shapiro-Wilk: Es una prueba de normalidad ampliamente utilizada que evalúa si los datos se ajustan a una distribución normal. La prueba de Shapiro-Wilk calcula una estadística de prueba basada en los coeficientes de regresión de los datos en función de los cuantiles teóricos esperados de una distribución normal. Un valor p obtenido de esta prueba mayor que un nivel de significancia dado implica que los datos se ajustan a una distribución normal.

Prueba de normalidad de Lilliefors: Es una variante modificada de la prueba de Kolmogorov-Smirnov que se utiliza específicamente para muestras pequeñas. Es una prueba no paramétrica que compara la distribución empírica de los datos con una distribución

normal teórica. Un valor p obtenido de esta prueba mayor que un nivel de significancia predefinido implica que los datos se ajustan a una distribución normal.

Gráficos de densidad e histogramas: Los gráficos de densidad y los histogramas son herramientas visuales que pueden ayudar a evaluar la normalidad de los datos. En una distribución normal, la forma de la densidad o del histograma debe tener una forma de campana simétrica alrededor de la media.

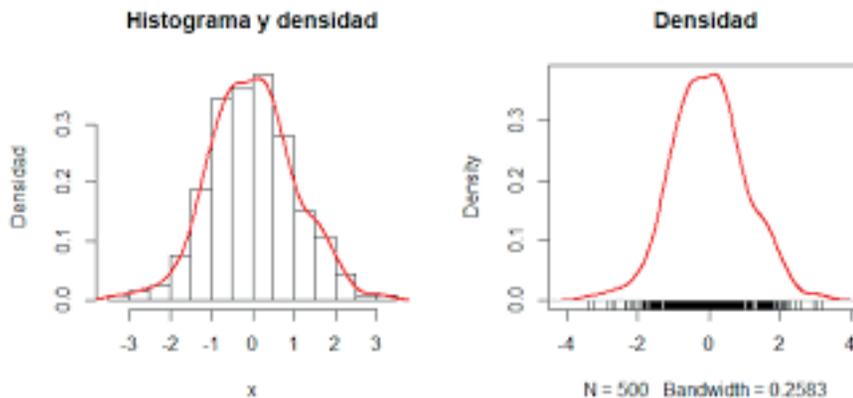


Figura 3.3

Gráficos de densidad e histograma

Nota. Elaboración propia. Gráficos realizados con R studio.

Es importante tener en cuenta que ningún método es perfecto y que los resultados de las pruebas de normalidad deben interpretarse con precaución. En algunos casos, las pruebas de normalidad pueden no ser sensibles para detectar desviaciones leves de la normalidad o pueden arrojar resultados incorrectos debido al tamaño de la muestra o a la presencia de datos atípicos. Por lo tanto, es recomendable utilizar múltiples métodos y herramientas visuales para evaluar la normalidad de los datos.

Además, en algunos análisis estadísticos, la normalidad de los datos puede no ser estrictamente necesaria. Por ejemplo, muchos métodos basados en la inferencia robusta, como los intervalos de confianza robustos o las pruebas no paramétricas, no requieren la suposición de normalidad de los datos. Sin embargo, en otros casos, como en algunas pruebas de hipótesis paramétricas o en la regresión lineal, la normalidad de los datos puede ser un supuesto importante.

En este tema en particular, se dispone de una serie de videos que muestran las generalidades y métodos a consideración sobre la estadística inferencial.



Escanea el
código QR

CAPÍTULO 4. CONTRASTE DE HIPÓTESIS



De forma similar a la estimación de parámetros poblacionales a través de intervalos de confianza, los contrastes de hipótesis son técnicas estadísticas utilizadas para inferir información sobre los parámetros de una población a partir de una muestra. El proceso de inferencia implica formular una afirmación sobre uno o varios parámetros poblacionales (hipótesis) y luego evaluar si los datos observados en la muestra son compatibles con esa afirmación. Si los datos observados en la muestra no son consistentes con la hipótesis planteada, se puede inferir que hay evidencia en contra de la hipótesis y se pueden realizar conclusiones al respecto.

Ejemplo 4.1

En un estudio se recolectaron datos sobre el nivel de lactato en sangre de un grupo de 46 deportistas seleccionados aleatoriamente. El promedio de nivel de lactato fue de 6,28 mmol/L con una desviación típica de 2,1 mmol/L. ¿Proporcionan estos datos evidencia suficiente que indique que el nivel promedio de lactato en este tipo de deportistas es superior a 4 mmol/L?

Ejemplo 4.2

De los 250 individuos a los que se les administró un determinado tratamiento, 180 respondieron de forma positiva. ¿Puede considerarse que la proporción de éxito del tratamiento es del 80%?

En el primer ejemplo, se considera que la población se compone de todos los individuos deportistas, y el parámetro sobre el que se hace la afirmación es la media del nivel de lactato en sangre en mmol/L. La hipótesis establecida es que la media es menor o igual a 4 mmol/L.

En contraste con el ejemplo anterior, en este segundo caso la población de estudio se compone de individuos afectados o que podrían estar afectados por la enfermedad que el tratamiento pretende tratar. El parámetro poblacional que se quiere evaluar es la proporción de éxito del tratamiento, y se establece una hipótesis según la cual esta proporción es del 80% o 0,8. En ambos casos, el resultado del contraste de hipótesis permitirá validar o rechazar la hipótesis planteada. En este sentido los pasos necesarios para realizar el contraste de hipótesis se describen a continuación.

1. Definición de las hipótesis de contraste

Se busca contrastar una hipótesis sobre un parámetro que representa la media poblacional. La pregunta es si la media poblacional puede considerarse como 4 mmol o si, en cambio, difiere de 4 mmol. Estas dos opciones se plasman en las siguientes hipótesis a contrastar:

$$H_0 : \mu = 4 \text{ m / mol}$$

$$H_1 : \bar{\mu} \neq 4 \bar{m} / \text{mol}$$

Donde H_0 se denomina *Hipótesis nula* y H_1 *Hipótesis alterna*. En este sentido la idea del contraste de hipótesis es

2. Determinar el estadístico de contraste para confirmar o rechazar la hipótesis nula teniendo en cuenta la información proporcionada en los datos de la muestra observada

En el caso del contraste de una media, cuando se desconoce la desviación típica poblacional, el estadístico de contraste utilizado es:

$$z = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{6,28 - 4}{2,1 / \sqrt{46}} = 7,36$$

Donde μ es el valor de la media que se estableció en la hipótesis nula, \bar{x} es la media muestral, S es el estimador puntual de la desviación estándar y n el tamaño de la muestra. Es importante notar que debido a que la desviación típica es siempre un valor positivo, por lo tanto, el estadístico de contraste solo puede tener valores mayores o iguales a 0; si la media observada en la muestra coincide exactamente con la media propuesta en la hipótesis nula. A medida que la diferencia entre la media observada y la media de la esperada de la hipótesis nula aumenta, el valor del estadístico de contraste también aumenta en términos absolutos. Por lo tanto, los valores del estadístico de contraste cercanos a 0 indican un mayor respaldo a la hipótesis nula, mientras que valores alejados de 0 indican evidencia en contra de dicha hipótesis. Es importante determinar cuándo la discrepancia es lo suficientemente grande como para rechazar la hipótesis nula, y en este sentido, el siguiente paso en consideración puede resultar útil.

3. Obtener información acerca de la distribución de probabilidad relacionada con la medida de discrepancia o estadístico de contraste utilizado

Es fundamental tener un conocimiento claro de la distribución de probabilidad que rige el comportamiento del estadístico de contraste para llevar a cabo de manera adecuada el contraste de hipótesis. En el ejemplo dado, se cuenta con 46 datos con una distribución normal y mediante el teorema de límite central puede asumirse que esta distribución es simétrica. En consecuencia, si la hipótesis nula resultara verdadera, el estadístico de contraste se situaría en la región central de la distribución, lo que implicaría que es altamente improbable observar valores en cualquiera de los extremos de dicha distribución.

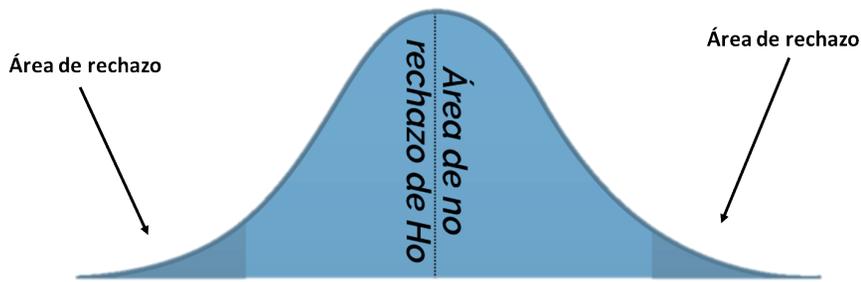


Figura 4.1

Pruebas de hipótesis

Nota. Elaboración propia.

4. Establecer nivel de significación del contraste

Para establecer qué valores del estadístico de contraste tienen una probabilidad muy baja de ocurrir si la hipótesis nula es verdadera, se define el nivel de significancia del contraste, que también se conoce como nivel de significancia a priori. Este nivel suele fijarse en 0,05, aunque en algunos casos se puede utilizar 0,01 o incluso 0,001, dependiendo del contexto. En el ejemplo dado, se ha decidido utilizar un nivel de significación a priori de 0,05. Esto implica que se considerarán como evidencia en contra de la hipótesis nula aquellos valores del estadístico de contraste que sean iguales o más extremos que lo esperado bajo el nivel de significación a priori de 0,05, lo que nos ayudará a tomar una decisión precisa en el contraste de hipótesis.

5. Construcción regla de decisión

Si se establece un nivel de significancia de 0,05, se deben descartar los valores del estadístico de contraste que tengan una probabilidad menor a 0,05. Los valores menos probables del estadístico de contraste, en caso de que la hipótesis nula sea verdadera, se encuentran en los extremos de la distribución, por lo que estas áreas se definen como la región crítica o región de rechazo de la hipótesis nula (ver Figura 4.1).

De forma complementaria, se define como región de aceptación de la hipótesis nula a la región comprendida entre $t_{0,025}$ y $t_{0,975}$. Por lo tanto, los valores $t_{0,025}$ y $t_{0,975}$ sobre una distribución t_{45} que determinan la región de aceptación y de rechazo de la hipótesis nula en este caso son:

$t_{0,025} = -2,014$ y $t_{0,975} = 2,014$. La forma en que se tomará la decisión se definirá como sigue:

- Si $EC > 2,014$ o $EC < -2,014 \rightarrow$ Se rechaza la hipótesis nula H_0 .
- Si $-2,014 \leq EC \leq 2,014 \rightarrow$ Se acepta la hipótesis nula H_0 .

6. Aplicación de la regla de decisión

Por último, es necesario utilizar la regla de decisión del contraste de hipótesis, lo que implica definir la región en la que se ubica el valor del estadístico de contraste calculado en el paso 2.

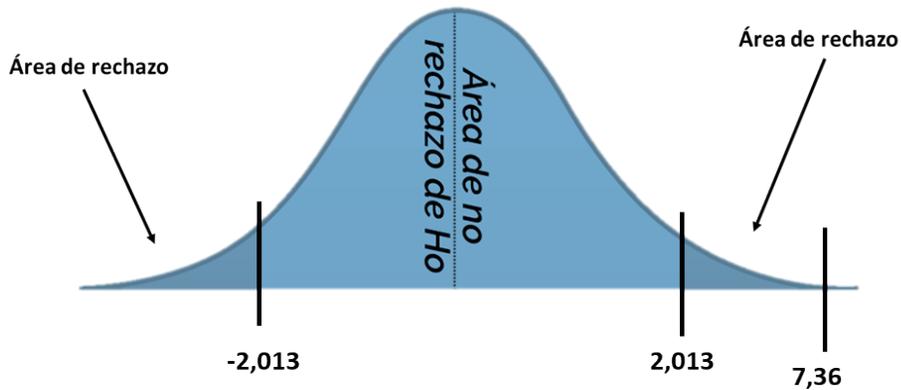


Figura 4.2

Aplicación regla de decisión

Nota. Elaboración propia.

De acuerdo con lo que se muestra en la Figura 4.2, el valor del estadístico de contraste EC es 7,36 y se ubica dentro de la región crítica de rechazo de la hipótesis nula ($7,36 > 2,013$). Por lo tanto, siguiendo la regla de decisión, se concluye que la hipótesis nula debe ser rechazada, lo que sugiere que hay suficiente evidencia para afirmar que el valor promedio del lactato es significativamente diferente de 4 mmol/L.

Errores en el contraste de hipótesis

Como ya se ha mencionado en capítulos anteriores, la estadística es una disciplina fundamental en la investigación científica, la cual busca analizar y comprender los datos para tomar decisiones basadas en evidencia. Sin embargo, como en cualquier disciplina, la estadística no está exenta de errores. En este apartado, se analizarán dos tipos de errores comunes en estadística.

Error tipo 1:

El Error tipo 1, también conocido como error de falsos positivos, ocurre cuando se rechaza una hipótesis nula siendo esta verdadera. En otras palabras, se llega a la conclusión de que hay una diferencia o efecto significativo, cuando en realidad no la hay. Este error se

produce cuando se establece un nivel de significancia demasiado bajo, lo que lleva a una mayor probabilidad de cometer este tipo de error.

Ejemplo 4.3

Supongamos que un grupo de investigadores está llevando a cabo un estudio para investigar si un nuevo tratamiento para la hipertensión arterial es más efectivo que el tratamiento estándar. La hipótesis nula (H_0) es que no hay diferencia significativa entre los dos tratamientos, mientras que la hipótesis alternativa (H_1) es que el nuevo tratamiento es más efectivo.

Los investigadores deciden utilizar un nivel de significancia (α) de 0.05, lo que significa que están dispuestos a aceptar una probabilidad de cometer un error tipo 1 del 5%. Realizan el estudio y obtienen los resultados. Al analizar los datos, calculan un valor de p (probabilidad) de 0.03, que es menor que el nivel de significancia establecido.

Con base en estos resultados, los investigadores deciden rechazar la hipótesis nula y concluir que el nuevo tratamiento es efectivo. Sin embargo, más tarde se descubre que debido a un error en la recopilación de datos, se incluyeron datos incorrectos en el análisis, lo que llevó a un resultado falso.

En este caso, el error tipo 1 se cometió al rechazar la hipótesis nula y concluir que el nuevo tratamiento es efectivo, cuando en realidad no había evidencia suficiente para hacer tal afirmación. Este error podría tener consecuencias significativas en la toma de decisiones clínicas, como la adopción de un nuevo tratamiento en la práctica médica, basada en resultados incorrectos.

Error tipo 2:

El Error tipo 2, también conocido como error de falsos negativos, ocurre cuando se acepta una hipótesis nula siendo falsa. En otras palabras, se llega a la conclusión de que no hay una diferencia o efecto significativo, cuando en realidad sí lo hay. Este error se produce cuando se establece un nivel de significancia demasiado alto, lo que lleva a una menor probabilidad de detectar efectos reales.

Ejemplo 4.4

Imaginemos que un grupo de investigadores está realizando un estudio para investigar si un nuevo medicamento para reducir el colesterol es más efectivo que el medicamento actual en el mercado. La hipótesis nula (H_0) es que no hay diferencia significativa entre los dos medicamentos, mientras que la hipótesis alternativa (H_1) es que el nuevo medicamento es más efectivo.

Los investigadores deciden reclutar a 200 participantes para el estudio, dividiéndolos en dos grupos: uno que recibirá el medicamento actual y otro que recibirá el nuevo medicamento. Realizan el estudio y recopilan los datos. Al analizar los resultados, obtienen un valor de p (probabilidad) de 0.08, que es mayor que el nivel de significancia establecido de 0.05.

Con base en estos resultados, los investigadores no logran rechazar la hipótesis nula y concluyen que no hay evidencia suficiente para afirmar que el nuevo medicamento es más efectivo que el actual. Sin embargo, más tarde se deduce que debido a errores en la medición de los niveles de colesterol, algunos de los participantes del grupo del nuevo medicamento recibieron dosis más bajas de lo previsto, lo que afectó la capacidad del estudio para detectar diferencias significativas.

Es importante tener en cuenta que la elección del nivel de significancia, que determina la probabilidad de cometer errores tipo 1, y el tamaño de muestra, que influye en la probabilidad de cometer errores tipo 2, son decisiones críticas en cualquier análisis estadístico. Una selección inapropiada de estos parámetros puede aumentar la probabilidad de cometer uno de estos errores y llevar a conclusiones incorrectas.

A continuación, se presenta una tabla que incluye la descripción, consecuencias y estrategias para mitigar los errores tipo 1 y tipo 2 en estadística:

Tabla 4.1

Tipos de errores en la estadística

Error	Descripción	Consecuencias	Cómo se puede mitigar
Error tipo 1	Rechazar incorrectamente la hipótesis nula (H_0) cuando es verdadera	Conclusión incorrecta de que hay una diferencia o efecto significativo cuando no lo hay	<ul style="list-style-type: none"> • Elegir un nivel de significancia apropiado (α) antes de realizar el estudio • Validar cuidadosamente los resultados antes de sacar conclusiones finales • Replicar el estudio para confirmar los resultados
Error tipo 2	Aceptar incorrectamente la hipótesis nula (H_0) cuando es falsa	Conclusión incorrecta de que no hay diferencia o efecto significativo cuando sí lo hay	<ul style="list-style-type: none"> • Asegurarse de que el tamaño de muestra sea adecuado para detectar diferencias significativas • Utilizar pruebas estadísticas apropiadas con poder estadístico adecuado • Validar cuidadosamente los procedimientos de medición y análisis • Replicar el estudio con un tamaño de muestra mayor si es posible

Nota. Elaboración propia.

A continuación, se proporciona un enlace a un video en YouTube que brinda un contexto más completo sobre el concepto que se discutió anteriormente.



Contrastes bilaterales y unilaterales

Los contrastes de hipótesis en estadística pueden ser unilaterales o de una cola, y bilaterales o de dos colas, dependiendo de cómo se planteen las hipótesis. En el caso de que el interés se centre en contrastar la hipótesis de que un determinado parámetro de la población tome exactamente un valor dado frente a la hipótesis de que el parámetro tome un valor distinto al propuesto, entonces el contraste será bilateral o de dos colas. Esto implica que se estará evaluando la posibilidad de que el valor del parámetro sea mayor o menor al valor propuesto en la hipótesis nula, lo que amplía el rango de posibles resultados.

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$

Nótese que el área de rechazo definida por un contraste de este tipo (Figura 4.1) quedaría situada a ambos extremos de la distribución, puesto que debería rechazarse la hipótesis nula tanto cuando $u > u_0$ como cuando $u < u_0$. Por otra parte, si la hipótesis se plantea de tal forma que se busca determinar únicamente que el parámetro de estudio de la población tome un valor superior (o análogamente inferior) a un valor dado, el contraste será unilateral o de una cola.

$$H_0 : \theta \leq \theta_0 \quad H_0 : \theta \geq \theta_0$$

$$H_1 : \theta > \theta_0 \quad H_1 : \theta < \theta_0$$

En este ejemplo específico, la región crítica para el contraste se ubicaría en el extremo derecho de la distribución si como resultado del contraste se rechaza H_0 cuando $u > u_0$, o, por el contrario, a la izquierda de la distribución si rechaza H_0 cuando $u < u_0$.

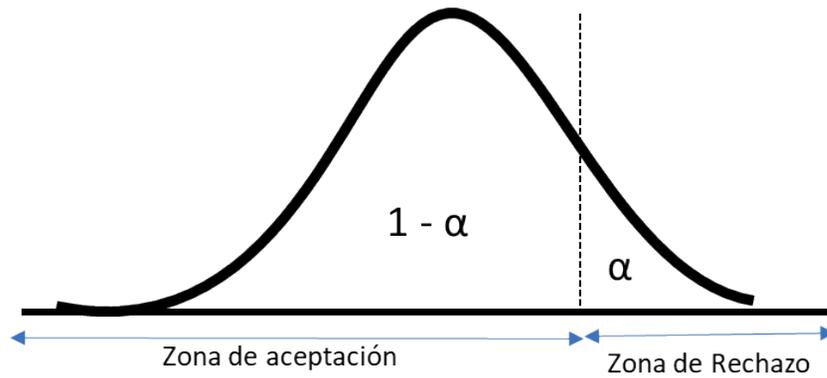


Figura 4.3

Contraste de hipótesis unilateral

Nota. Elaboración propia.

Si se trabaja con los datos del ejemplo 4.1, para demostrar que el nivel promedio de lactato en los deportistas es superior a 4 mmol/L es colocar este supuesto en la hipótesis alternativa. Las hipótesis podrían formularse del siguiente modo para el contraste unilateral:

$$H_0 : \theta \leq 4$$

$$H_1 : \theta > 4$$

Es relevante tener claridad sobre el tipo de contraste a utilizar, ya que esto afectará la interpretación de los resultados y la elección de la prueba estadística adecuada. Además, es importante considerar el tamaño de muestra, el nivel de significancia y el poder estadístico al realizar contrastes de hipótesis, con el fin de obtener conclusiones válidas y confiables a partir del análisis estadístico.

En este contexto, tanto el contraste como la distribución de probabilidad corresponden exactamente a lo descrito en el contraste bilateral previamente llevado a cabo.

$$z = \frac{x - \mu}{s / \sqrt{n}} = \frac{6,28 - 4}{2,1 / \sqrt{46}} = 7,36$$

Si se asume que los datos siguen una distribución t de Student con n-1 grados de libertad y el mismo nivel de significancia que en el ejemplo original, la región crítica de contraste será modificada. La nueva regla de decisión quedará de la siguiente manera:

- Si $z > 1,677$ se rechaza la hipótesis nula
- Si $z < 1,677$ se acepta la hipótesis nula

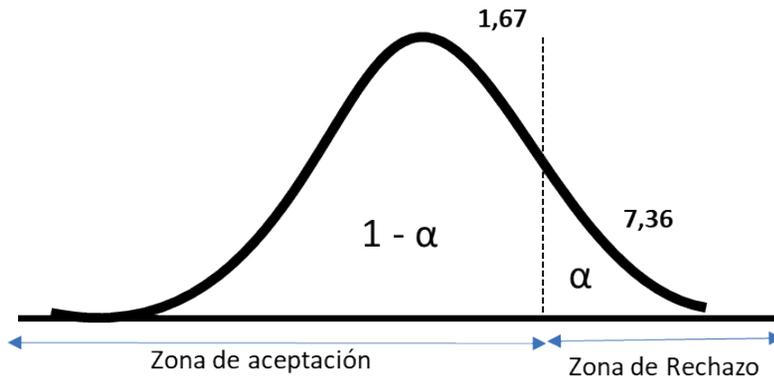


Figura 4.4

Contraste de hipótesis unilateral

Nota. Elaboración propia.

Nótese que con los contrastes unilaterales la región crítica se agranda en dirección a la discrepancia observada, en los contrastes bilaterales se considera la posibilidad de que la discrepancia pueda ocurrir en ambas direcciones. Esto significa que los contrastes bilaterales son más conservadores y permiten extraer conclusiones en cualquier dirección de discrepancia de los datos observados.

En el ejemplo anterior, se menciona que, si se hubiera realizado un contraste bilateral en lugar de un contraste unilateral, se habría concluido que la hipótesis nula debía rechazarse, lo que indicaría que el valor del promedio poblacional de lactato es significativamente distinto de 4 mmol/L, ya sea mayor o menor. Esto resalta la ventaja de los contrastes bilaterales, ya que no se limitan a una dirección específica y consideran todas las posibles discrepancias en los datos.

Sin embargo, es importante tener en cuenta que la elección entre un contraste unilateral y un contraste bilateral depende del contexto y del conocimiento previo sobre el fenómeno objeto de estudio. Los contrastes unilaterales pueden ser apropiados cuando se tiene una clara dirección o expectativa sobre el efecto o la relación que se espera encontrar (por ejemplo, valoración de un nuevo tratamiento que solo puede dejar igual o mejorar el estado de salud del paciente, pero no empeorarlo). Por otro lado, los contrastes bilaterales son más apropiados cuando no se tiene una dirección específica en mente y se quiere considerar la posibilidad de discrepancias en ambas direcciones.

Contraste de hipótesis sobre una proporción

El ejemplo 4.2 planteaba la pregunta de si la proporción de éxito del tratamiento, basada en 180 respuestas positivas de un total de 250 individuos que recibieron el tratamiento, es del 0,8 o del 80%. Para responder a esta pregunta, se requerirá realizar un contraste de hipótesis sobre la proporción poblacional de respuesta positiva al tratamiento, siguiendo los pasos establecidos previamente.

$$H_0 : p = 0,8$$

$$H_1 : p \neq 0,8$$

El indicador de discrepancia o estadístico de contraste en el caso de una proporción poblacional se refiere a la medida utilizada para comparar la proporción observada en la muestra con la proporción esperada bajo la hipótesis nula. Esta será:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1 - p_0)}{n}}}$$

Donde p_0 es el valor de la proporción poblacional que se describe en la hipótesis nula, \hat{p} es la proporción calculada a partir del conjunto de datos de la muestra y n es el tamaño de la muestra. En este caso se tendrá que:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}} = \frac{\frac{180}{250} - 0,8}{\sqrt{\frac{\frac{180}{250} \left(1 - \frac{180}{250}\right)}{250}}} = 2,81$$

Se puede notar que en este caso se satisfacen las condiciones para utilizar la aproximación normal. En caso contrario, se necesitaría usar la distribución binomial exacta. Al utilizar un nivel de significancia convencional de $\alpha=0,05$ y dado que el valor del estadístico de contraste ($z=2,81$) supera el valor crítico de 1,96, se encuentra en la región de rechazo del contraste, lo que implica el rechazo de la hipótesis nula. En consecuencia, se puede afirmar que la proporción de individuos en la población con respuesta positiva al tratamiento es significativamente diferente de 0,8 (o del 80%). Debido a que el rechazo se produjo en el extremo derecho de la distribución, se puede inferir que la proporción poblacional es significativamente mayor que 0,8.

CAPÍTULO 5. PRUEBAS ESTADÍSTICAS



Prueba Chi-cuadrado

La prueba estadística de Chi-cuadrado es una herramienta ampliamente utilizada en la estadística inferencial para analizar datos categóricos, es decir, datos que se pueden clasificar en categorías o grupos. Es especialmente útil cuando se desea determinar si existe una asociación significativa entre dos o más variables categóricas. Si se trabaja con los datos del ejemplo 4,2 en primer lugar será necesario construir una tabla 2x2 como se muestra en la tabla 5.1.

Una posible hipótesis de contraste puede ser que las proporciones de respuestas positivas al tratamiento sean iguales en hombres y mujeres, en comparación con la alternativa de que sean diferentes. Por lo tanto, se plantea la siguiente hipótesis:

$$H_0 : P_1 = P_2$$

$$H_1 : P \neq P_2$$

Para calcular el estadístico de contraste, es necesario considerar cuáles serían las frecuencias esperadas en cada celda de la tabla si la hipótesis nula fuera verdadera. En caso de que las dos proporciones de respuestas positivas al tratamiento fueran iguales en hombres y mujeres, habría una única proporción global de respuestas positivas al tratamiento, la cual sería:

$$\hat{p} = \frac{180}{250} = 0,72$$

Posteriormente, se esperaría que, tanto en hombres como en mujeres, el 72% de los individuos muestre una respuesta positiva, mientras que el 28% tenga una respuesta negativa en ambos grupos. Por lo tanto, se obtendría la siguiente situación:

$$\text{Frecuencia esperada respuesta positiva}(H) = 0,72 * 150 = \frac{180 * 150}{250} = 108$$

Tabla 5.1

Distribución de frecuencias de respuesta al tratamiento según sexo

Respuesta al tto	Hombres	Mujeres	Total
Positiva	110	70	180
Negativa	40	30	70

Nota. Elaboración propia.

Tabla 5.2

Frecuencias esperadas bajo la hipótesis nula

Respuesta al tto	Hombres	Mujeres	Total
Positiva	108	70	180
Negativa	42	28	70

Nota. Elaboración propia.

$$Frecuencia\ esperada\ respuesta\ positiva(m) = 0,72 * 100 = \frac{180 * 100}{250} = 72$$

De esta manera, es posible elaborar una tabla con las frecuencias que se esperan. En la tabla 5.2 se presentan todas las frecuencias esperadas obtenidas. Una forma rápida de calcular los valores esperados en cada celda se basa en los totales por fila, columna y total. Por ejemplo, para obtener la frecuencia esperada E_{ij} se calculará:

$$E_{ij} = \frac{Total\ por\ fila - total\ por\ columna}{total}$$

En última instancia, la medida de discrepancia utilizada en este caso se calculará como la diferencia entre las frecuencias observadas en cada una de las categorías y los valores esperados asumiendo que las proporciones son iguales, de acuerdo con la hipótesis nula. De esta manera, la expresión matemática del estadístico de contraste quedará de la siguiente forma:

$$x^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Así, se tendrá que:

$$x^2 = \frac{(110 - 108)^2}{108} + \frac{(70 - 72)^2}{72} + \frac{(40 - 42)^2}{42} + \frac{(30 - 28)^2}{28} = 0.331$$

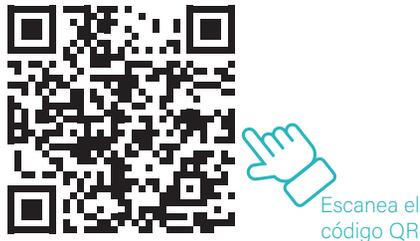
La distribución del estadístico de contraste seguirá un modelo de distribución ji-cuadrado con 1 grado de libertad. Es necesario que las frecuencias esperadas en cada celda sean iguales o superiores a 5 para garantizar la validez de la distribución muestral. En caso de que alguna o algunas de las celdas tengan una frecuencia esperada inferior a 5, se podría utilizar la distribución ji-cuadrado, siempre y cuando el porcentaje de celdas con esta situación no supere el 20%. En el ejemplo dado, todas las celdas tienen valores esperados superiores a 5.

Si las frecuencias esperadas bajo la hipótesis nula coinciden con las observadas, el valor del estadístico de contraste sería 0. Sin embargo, a medida que las diferencias entre las

frecuencias esperadas y observadas sean mayores, mayor será la evidencia en contra de la hipótesis nula. Por lo tanto, la región crítica de contraste estará ubicada únicamente en el extremo derecho de la distribución.

Dado que el valor del estadístico de contraste $x^2 = 0,331$ es menor que el valor crítico de la región de rechazo (3,841), se puede concluir que la hipótesis nula será aceptada. Por lo tanto, no se encontró evidencia suficiente para demostrar diferencias en las proporciones de respuesta positiva al tratamiento entre hombres y mujeres.

Con el objetivo de simplificar el proceso estadístico, el presente libro busca brindar una comprensión clara al lector sobre cómo surgen matemáticamente los valores de los diferentes estadísticos de contraste; sin embargo, se utiliza el software estadístico SPSS, el cual facilita el proceso ejecutando todo lo acá planteado de manera rápida y sencilla. A continuación, se proporciona un enlace en el cual se explican detalladamente los supuestos para el análisis de eventos.



Prueba t para muestras independientes

La prueba t para muestras independientes es una técnica estadística utilizada para comparar las medias de dos grupos diferentes o mutuamente excluyentes, con el objetivo de determinar si existen diferencias significativas entre ellos. Es una herramienta ampliamente utilizada en la investigación y análisis de datos en diversas disciplinas, como la psicología, la medicina, la educación y las ciencias sociales.

En primer lugar, es importante entender que la prueba t para muestras independientes se utiliza cuando se desea comparar dos grupos diferentes, donde los individuos de un grupo no están relacionados o emparejados con los individuos del otro grupo. Por ejemplo, se podría utilizar para comparar las calificaciones de dos grupos de estudiantes expuestos a diferentes métodos de enseñanza, o para comparar los niveles de ansiedad en dos grupos de pacientes sometidos a diferentes tratamientos médicos.

La prueba t para muestras independientes se basa en el supuesto de que las poblaciones subyacentes de ambos grupos siguen una distribución normal, es decir, que las puntuaciones de ambos grupos se distribuyen en forma de campana con una media y una desviación estándar bien definidas. Además, se asume que las varianzas de ambos grupos son iguales, lo cual se conoce como homogeneidad de varianzas. Es importante tener en

cuenta estos supuestos antes de realizar la prueba t para muestras independientes, ya que su validez y precisión dependen de que estos supuestos se cumplan.

Una vez que se tienen los datos de ambos grupos, se puede proceder a realizar la prueba t. Esta prueba compara las medias de los dos grupos y determina si las diferencias observadas son lo suficientemente grandes como para considerarse significativas o si podrían haberse producido por casualidad. El resultado de la prueba t es un valor t, que se compara con un valor crítico para determinar si la diferencia entre las medias es significativa.

La fórmula para calcular la prueba t para muestras independientes es la siguiente:

$$t = \frac{M_1 - M_2}{\sqrt{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Donde:

M1 y M2 son las medias de los dos grupos que se están comparando.

S^2 Es una estimación combinada de la varianza de ambos grupos, llamada varianza combinada. se calcula:

$$S^2 = \frac{\sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2 + \sum_{j=1}^{n_2} (x_j - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

n_1 y n_2 son los tamaños de muestra de los dos grupos.

Una vez que se ha calculado el valor t, se compara con el valor crítico correspondiente, que se encuentra en tablas de distribución t o puede calcularse utilizando un software estadístico. Si el valor t calculado es mayor que el valor crítico, se concluye que hay una diferencia significativa entre las medias de los dos grupos, lo cual implica que hay evidencia estadística para afirmar que las medias no son iguales. Por otro lado, si el valor t calculado es menor que el valor crítico, no se puede afirmar que exista una diferencia significativa entre las medias, lo cual implica que no hay evidencia estadística para afirmar que las medias son diferentes.

Es importante destacar que el resultado de la prueba t no establece la dirección de la diferencia entre las medias. Es decir, no nos indica si una media es mayor o menor que la otra. Simplemente nos indica si hay una diferencia significativa o no. Para determinar la dirección de la diferencia, se debe realizar un análisis adicional.

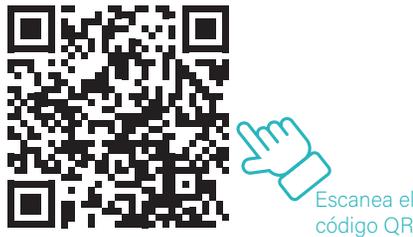
Además de comparar el valor t calculado con el valor crítico, es común utilizar un nivel de significancia o nivel alfa para tomar decisiones. El nivel de significancia es la probabilidad de cometer un error tipo I, es decir, rechazar la hipótesis nula cuando es verdadera. El valor típico utilizado para el nivel de significancia es 0.05, lo que significa que hay un 5% de probabilidad de cometer un error tipo I. Si el valor p obtenido en la prueba t es menor que el nivel de significancia establecido, se rechaza la hipótesis nula y se concluye

que hay una diferencia significativa entre las medias. Si el valor p es mayor que el nivel de significancia, no se rechaza la hipótesis nula y no se puede afirmar que haya una diferencia significativa.

Es importante mencionar que el tamaño de muestra de cada grupo juega un papel importante en la interpretación de los resultados de la prueba t . Muestras más grandes tienden a tener más poder estadístico, lo que significa que son más capaces de detectar diferencias significativas entre las medias. Por lo tanto, es importante tener en cuenta el tamaño de muestra al interpretar los resultados de la prueba t .

Es fundamental comprender que la prueba t para muestras independientes solo nos indica si hay una diferencia estadísticamente significativa entre las medias de los dos grupos, pero no nos proporciona información sobre la magnitud o importancia clínica o práctica de la diferencia. Para evaluar la relevancia práctica de la diferencia, se requiere un análisis adicional que considere el contexto del estudio y los objetivos de investigación.

A continuación, se proporciona una lista de videos que ofrecen una explicación más detallada de los supuestos y la aplicación de la prueba t para muestras independientes en el contexto del software SPSS, con el objetivo de facilitar una comprensión más completa de los conceptos previamente planteados.



Prueba t para muestras relacionadas

En estadística, las pruebas t para muestras relacionadas son una herramienta valiosa para analizar datos en situaciones donde se comparan dos conjuntos de observaciones relacionadas o emparejadas. Estas pruebas permiten a los investigadores evaluar si hay una diferencia significativa entre las medias de dos grupos de datos relacionados o repetidos en el tiempo, en lugar de comparar medias de dos grupos independientes, como se hace en las pruebas t para muestras independientes.

Las pruebas t para muestras relacionadas se aplican en diversas situaciones en las que los datos se obtienen de la misma muestra o grupo de individuos en dos momentos diferentes o en dos condiciones diferentes. Por ejemplo, en estudios pre y post intervención, donde se mide una variable antes y después de aplicar una intervención o tratamiento, o en estudios en los que se comparan dos técnicas o condiciones diferentes en el mismo grupo de individuos. También se utilizan en estudios de casos y controles, en los cuales

los casos y los controles se emparejan en función de ciertas características antes de realizar las mediciones.

El análisis de datos emparejados o relacionados tiene ventajas importantes en comparación con el análisis de datos independientes. Al utilizar pruebas t para muestras relacionadas, se pueden controlar factores de confusión y variabilidad individual que pueden afectar a las mediciones, ya que cada individuo sirve como su propio control. Esto reduce la variabilidad en la estimación de los efectos y aumenta la precisión del análisis.

El procedimiento para realizar una prueba t para muestras relacionadas es similar al de la prueba t para muestras independientes en términos de cálculo del estadístico t y la obtención del valor p. Sin embargo, en este caso, la diferencia entre las mediciones emparejadas o relacionadas es el enfoque central del análisis.

La fórmula planteada para determinar el estadístico es:

$$t = \frac{XD - \mu D}{\frac{S_d}{\sqrt{n}}}$$

Donde:

t: Es el valor del estadístico de prueba t.

XD: Es la media de las diferencias entre las observaciones relacionadas.

μD : Es la media esperada de las diferencias (generalmente se asume que es 0, lo que implica que no hay diferencia entre las 2 muestras).

Sd: Es la desviación estándar de las diferencias entre las observaciones relacionadas.

n: Es el número de pares de observaciones relacionadas.

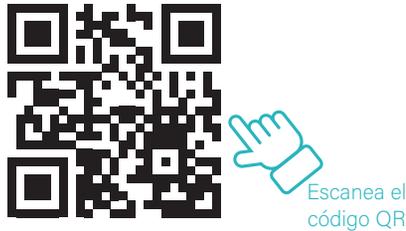
El estadístico t se calcula dividiendo la diferencia promedio entre las mediciones emparejadas por la desviación estándar de las diferencias y se compara con un valor crítico o se utiliza un nivel de significancia establecido para determinar si hay una diferencia significativa entre las medias. Un valor t calculado mayor que el valor crítico o un valor p menor que el nivel de significancia establecido indicaría una diferencia estadísticamente significativa entre las medias.

Al igual que en las pruebas t para muestras independientes, el valor p obtenido en una prueba t para muestras relacionadas también se utiliza para tomar decisiones. Si el valor p es menor que el nivel de significancia establecido, se rechaza la hipótesis nula y se concluye que hay una diferencia significativa entre las medias. Si el valor p es mayor que el nivel de significancia, no se rechaza la hipótesis nula y no se puede afirmar que haya una diferencia significativa.

Es importante tener en cuenta que las pruebas t para muestras relacionadas también tienen supuestos que deben cumplirse para garantizar la validez de los resultados. Uno de

los supuestos clave es que las diferencias entre las mediciones emparejadas deben seguir una distribución normal. Además, la correlación entre las mediciones emparejadas debe ser tomada en cuenta al interpretar los resultados.

En el siguiente video se explica con mayor detalle los supuestos a tener en cuenta a la hora de aplicar el estadístico t student para muestras relacionadas.



Y un video con un ejemplo práctico.



Comparativa de tres o más grupos independientes (ANOVA)

En muchas ocasiones, los investigadores buscan analizar la posible relación entre una variable cuantitativa y una variable cualitativa o factor. Por lo general, se compara la media de la variable cuantitativa en cada grupo o nivel de la variable cualitativa. Si se encuentran diferencias significativas entre las medias, se puede inferir una relación entre las variables, ya que la variable cuantitativa muestra un comportamiento diferencial en cada uno de los grupos analizados.

Por ejemplo, se puede plantear la pregunta de investigación sobre si existen diferencias en los niveles promedio de colesterol (en mg/100 ml) según el nivel de obesidad, clasificando la obesidad en tres grupos: delgado-normal ($IMC < 25$), sobrepeso ($25 \leq IMC < 30$) y obesidad ($IMC \geq 30$). Otra pregunta de investigación podría ser si existe una relación entre el nivel de colesterol (en mg/100 ml) y el consumo de alcohol, clasificando el consumo de alcohol en tres niveles: nunca, moderado y alto. En ambos casos, se puede utilizar el análisis de diferencia de medias con ANOVA y pruebas post hoc para comparar las medias de la variable cuantitativa en cada grupo o nivel de la variable cualitativa y determinar si hay diferencias significativas.

En resumen, en el análisis de diferencia de medias con ANOVA se busca comparar las medias de una variable cuantitativa en tres o más grupos o niveles de una variable cualitativa, y se utilizan pruebas post hoc para identificar cuáles de los grupos difieren significativamente entre sí. El ANOVA es una extensión natural de la prueba t para muestras independientes, pero aplicada a la comparación de 3 o más medias en lugar de solo 2.

ANOVA de efectos fijos y aleatorios

El ANOVA puede ser de efectos fijos o de efectos aleatorios, o incluso una combinación de ambos, dependiendo de las características de la variable que define los grupos o factores. Cuando los niveles o grupos del factor son los únicos de interés o los únicos que pueden tener un efecto sobre la variable dependiente, se denominará factor fijo. Por ejemplo, variables como sexo, estado civil, etc., se consideran factores fijos. En cambio, si los niveles del factor constituyen una muestra de una población de niveles más amplia, se denominará efecto aleatorio. Por ejemplo, si se disponen de observaciones de determinadas características para una muestra de ciudades, países, hospitales, etc., se consideraría un efecto aleatorio.

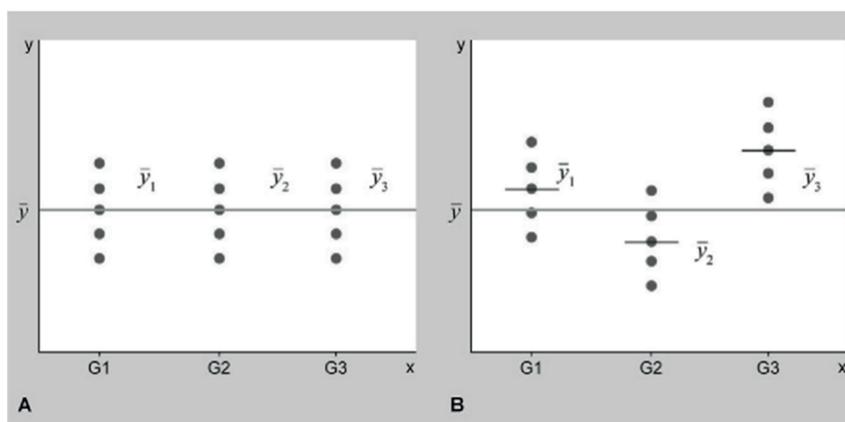


Figura 5.1

Diagrama de dispersión de una variable cuantitativa (y) según el grupo de pertenencia (G1, G2 o G3)

Nota. Tomado de Missau et al. (2015).

En la Figura 5.1 se pueden visualizar dos escenarios derivados de la comparación de la distribución de una variable cuantitativa en tres grupos, definidos por una variable cualitativa. Por ejemplo, la variable “y” podría representar el nivel de colesterol y la variable “x” el nivel de obesidad. En la figura 5.1A, las medias de los niveles de colesterol en cada

grupo o nivel de obesidad son iguales, lo que resulta en una coincidencia con la media global. En cambio, en la figura 5.2B se observa que las medias de los niveles de colesterol difieren entre los grupos, lo que implica una diferencia con la media global. Es importante tener en cuenta que estos resultados se obtienen de una muestra de una población mucho mayor y que el objetivo es establecer si existen diferencias significativas entre las medias poblacionales a través de pruebas de hipótesis, donde la hipótesis nula sería evaluada. La construcción de la hipótesis nula sería:

$$H_0 = \mu_1 = \mu_2 = \mu_3$$

Por lo tanto, la hipótesis alternativa contemplaría el caso en el que al menos una de las medias poblacionales es diferente del resto.

Coeficiente de determinación

En el proceso de la descomposición de la variabilidad de la variable de estudio surge una medida que puede ser interpretada como la proporción de variabilidad de la variable cuantitativa o dependiente explicada por la variable que define los grupos o factores. Esta recibe el nombre de coeficiente de determinación. Por lo tanto, se tiene que:

$$\text{Coeficiente de determinación} = R^2 = \frac{VE}{VNE} = \frac{\sum(y_i - \bar{y})^2}{\sum(y_{ij} - y_i)^2} * 100$$

Donde:

VE: Variabilidad explicada por la variable explicativa o factor

VNE: Variabilidad no explicada por la variable explicativa o factor.

Para ilustrar todo lo anterior se propone el siguiente ejemplo.

Ejemplo 5.1

En un estudio se recopilaron datos acerca del nivel de colesterol en mg/100 ml y el nivel de obesidad (clasificado como delgado-normal, sobrepeso y obesidad) de un grupo de 25 pacientes. Los resultados obtenidos se presentan en la tabla 5.3

Tabla 5.3

Datos nivel obesidad y nivel de colesterol

Nivel de obesidad	Nivel de colesterol	n
Delgado-normal	145, 160, 166, 173, 202, 218, 224, 251, 172	9
Sobrepeso	156, 232, 237, 239, 245, 247, 302, 241, 235, 240	10
Obesidad	154, 228, 236, 242, 245, 266, 269, 305, 280, 275, 304	11

Nota. Tomado de Missau et al. (2015).

Si se calculan las medias del nivel de colesterol en cada uno de los grupos o niveles del factor obesidad y la media global se obtendrá que:

$$\underline{Y} = \frac{145 + 160 + 166 + 173 \dots 304}{30} = 229,63$$

$$\underline{Y}_1 = \frac{145 + 160 + 166 + 173 + 202 + 218 + 224 + 251 + 172}{9} = 190,11$$

$$\underline{Y}_2 = \frac{156 + 232 + 237 + 239 + 245 + 247 + 302 + 241 + 235 + 240}{10} = 237,4$$

$$\underline{Y}_3 = \frac{154 + 228 + 236 + 242 + 245 + 266 + 269 + 305 + 280 + 275 + 304}{11} = 254,91$$

A continuación, se calculan las sumas de cuadrados correspondientes a las variabilidades totales, explicada y no explicada, de la siguiente forma:

$$VE = \sum(y_i - \bar{y})^2 = 9(190,11 - 229,63)^2 + 10(237,11 - 229,63)^2 + 11(254,91 - 229,63)^2$$

$$VNE = \sum(y_{ij} - \bar{y}_i)^2 = (145 - 190,11)^2 + (160 - 190,11)^2 \dots$$

$$(304 - 190,11)^2 + (237 - 229,63)^2 + (232 - 237,4)^2 \dots$$

$$(240 - 237,4)^2 + (154 - 254,91)^2 + (228 - 254,91)^2 \dots (304 - 254,91)^2$$

Al hacer los cálculos se obtendrá que:

$$VT = 60416,967; VNE 38728,198; VE = 21688,769$$

donde la suma de la variabilidad explicada por el factor y la no explicada dará como resultado la variabilidad total:

$$VT = VE + VNE = 21688,769 + 38728,198 = 60416,967$$

Por su parte, el coeficiente de determinación quedará de la siguiente forma:

$$\frac{VE}{VNE} = \frac{21688,769}{38728,198} = 0,56$$

Este resultado implicaría que el factor nivel de obesidad lograría explicar el 56% de la variabilidad observada en la variable dependiente nivel de colesterol.

Inferencia y tabla ANOVA

Cuando se realiza la comparación de tres o más medias a partir de datos recopilados en una muestra aleatoria de una población, es importante tener en cuenta que las medias muestrales calculadas en cada uno de los grupos o niveles del factor son solo una de las muchas medias posibles que podrían haberse obtenido a partir de diferentes muestras posibles de la población en el proceso de muestreo. Esto significa que para cada muestra se obtendrán medias que podrían ser similares o diferentes a las de otra muestra, ya que los datos iniciales pueden variar.

La Figura 5.2 ilustra cómo las medias de los grupos pueden diferir en función de la muestra aleatoria seleccionada. Por lo tanto, las medias muestrales de los grupos son variables aleatorias que varían de una muestra a otra en la población. Es importante tener en cuenta esta variabilidad inherente al proceso de muestreo al interpretar y generalizar los resultados obtenidos al comparar medias de tres o más grupos en un estudio.

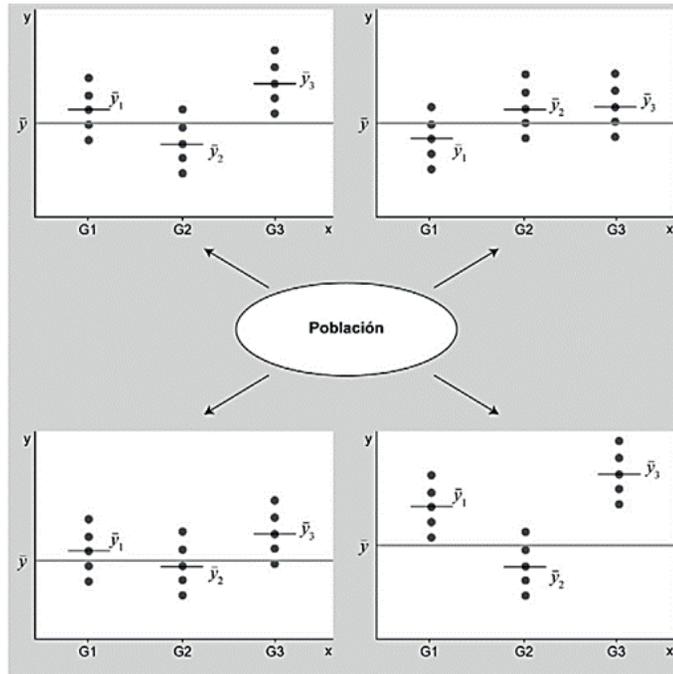


Figura 5.2

Diferencias en las medias en función de la muestra seleccionada

Nota. Tomado de Missau et al. (2015).

Un contraste de interés busca determinar si todas las medias poblacionales de los diferentes grupos o niveles del factor son iguales o si alguna o algunas de ellas son diferentes de las demás. Si se acepta la hipótesis nula, no habría evidencia de diferencias significativas entre las medias poblacionales. Sin embargo, si se rechaza la hipótesis nula, esto indicaría que al menos una de las medias poblacionales es diferente. El estadístico de contraste se utiliza en este caso para realizar dicha evaluación.

$$f = \frac{VE / (k - 1)}{VNE / (n - k)}$$

El estadístico de contraste utilizado en este caso tiene una distribución F de Snedecor con $k-1$ y $n-k$ grados de libertad, donde k representa el número de grupos y n es el tamaño de la muestra global ($n = n_1 + n_2 + n_3$, donde n es el tamaño de la muestra en cada grupo). En caso de que se cumplan las hipótesis necesarias que se abordarán más adelante, este estadístico de contraste se distribuirá de esta manera.

Es importante destacar que cuando la variabilidad explicada por el modelo es igual a 0, el valor del estadístico de contraste también será 0, lo que concuerda con la hipótesis nula

de que no existen diferencias significativas entre las medias de los grupos. En contraste, si la variabilidad explicada aumenta (numerador mayor), la variabilidad no explicada debe disminuir (denominador menor), lo que conducirá a un valor del estadístico de contraste cada vez mayor. Si las medias de los grupos son significativamente diferentes, esto podría resultar en el rechazo de la hipótesis nula.

Por lo general, los resultados se presentan en una tabla ANOVA (Tabla 5.4) que muestra la secuencia de razonamiento que conduce a la construcción del estadístico de contraste, el cual se basa principalmente en la varianza explicada entre grupos y la varianza no explicada dentro de los grupos.

Tabla 5.4

Tabla ANOVA

Fuente	Suma de cuadrados	Media de cuadrados	Cociente de varianzas
Variabilidad entre grupos (VE)	$\sum(y_i - \bar{y})^2$	$\frac{\sum(y_i - \bar{y})^2}{k-1}$	$f = \frac{VE / (k-1)}{VNE / (n-k)}$
Variabilidad intra-grupos (VNE)	$\sum(y_{ij} - \bar{y}_i)^2$	$\frac{\sum(y_{ij} - \bar{y}_i)^2}{n-k}$	
Total (VT)	$\sum(y_{ij} - \bar{y}_i)^2$	$\frac{\sum(y_{ij} - \bar{y}_i)^2}{n-1}$	

Nota. Elaboración propia.

Se obtendría un estadístico de contraste $EC = 7,56$, que al verificar en las tablas de la distribución F de Snedecor con 2 y 27 grados de libertad, se obtendría un valor de p de 0,002. Al ser el valor de p del contraste menor al nivel usual de 0,05, se rechazaría la hipótesis nula y se llegaría a la conclusión de que al menos una media poblacional difiere de las demás, tal y como se muestra en la Tabla 5.5.

Tabla 5.5

Tabla ANOVA para el ejemplo 5.1

Fuente	Suma de cuadrados	Media de cuadrados	Cociente de varianzas
Variabilidad entre grupos (VE)	21.688,769	10.844,38	$\frac{10.844,38}{1.434,38} = 7,56$
Variabilidad intragrupos (VNE)	38.728,198	1.434,38	
Total (VT)	60.416,96	2.083,34	

Nota. Elaboración propia.

Ahora bien, es importante tener en cuenta que la aplicación del método de ANOVA para comparar tres o más medias de muestras independientes requiere que se cumplan ciertos requisitos en relación con las variables y observaciones involucradas. Estos requisitos se pueden resumir en los siguientes puntos:

- ⊙ La distribución de las observaciones en cada uno de los grupos o niveles del factor debe seguir una distribución normal. En la práctica, se recomienda contar con un mínimo de 30 datos por grupo para asegurar una aproximación a la normalidad.
- ⊙ La varianza debe ser homogénea en todos los grupos o niveles del factor, lo que implica que las variabilidades de las muestras deben ser similares.
- ⊙ Las observaciones en cada uno de los grupos o niveles del factor deben constituir una muestra aleatoria y ser independientes entre sí.

Para comprobar estas hipótesis, se utilizan recursos gráficos y analíticos, tales como

- ⊙ Criterios gráficos: se pueden utilizar histogramas y gráficos P-P normal para la variable cuantitativa en su totalidad y para cada uno de los grupos o niveles del factor considerados, así como gráficos comparativos de cajas (box-plots) y diagramas de dispersión de la variable cuantitativa para cada uno de los grupos o niveles del factor. También se pueden usar gráficos de secuencia, dependiendo de la forma en que se hayan obtenido los datos.
- ⊙ Criterios analíticos: se pueden aplicar pruebas estadísticas como el Test de Kolmogorov-Smirnov o el contraste ji-cuadrado de bondad de ajuste para comprobar la hipótesis de normalidad de los datos. Además, se puede utilizar el Test de Levene para comprobar la hipótesis de homogeneidad de varianzas, el cual es menos exigente en términos de la hipótesis de normalidad que la prueba de Barlett.

Estos recursos gráficos y analíticos son útiles en la comprobación de las hipótesis necesarias para la aplicación del método de ANOVA en la comparación de medias de muestras independientes.

Pruebas post hoc

Ahora bien, se dejó plasmado anteriormente que el análisis de varianza (ANOVA) es una técnica estadística utilizada para comparar tres o más medias de diferentes grupos o niveles de un factor. Esta es una herramienta poderosa para determinar si existe alguna diferencia significativa entre las medias de los grupos en estudio. Sin embargo, cuando se encuentra una diferencia significativa en los resultados del ANOVA, es necesario realizar pruebas post hoc para identificar cuáles de las medias difieren entre sí.

El término “post hoc” proviene del latín y significa “después de esto”. En el contexto del análisis de varianza, las pruebas post hoc se realizan después de obtener un resultado significativo en el ANOVA para determinar qué medias son diferentes entre sí. Estas pruebas se utilizan para realizar comparaciones múltiples entre las medias de los grupos, ya que el ANOVA sólo determina si hay diferencias significativas en general, pero no especifica cuáles son las diferencias.

Existen varias pruebas post hoc disponibles, como la prueba de Tukey, la prueba de Bonferroni, la prueba de Scheffé, la prueba de Dunnett, entre otras. Estas pruebas ajustan el nivel de significancia para tener en cuenta el hecho de que se están realizando múltiples comparaciones. Por lo tanto, son más conservadoras en términos de la detección de diferencias significativas, para evitar errores tipo I o falsos positivos.

La elección de la prueba post hoc a utilizar depende del diseño del estudio, el tamaño de la muestra y los supuestos del modelo estadístico. Es importante tener en cuenta que, al realizar múltiples comparaciones, aumenta la probabilidad de encontrar diferencias significativas al azar. Por lo tanto, es fundamental interpretar los resultados de las pruebas post hoc con precaución y considerar el contexto del estudio en su totalidad.

Par dar mayor claridad a todo lo planteado anteriormente, se sugiere revisar la siguiente lista de reproducción en YouTube.

ANOVA para muestras independientes:



Escanea el
código QR

ANOVA para muestras relacionadas:



Escanea el código QR

Coeficiente de correlación de Pearson

En el ámbito de la estadística, el coeficiente de correlación de Pearson es una medida que permite evaluar la dependencia lineal entre dos variables cuantitativas aleatorias. Una de las ventajas de esta medida es que, a diferencia de la covarianza, es independiente de la escala de medida de las variables analizadas.

De forma más sencilla, el coeficiente de correlación de Pearson se puede definir como un indicador que se emplea para medir el grado de relación entre dos variables continuas y cuantitativas. Es decir, permite evaluar la magnitud y dirección de la relación existente entre dos variables, siempre y cuando ambas sean numéricas y continuas.

El coeficiente de correlación de Pearson se calcula utilizando la siguiente fórmula:

$$r = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) * \text{var}(Y)}} = \frac{\sum (X_i - \bar{X})(y_i - \bar{y})}{\sqrt{[\sum (X_i - \bar{X})^2][\sum (y_i - \bar{y})^2]}}$$

La fórmula puede parecer un poco complicada, pero básicamente lo que hace es medir la relación entre dos variables. Si el valor de r es positivo, significa que hay una relación positiva entre las dos variables, es decir, cuando una variable aumenta, la otra también lo hace. Si el valor de r es negativo, significa que hay una relación negativa entre las dos variables, es decir, cuando una variable aumenta, la otra disminuye. Si el valor de r es 0, significa que no hay relación entre las dos variables.

El valor del coeficiente de correlación de Pearson oscila entre -1 y 1. Cuanto más cercano a 1 o -1 esté el valor de r , mayor será la correlación entre las dos variables. Si el valor de r es cercano a 0, significa que no hay una correlación fuerte entre las dos variables.

Un valor de r de 1 indica una correlación positiva perfecta entre las dos variables, es decir, cuando una variable aumenta, la otra también lo hace en una proporción constante. Un valor de r de -1 indica una correlación negativa perfecta entre las dos variables, es decir, cuando una variable aumenta, la otra disminuye en una proporción constante (Figura 5.1).

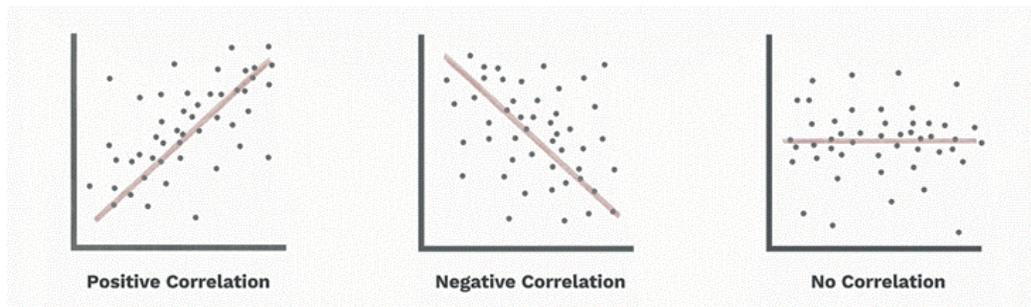


Figura 5.3

Coeficiente de correlación de Pearson

Nota. Tomado de Jason (2021).

Ejemplo 5.2

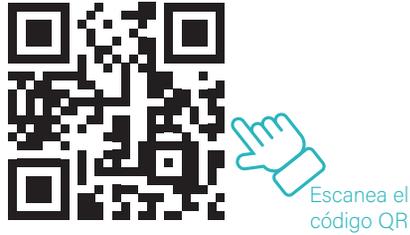
Un estudio reciente publicado en la revista *Nutrients* encontró una correlación positiva significativa entre el consumo de azúcar y la obesidad en niños y adolescentes. Los investigadores utilizaron el coeficiente de correlación de Pearson para medir la relación entre estas dos variables. Encontraron que el valor de r era 0.63, lo que indica una correlación positiva moderadamente fuerte entre el consumo de azúcar y la obesidad.

Sin embargo, a pesar de todo lo mencionado anteriormente, existen algunos supuestos que deben cumplirse para que su uso sea apropiado y los resultados sean confiables:

- ⦿ Las dos variables deben estar distribuidas normalmente: Para utilizar el coeficiente de correlación de Pearson, se asume que las dos variables están distribuidas normalmente. Si no es así, los resultados pueden ser sesgados o poco fiables. Es importante verificar la normalidad de la distribución de las variables antes de aplicar este coeficiente.
- ⦿ Ambas variables deben ser continuas: El coeficiente de correlación de Pearson sólo se puede utilizar para analizar la relación entre dos variables continuas y cuantitativas. No es adecuado para variables categóricas o variables que se miden en una escala ordinal.
- ⦿ Las variables deben ser linealmente relacionadas: El coeficiente de correlación de Pearson mide la relación lineal entre dos variables. Si la relación entre las variables no es lineal, el coeficiente de correlación de Pearson no es una medida adecuada y se necesitaría otro tipo de medida de correlación.
- ⦿ No deben existir valores atípicos: Los valores atípicos o extremos pueden afectar la relación entre las variables y, por lo tanto, pueden distorsionar el resultado del coeficiente de correlación de Pearson.

Por lo tanto, es importante tener en cuenta estos supuestos antes de aplicar el coeficiente de correlación de Pearson para evitar interpretaciones incorrectas de los resultados. Si alguno de estos supuestos no se cumple, se debe considerar el uso de otras medidas de correlación o técnicas estadísticas.

En el siguiente video de YouTube se observa una explicación práctica sobre cómo calcular y analizar el coeficiente de correlación de Pearson.



CAPÍTULO 6. PRUEBAS NO PARAMÉTRICAS



En estadística, las pruebas no paramétricas son una alternativa a las pruebas paramétricas que se utilizan cuando los datos no cumplen con los supuestos de normalidad y homogeneidad de varianza. Las pruebas no paramétricas no requieren que los datos sigan una distribución normal y, por lo tanto, son útiles cuando los datos son muy asimétricos o cuando se sospecha que hay valores atípicos o datos censurados.

Las pruebas no paramétricas son también conocidas como pruebas de distribución libre porque no asumen una distribución específica de los datos. En su lugar, se basan en una estadística de prueba que se calcula a partir de los datos observados y se compara con una distribución de referencia que es independiente de la distribución subyacente de los datos.

Prueba de rangos de signos de Wilcoxon

El test de rangos de Wilcoxon es una prueba no paramétrica que se utiliza para comparar las diferencias entre dos medidas de la misma muestra. Esta prueba es útil cuando los datos son pareados y no se puede suponer que los datos sigan una distribución normal. La prueba se basa en la hipótesis nula de que no hay diferencia significativa entre las dos medidas y la hipótesis alternativa de que hay una diferencia significativa entre ellas.

La prueba de Wilcoxon es una prueba de rango que se basa en la disposición de las diferencias de los datos pareados. En primer lugar, se calculan las diferencias entre las dos medidas pareadas. A continuación, se ordenan las diferencias de menor a mayor en valor absoluto y se les asigna un rango. Si hay empates, se les asigna el rango promedio.

La estadística de prueba para la prueba de Wilcoxon es la suma de los rangos de las diferencias positivas o la suma de los rangos de las diferencias negativas, dependiendo de la dirección de la hipótesis alternativa. Si la hipótesis alternativa es que la medida 1 es mayor que la medida 2, entonces se utiliza la suma de los rangos de las diferencias positivas. Si la hipótesis alternativa es que la medida 2 es mayor que la medida 1, entonces se utiliza la suma de los rangos de las diferencias negativas.

La prueba de Wilcoxon se compara con la distribución de referencia de la suma de los rangos de las diferencias positivas o negativas. La distribución de referencia se calcula a partir de la suma de los rangos de las diferencias positivas o negativas para todas las permutaciones posibles de las medidas pareadas. El número de permutaciones posibles es igual a 2 elevado al número de pares de datos.

La fórmula establecida para el estadístico de Wilcoxon es:

$$W = \min(R+; R-)$$

donde R^+ es la suma de los rangos de las diferencias positivas y R^- es la suma de los rangos de las diferencias negativas.

Si el valor de la estadística de prueba es mayor que el valor crítico de la distribución de referencia, entonces se rechaza la hipótesis nula y se concluye que hay una diferencia significativa entre las dos medidas. Si el valor de la estadística de prueba es menor que el valor crítico de la distribución de referencia, entonces se acepta la hipótesis nula y se concluye que no hay una diferencia significativa entre las dos medidas.

El test de rangos de Wilcoxon es una prueba robusta que no requiere que los datos sigan una distribución normal y es menos sensible a los valores atípicos que otras pruebas paramétricas. Sin embargo, esta prueba tiene menos poder estadístico que las pruebas paramétricas cuando los datos sí siguen una distribución normal.

Ejemplo 6.1

Supongamos que un grupo de 10 estudiantes ha tomado un examen de matemáticas en dos ocasiones diferentes (antes y después de un curso de tutoría) para evaluar si el curso ha mejorado su rendimiento en el examen. Los resultados de los estudiantes se presentan en la siguiente tabla:

Tabla 6.1

Notas estudiantes

Estudiante	Antes de la tutoría	Después de la tutoría
1	60	70
2	75	80
3	80	85
4	65	70
5	70	75
6	55	60
7	85	90
8	90	95
9	75	80
10	65	70

Nota. Elaboración propia.

Para realizar el test de rangos de Wilcoxon, primero se deben calcular las diferencias entre las dos medidas para cada estudiante:

Tabla 6.2

Diferencia entre las medidas relacionadas

Estudiante	Antes de la tutoría	Después de la tutoría	Diferencia
1	60	70	10
2	75	80	5
3	80	85	5
4	65	70	5
5	70	75	5
6	55	60	5
7	85	90	5
8	90	95	5
9	75	80	5
10	65	70	5

Se asignan los rangos a cada diferencia. Para las diferencias repetidas, se asigna el rango promedio:

Tabla 6.3

Tabla descriptiva test Wilcoxon

Estudiante	Nota en matemáticas	Nota en tutoría	Diferencia	Rango
1	70	65	5	9.5
2	85	75	10	6.5
3	90	80	10	6.5
4	80	75	5	9.5
5	75	70	5	9.5
6	70	65	5	9.5
7	95	90	5	9.5
8	85	80	5	9.5
9	75	80	-5	3.5
10	65	70	-5	3.5

Nota. Elaboración propia.

Se suman los rangos de las diferencias positivas y de las diferencias negativas:

$$R+ = 1.5 + 1.5 + 1.5 + 1.5 + 1.5 + 1.5 + 1.5 + 1.5 = 12$$

$$R- = 9.5 + 9.5 = 19$$

La estadística de prueba W es el mínimo entre $R+$ y $R-$:

$$W = \min (R+, R-) = \min (12, 19) = 12$$

Para evaluar la significancia estadística de W , se utiliza una tabla de valores críticos del test de rangos de Wilcoxon. La tabla se basa en el número de pares de datos, en este caso $n = 10$, y en un nivel de significancia α , que se suele establecer en 0.05.

La tabla indica que el valor crítico para $n = 10$ y $\alpha = 0.05$ es 11. Como W (12) es mayor que el valor crítico (11), se puede rechazar la hipótesis nula y concluir que hay evidencia estadística para afirmar que el curso de tutoría ha mejorado el rendimiento de los estudiantes en el examen de matemáticas.

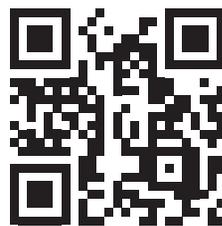
En los siguientes videos se profundizan y ponen en práctica los temas tratados en esta sección:

Supuestos teóricos:



Escanea el código QR

Ejemplo práctico:



Escanea el código QR

Prueba de McNemar

El estadístico de McNemar es una prueba no paramétrica que se utiliza para evaluar la diferencia en la proporción de ocurrencias de un evento entre dos grupos relacionados. Se utiliza comúnmente en estudios de casos y controles, donde se comparan dos grupos de

sujetos que han sido apareados de alguna manera, como, por ejemplo, mediante el género, la edad, la condición médica, entre otros.

La prueba de McNemar se basa en una tabla de contingencia 2x2 que contiene el número de casos que presentan o no el evento de interés en ambos grupos. El estadístico de prueba de McNemar se calcula como la diferencia de proporciones entre los casos discordantes (es decir, aquellos casos en los que se observa el evento en uno de los grupos, pero no en el otro) dividida por la suma de los casos discordantes.

La fórmula para el estadístico de McNemar es la siguiente:

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

donde:

b: número de casos en los que se observa el evento en el primer grupo, pero no en el segundo.

c: número de casos en los que se observa el evento en el segundo grupo, pero no en el primero.

El estadístico de McNemar sigue una distribución chi-cuadrado con un grado de libertad, por lo que se puede utilizar para calcular el valor p y determinar si la diferencia en las proporciones es estadísticamente significativa.

Ejemplo 6.2

Supongamos que se realizó un estudio para evaluar la efectividad de un nuevo tratamiento para reducir los síntomas de la migraña en un grupo de pacientes. Se registró si cada paciente experimentó una reducción en el número de episodios de migraña después de recibir el tratamiento, en comparación con su número de episodios antes del tratamiento. Los datos se presentan en una tabla de contingencia 2x2, como se muestra a continuación:

Tabla 6.4

Tabla de asociación o 2 x 2

	Post tratamientos	Pretratamientos
Menos episodios	20	30
Igual o más episodios	10	40

Nota. Elaboración propia.

En este caso, el objetivo es determinar si el tratamiento es efectivo en la reducción de los episodios de migraña en los pacientes.

Para aplicar la prueba de McNemar, se calculan los valores de b y c , que representan los casos en los que el evento de interés (reducción en los episodios de migraña) ocurre en un grupo, pero no en el otro. En este ejemplo, b es igual a 20 y c es igual a 10. Por lo tanto, la fórmula para el estadístico de McNemar es:

$$M = \frac{[(b-c)^2]}{b+c} = \frac{[(20-10)^2]}{20+10} = 3,33$$

Para interpretar el valor del estadístico de prueba, se utiliza una tabla de distribución chi-cuadrado con un grado de libertad. En este caso, el valor crítico de chi-cuadrado para un nivel de significancia del 0.05 es 3.84. Como el valor de M (3.33) es menor que el valor crítico, no se puede rechazar la hipótesis nula de que no hay diferencia significativa en la proporción de pacientes que experimentan una reducción en los episodios de migraña antes y después del tratamiento.

Por lo tanto, no se puede concluir que el nuevo tratamiento sea efectivo en la reducción de los episodios de migraña en los pacientes, al menos no con un nivel de significancia del 0.05. Es importante tener en cuenta que se deben considerar otros factores y limitaciones del estudio antes de sacar conclusiones definitivas sobre la efectividad del tratamiento.

En conclusión, el estadístico de McNemar es una herramienta útil para evaluar la diferencia en la proporción de ocurrencia de un evento en dos grupos relacionados. Su uso es adecuado en estudios de casos y controles, y su interpretación requiere conocer tanto el valor del estadístico de prueba como el valor p asociado.

En los siguientes videos se ponen en práctica los temas tratados en esta sección:

Supuestos:

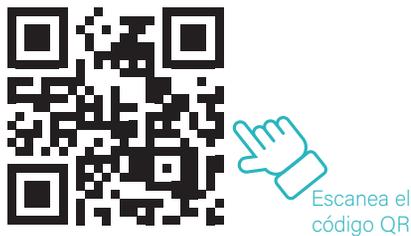


Escanea el código QR

Ejemplo:



Test de McNemar-Bowker y test de Bhapkar para variable politómica:



Prueba de Kruskal-Wallis

El test estadístico de Kruskal-Wallis es una prueba no paramétrica utilizada para comparar tres o más grupos independientes. Se utiliza para determinar si las medianas de las poblaciones de los grupos son iguales o diferentes entre sí.

La prueba de Kruskal-Wallis se basa en la clasificación de los datos de todas las muestras juntas, y luego se calcula la suma de los rangos de cada grupo. El estadístico de prueba se calcula utilizando esta suma de rangos y se compara con una distribución chi-cuadrado para determinar si hay una diferencia significativa entre los grupos.

$$H = \left[\frac{12}{N(n-1)} \sum \frac{R_j^2}{n} \right] - 3(n+1)$$

donde:

H es el estadístico de prueba de Kruskal-Wallis.

N es el número total de observaciones en todos los grupos juntos.

$\sum (R_j - (n+1)/2)^2$ es la suma de los cuadrados de las desviaciones de los rangos promedio de cada grupo j, donde R_j es la suma de los rangos de las observaciones en el grupo j.

12 es una constante que se utiliza en la fórmula.

$3(n+1)$ es un término de corrección que se utiliza para ajustar el estadístico de prueba para muestras pequeñas.

Es importante destacar que este test no requiere que los datos se distribuyan de manera normal, lo que lo hace útil para comparar grupos con distribuciones no normales o con tamaños de muestra pequeños.

La hipótesis nula de la prueba de Kruskal-Wallis se construye bajo la premisa de la no diferencia significativa entre las medianas de los grupos, mientras que la hipótesis alternativa es que al menos una mediana es diferente de las demás.

Si la prueba arroja un valor de p menor que el nivel de significancia establecido (por lo general, 0.05), entonces se rechaza la hipótesis nula y se concluye que hay una diferencia significativa entre las medianas de los grupos.

Es importante destacar que la prueba de Kruskal-Wallis solo indica que hay una diferencia significativa entre los grupos, pero no indica dónde se encuentra la diferencia. En caso de encontrar una diferencia significativa, se puede realizar una prueba de comparaciones múltiples para determinar qué grupos difieren significativamente entre sí.

Ejemplo 6.3

Supongamos que un investigador desea comparar la efectividad de tres tratamientos diferentes para reducir la ansiedad en pacientes con trastornos de ansiedad. Para ello, se reclutaron 15 pacientes con trastornos de ansiedad y se asignaron aleatoriamente a uno de los tres tratamientos diferentes. Después de una semana de tratamiento, se midió el nivel de ansiedad de cada paciente utilizando una escala de calificación de ansiedad.

Los datos obtenidos se presentan a continuación:

Tratamiento 1: 5 6 7 8 5

Tratamiento 2: 4 4 2 3 5

Tratamiento 3: 6 7 8 9 6

Para realizar el test de Kruskal-Wallis, primero debemos ordenar los datos de menor a mayor y asignar los rangos correspondientes a cada observación. Los rangos se asignan de manera que el valor más bajo tenga un rango de 1 y los rangos se asignan en orden ascendente. Si hay empates, se promedian los rangos.

Los datos ordenados y los rangos asignados se presentan a continuación:

2 3 4 4 5 5 5 6 6 7 7 8 8 9 9

1 2 3 4 6 7 8 9 10 11 12 13.5 13.5 15 15

Luego, calculamos la suma de los rangos para cada grupo:

Suma de rangos para Tratamiento 1: 31

Suma de rangos para Tratamiento 2: 18

Suma de rangos para Tratamiento 3: 46

A continuación, utilizamos la fórmula para calcular el estadístico de prueba H:

$$H = \left[\frac{12}{N(n-1)} \sum \frac{R^2}{n} \right] - 3(n+1) = \frac{12}{15(15-1)} * 49,63 - 3(15+1) = 6,617$$

El valor de H se compara con una distribución chi-cuadrado con k-1 grados de libertad, donde k es el número de grupos en el análisis. En este caso, k = 3 y los grados de libertad son 2.

Si se utiliza un nivel de significancia de 0.05, el valor crítico de la distribución chi-cuadrado con 2 grados de libertad es 5.99.

Como H (6.617) es mayor que el valor crítico de la distribución chi-cuadrado (5.99), se rechaza la hipótesis nula y se concluye que hay una diferencia significativa entre las medianas de los tres tratamientos.

A continuación, se exhiben dos videos: uno que conceptualiza los temas previamente abordados y otro que los aplica en la práctica mediante un ejercicio específico relacionado con el estadístico Kruskal-Wallis.

Video 1:



Video 2:



El estadístico de Jonckheere-Terpstra es una prueba no paramétrica que se utiliza para detectar si hay una tendencia significativa en un conjunto de datos ordenados. Esta prueba es útil cuando los datos no cumplen con los supuestos de normalidad y homogeneidad de varianza que requieren otras pruebas estadísticas paramétricas como ANOVA.

La prueba de Jonckheere-Terpstra es una prueba unilateral, lo que significa que solo detecta tendencias que aumentan o disminuyen. La hipótesis nula establece que no hay tendencia en los datos ordenados y la hipótesis alternativa establece que hay una tendencia significativa.

Para calcular la estadística de prueba J, primero se deben asignar los rangos a cada observación de acuerdo con su posición en la secuencia ordenada de los datos. Luego, se determinan los pesos de acuerdo con la hipótesis nula. Los pesos se utilizan para dar más peso a las diferencias de rango entre los grupos cercanos y menos peso a las diferencias de rango entre grupos más lejanos.

La estadística de prueba J, se calcula como la suma de los valores de un vector de pesos multiplicados por las diferencias de los rangos correspondientes de los grupos. La distribución de probabilidad de J depende del tamaño de la muestra y del número de grupos.

Una vez que se ha calculado la estadística de prueba J, se compara con los valores críticos de la distribución de probabilidad de J para determinar si hay una tendencia significativa. Si el valor de J es mayor que el valor crítico, se rechaza la hipótesis nula y se concluye que hay una tendencia significativa.

Finalmente, se calcula el valor p a partir de la distribución de probabilidad de J. Si el valor p es menor que el nivel de significancia establecido, se concluye que hay una tendencia significativa.

Ejemplo 6.4

Supongamos que un investigador desea determinar si hay una tendencia significativa en la concentración de contaminantes en el agua de un río en diferentes puntos de muestreo. El investigador ha recolectado datos de 4 puntos de muestreo diferentes y ha registrado la concentración de contaminantes en cada punto de muestreo. Los datos se muestran en la siguiente tabla:

Tabla 6.5

Descripción de la muestra

Punto de muestreo	Concentración de contaminantes
B	9
A	5
D	18
C	12

Nota. Elaboración propia.

Para realizar la prueba de Jonckheere-Terpstra, primero se deben ordenar los datos por concentración de contaminantes:

Tabla 6.6

Descripción de la muestra ordenada

Punto de muestreo	Concentración de contaminantes (ordenados)
A	5
B	9
C	12
D	18

Nota. Elaboración propia.

A continuación, se asignan rangos a los datos según su posición en la secuencia ordenada. En caso de empate, se les asigna el promedio de los rangos que les corresponden. Por lo tanto, los rangos asignados serían:

Tabla 6.7

Descripción de la muestra con rangos

Punto de muestreo	Concentración de contaminantes (ordenados)	Rangos
A	5	1
B	9	2

Punto de muestreo	Concentración de contaminantes (ordenados)	Rangos
C	12	3
D	18	4

Nota. Elaboración propia.

A continuación, se determinan los pesos de acuerdo con la hipótesis nula de que no hay tendencia en los datos. Los pesos se utilizan para dar más peso a las diferencias de rango entre los grupos cercanos y menos peso a las diferencias de rango entre grupos más lejanos. En este caso, los pesos serían:

Tabla 6.8

Diferencia de rangos

Puntos de muestreo	Pesos
A y B	-1
B y C	-1
C y D	-1

Nota. Elaboración propia.

La estadística de prueba J se calcula como la suma de los valores de un vector de pesos multiplicados por las diferencias de los rangos correspondientes de los grupos. En este caso, la estadística de prueba J se calcularía como:

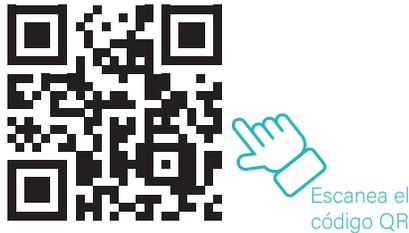
$$J = (-1)(2-1) + (-1)(3-2) + (-1)(4-3) = -3$$

Para comparar la estadística de prueba J, con los valores críticos de la distribución de probabilidad de J, se necesitan el tamaño de la muestra y el número de grupos. En este caso, el tamaño de la muestra es 4 y el número de grupos es 3 (ya que hay 3 diferencias de rango).

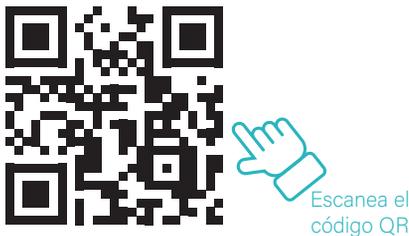
Si se utiliza un nivel de significancia del 5%, el valor crítico para J es 1. Para este ejemplo, como J es menor que el valor crítico, no se rechaza la hipótesis nula de que no hay tendencia en los datos ordenados. Por lo tanto, se concluye que no hay una tendencia significativa en la concentración de contaminantes en diferentes puntos de muestreo en el río.

En los siguientes videos se conceptualizan los temas previamente abordados y se ponen en práctica mediante un ejercicio específico en SPSS

Video 1:



Video 2:



Estadístico U de Mann-Whitney

El estadístico U de Mann-Whitney es una herramienta estadística no paramétrica utilizada para comparar la mediana de dos muestras independientes. Se utiliza para determinar si existe una diferencia significativa entre dos grupos de datos.

A diferencia de las pruebas paramétricas, como la prueba t de Student, la prueba U de Mann-Whitney es una prueba no paramétrica, lo que significa que no requiere la suposición de una distribución normal de los datos subyacentes. Sin embargo, aún existen algunos supuestos y consideraciones importantes para aplicar correctamente la prueba U de Mann-Whitney:

- Independencia de las muestras: las muestras comparadas deben ser independientes, es decir, que los valores en una muestra no deben estar relacionados o influir en los valores de la otra muestra.
- Homogeneidad de la varianza: aunque la prueba U de Mann-Whitney es robusta ante la presencia de valores atípicos, aún se debe tener cuidado para garantizar que las varianzas de ambas muestras sean similares.
- Escala ordinal de los datos: la prueba U de Mann-Whitney se utiliza comúnmente para datos ordinales o datos nominales convertidos en ordenados. Si los datos son de escala nominal o categóricos, se puede utilizar una prueba como la prueba exacta de Fisher.
- Muestra grande: para muestras pequeñas (por ejemplo, menos de 10 observaciones en cada muestra), el poder de la prueba puede ser limitado y se recomienda el uso de pruebas exactas.

Este estadístico se basa en la clasificación de los datos de ambas muestras en un solo conjunto y la asignación de rangos a cada uno de los valores. Después de asignar los rangos, se calcula la suma de los rangos para cada grupo. A continuación, se compara la suma de los rangos para determinar si existe una diferencia significativa entre las dos muestras. Su fórmula es:

$$U_1 = n_1 * n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 * n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

donde n_1 y n_2 son los tamaños de las dos muestras, y R_1 es la suma de los rangos de la muestra más pequeña.

La hipótesis nula del estadístico U de Mann-Whitney es que no existe diferencia significativa entre las medianas de las dos muestras. Si el valor calculado del estadístico U es menor que el valor crítico de la tabla de distribución de Mann-Whitney, se rechaza la hipótesis nula y se concluye que existe una diferencia significativa entre las medianas de las dos muestras.

Ejemplo 6.5

Supongamos que queremos comparar la eficacia de dos programas de entrenamiento en un grupo de atletas. Tenemos dos grupos de atletas, uno que ha seguido el programa A y otro que ha seguido el programa B. Queremos determinar si hay alguna diferencia significativa en el rendimiento de los atletas después de completar cada programa.

Supongamos que tenemos las siguientes calificaciones de cada atleta después de completar su respectivo programa:

Grupo A: 5, 8, 9, 10, 12

Grupo B: 3, 6, 7, 11, 13, 15

Primero, ordenamos todas las calificaciones de ambas muestras de menor a mayor:

Grupo A: 5, 8, 9, 10, 12

Grupo B: 3, 6, 7, 11, 13, 15

A continuación, asignamos un rango a cada una de las calificaciones. El atleta con la calificación más baja recibe el rango 1 y el atleta con la calificación más alta recibe el rango 6 (en el caso del grupo B, con 6 atletas, el atleta con la calificación más baja recibe el rango 1 y el atleta con la calificación más alta recibe el rango 6).

Grupo A: 1, 2, 3, 4, 5

Grupo B: 1, 2, 3, 4, 5, 6

A continuación, calculamos la suma de los rangos de cada grupo:

$$\text{Grupo A: } 1+2+3+4+5 = 15$$

$$\text{Grupo B: } 1+2+3+4+5+6 = 21$$

Luego, calculamos el valor de U utilizando la fórmula mencionada anteriormente:

$$U = n_1 \times n_2 + (n_1 \times (n_1 + 1)) / 2 - \text{suma de los rangos del grupo 1}$$

Donde n_1 es el tamaño de la muestra del grupo 1 (en este caso, 5) y n_2 es el tamaño de la muestra del grupo 2 (en este caso, 6).

$$U = (5 \times 6) + (5 \times (5 + 1)) / 2 - 15$$

$$U = 30 + 15 - 15$$

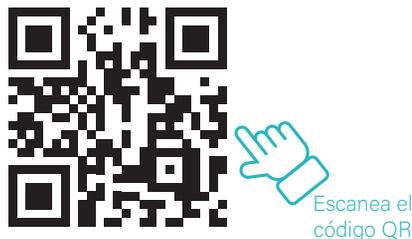
$$U = 30$$

Finalmente, comparamos el valor de U con el valor crítico de U correspondiente a un nivel de significancia de 0.05 (según la tabla de valores críticos de U). Si el valor de U es mayor que el valor crítico de U, entonces podemos concluir que hay una diferencia significativa entre los dos grupos en términos de rendimiento.

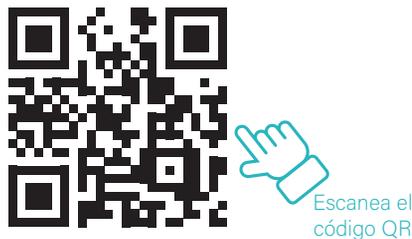
En este caso, el valor crítico de U es 9 (para $n_1 = 5$ y $n_2 = 6$). Como el valor de U (30) es mayor que el valor crítico de U (9), podemos rechazar la hipótesis nula de que las medianas de los grupos son iguales y concluir que hay una diferencia significativa entre los dos programas de entrenamiento en términos de rendimiento de los atletas.

En los siguientes videos se encontrará información más detallada sobre el estadístico anteriormente visto.

Video 1:



Video 2:



Test de Friedman

El test estadístico de Friedman es una técnica utilizada para comparar tres o más grupos en una variable dependiente medida en una escala ordinal. El objetivo de este test es determinar si hay diferencias significativas entre los grupos en la variable dependiente.

El proceso del test de Friedman implica el siguiente procedimiento: primero se clasifican los datos en grupos y se ordenan de menor a mayor en cada grupo. Luego se calcula la suma de los rangos para cada grupo y se obtiene un estadístico de prueba conocido como Chi-cuadrado de Friedman. Este se calcula de la siguiente manera:

$$X^2 = \left[\frac{12}{nk(n+1)} \sum R_j^2 \right] - 3n(n+1)$$

donde:

χ^2 : el valor del estadístico de prueba Chi-cuadrado de Friedman.

k: el número de tratamientos (en este caso, el número de grupos o equipos).

n: el número de observaciones (en este caso, el número de eventos).

R_j : la suma de los rangos de los datos de cada tratamiento (en este caso, cada equipo) en el evento j.

La fórmula implica ordenar los datos para cada tratamiento (equipo) por evento y calcular los rangos correspondientes. Luego se suman los rangos para cada tratamiento en cada evento y se elevan al cuadrado. La suma total de los cuadrados de los rangos para cada tratamiento se divide por 12 veces el número de tratamientos, multiplicado por el número de observaciones al cuadrado menos 1. A este valor se le resta 3 veces el número de tratamientos por el número de observaciones al cuadrado más el número de observaciones.

Este estadístico se compara con una distribución Chi-cuadrado con grados de libertad igual al número de grupos menos uno. Si el valor obtenido del estadístico es mayor que el valor crítico de la distribución Chi-cuadrado, entonces se rechaza la hipótesis nula de que no hay diferencias significativas entre los grupos.

El test de Friedman es útil en situaciones en las que se desea comparar tres o más grupos en una variable ordinal y no se pueden asumir supuestos de normalidad o igualdad de varianzas. Es una alternativa no paramétrica al ANOVA de un factor.

Es importante tener en cuenta que el test de Friedman no identifica qué grupo o grupos difieren significativamente. En caso de que se encuentre una diferencia significativa, se pueden realizar pruebas de comparación post hoc para determinar dónde se encuentra la diferencia significativa.

Ejemplo 6.6

Imaginemos que se desea comparar el desempeño de tres equipos deportivos en una competencia. Cada equipo participó en 5 eventos diferentes, y se registró su posición final en cada evento (1º, 2º, 3º, etc.) para crear una escala ordinal de 1 a 5.

Los datos se presentan en la siguiente tabla:

Tabla. 6.10

Descripción de eventos

Evento	Equipo 1	Equipo 2	Equipo 3
1	2	4	1
2	1	5	3
3	3	2	1
4	5	3	2
5	4	1	3

Nota. Elaboración propia.

El primer paso es ordenar los datos de cada equipo de menor a mayor posición final, creando una escala ordinal de 1 a 5. Los datos ordenados se muestran en la siguiente tabla:

Tabla. 6.11

Descripción de eventos ordenado

Evento	Equipo 1	Equipo 2	Equipo 3
1	2	1	1
3	3	2	2
2	1	3	3
5	4	1	3
4	5	3	4

Nota. Elaboración propia.

A continuación, se calculan los rangos de cada equipo para cada evento. El rango es simplemente el número de posición final de cada equipo dentro de su grupo. Si hay empates, se asigna el promedio de los rangos correspondientes. Los rangos se muestran en la siguiente tabla:

Tabla. 6.12

Descripción de eventos con rangos

Evento	Equipo 1	Equipo 2	Equipo 3
1	2	1	1
3	2	1.5	1.5
2	1.5	2	2
5	3	1.5	3
4	3	2	4

Nota. Elaboración propia.

Luego se suman los rangos de cada equipo para obtener una suma total de rangos para cada equipo. En este caso, la suma de rangos para cada equipo es:

Equipo 1: 12.5

Equipo 2: 8

Equipo 3: 12

Finalmente, se utiliza el estadístico de prueba Chi-cuadrado de Friedman para determinar si hay diferencias significativas entre los grupos. Para este ejemplo, el valor del estadístico de prueba es 3.6, y el valor crítico de la distribución Chi-cuadrado con dos grados de libertad al nivel de significancia del 5% es 5.99. Como el valor obtenido del estadístico de prueba es menor que el valor crítico, no se rechaza la hipótesis nula de que no hay diferencias significativas entre los grupos.

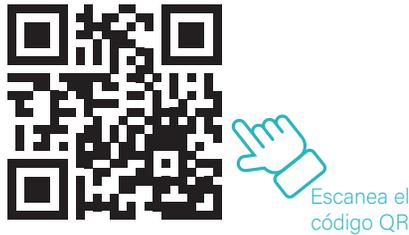
Para una mayor comprensión del estadístico y una manera sencilla del cálculo mediante el software SPSS, se recomienda ver los siguientes videos

Video 1:



Escanea el código QR

Video 2:



Test de Q de Cochran

La prueba de la Q de Cochran es una herramienta estadística que permite verificar la igualdad de varias muestras relacionadas en una variable dicotómica. Se trata de una prueba que puede ser considerada como una extensión del test de McNemar, ya que permite comparar más de dos poblaciones.

El objetivo de esta prueba es contrastar la hipótesis nula de igualdad de proporciones entre las muestras. Para ello, se utiliza un estadístico de test específico que se calcula a partir de la varianza observada entre las muestras y la varianza esperada bajo la hipótesis nula.

Es importante tener en cuenta que la prueba de la Q de Cochran solo es adecuada para datos categóricos binarios y para muestras relacionadas. Además, en caso de obtener una diferencia significativa entre las muestras, es necesario llevar a cabo un análisis posterior para determinar cuál de ellas presenta una proporción significativamente diferente al resto.

Este se calcula de la siguiente manera:

$$Q = \frac{k(k-1)\sum(g_i - \bar{g})^2}{\sum L_i - \sum L_i^2}$$

Donde:

K= número de tratamientos

N= tamaño muestra

g_i = número de éxito en el tratamiento i ésimo

\bar{g} = promedio de éxito por tratamiento

L_i = numero de éxito del individuo i ésimo

En resumen, la prueba de la Q de Cochran es una herramienta útil para comparar la igualdad de varias muestras relacionadas en una variable dicotómica. En los siguientes links encontrarán con mayor profundidad más información sobre el estadístico visto anteriormente.

Video 1:



Escanea el código QR

Video 2:



Escanea el código QR

Video 3:



Escanea el código QR

CAPÍTULO 7. REVISIÓN SISTEMÁTICA Y METAANÁLISIS



Revisión sistemática

Siempre que se habla de metaanálisis es imposible descartar el componente de la revisión sistemática, ya que es en este proceso en el cual se obtienen los datos que posteriormente serán tratados a través de programas estadísticos. Las revisiones sistemáticas son un tipo de estudio de análisis secundario de la información que propende por el establecimiento de protocolos de revisión del estado del arte con el objetivo de que sean replicables y comprobables. Al promover procesos estructurados de revisión evitan la posibilidad de sesgo por parte de los investigadores que puedan alterar los resultados de las búsquedas.

Ejemplo 7.1

Dos investigadores, uno de Latinoamérica y otro de Europa, quieren revisar qué existe en la literatura sobre el ejercicio físico y sus efectos sobre la cognición en población adulta mayor. De no existir un método eficaz de sistematización los resultados de cada investigador serían diferentes, seguramente sesgados por idioma y experiencia en el campo. Sin embargo, es posible que ambos encuentren los mismos resultados si se respetan los pasos planteados en las revisiones sistemáticas.

Es de esta forma que surgen instituciones que establecen los pasos de las revisiones sistemáticas como son PRISMA y COCHRANE. Para efectos prácticos de este capítulo, solo nos enfocaremos en la guía PRISMA, particularmente en el proceso de búsqueda, extracción de la información y, claro está, el metaanálisis.

Proceso de búsqueda de artículos

Si partimos con el objetivo de realizar un metaanálisis es imprescindible que nuestra búsqueda se adecúe a ello. Durante las revisiones sistemáticas, cuando se efectúa una búsqueda existen diferentes métodos para establecer la cadena de búsqueda que sería nuestra forma de comunicación con las bases de datos. Uno de estos métodos o estrategias es la PICO que, por sus siglas en inglés, significa (P)oblación, (I)ntervención, (C)omparación y (O)Variable. Cada una de estas siglas hace referencia a un apartado importante dentro de un artículo, que definiremos a partir de este momento como categorías.

En cada Categoría, es necesario combinar palabras clave agrupadas con operadores Booleanos (AND, OR y NOT), organizadas por comillas y paréntesis. Esta estructura permite enfocar la cadena de búsqueda a encontrar lo que realmente necesitamos, pero sin dejar de lado artículos que puedan llegar a ser importantes por culpa de algún sesgo o decisión del investigador.

Ejemplo 7.2

Retomando el ejemplo anterior, donde el objetivo es determinar los efectos de la actividad física sobre la cognición en adultos mayores aparentemente sanos, lo que se buscaría es una cadena similar a esta: P: (“older adults” OR “Elderly” OR “Elder”) AND I:(“exercise” OR “Physical exercise” OR “Physical Activity”) AND C:(“Control Group” OR “Placebo”) AND O:(“Cognition” OR “Global Cognition”). Donde el AND se utiliza para unir las categorías que están agrupadas por el paréntesis y formadas por las palabras clave entre comillas y unidas por OR.

Una vez la cadena de búsqueda es puesta a prueba, se puede empezar con la selección de los estudios basados en los criterios de inclusión que planteen los investigadores, por posteriormente iniciar la extracción de la información.

Extracción de la información

Este proceso es crucial para el desarrollo de un metaanálisis, ya que busca extraer de los estudios incluidos la información que posteriormente sufrirá el análisis secundario. Para esto existen muchos métodos, matrices y formas; sin embargo, no hay una que sea mejor que otra. Todo depende de la necesidad de los investigadores, sus criterios de inclusión y objetivos, junto con su habilidad y experiencia revisando artículos científicos, ya que para este proceso se requiere la lectura del artículo a texto completo.

Ejemplo 7.3

Pensemos que después de correr la cadena de búsqueda previamente diseñada encontramos 600 artículos de los cuales incluimos 5, con esto en mente, procederemos a extraer la información. Para dar respuesta a nuestro objetivo, tendremos que extraer información sobre autores y años de publicación, tamaño de la muestra, tipo de actividad física efectuada y los resultados obtenidos, en términos de instrumento empleado y datos observados. Por lo que nuestra tabla quedaría de la siguiente manera:

Tabla 7.1

Extracción de la información para metaanálisis

Autores y Año	Muestra	Intervención	Instrumento	Medición
Autor 1, 2020	GC=30 EG=39	Actividad física aeróbica de alta intensidad	MMSE	EG T0=21,4 (3,94) T1=21,0 (4,38) T2= 20,4 (4,77)

Autores y Año	Muestra	Intervención	Instrumento	Medición
Autor 2, 2021	GC= 17 EG= 16	Ejercicio físico	MMSE	T0=27.2 (1.9) T1=27.5 (1.9) T2=28.3 (1.3) T3=28.3 (1.5) T4=28.2 (1.5) T5=27.6 (1.6)
Autor 3, 2014	GC= 27 GI= 27	Ejercicio de resistencia progresivo	ADAS-Cog	T0 = 8.02 (6.87, 9.17) T1 = 6.26 (5.11,7.41)* T2 = 5.76 (4.59. 6.92)
Autor 4, 2019	GC= 15 GI= 16	Entrenamiento orientado a la tarea	MMSE	T0= 16.88 T1= 3.56* T2= 3.75d*
Autor 5, 2018	GC= 165 GI= 329	Entrenamiento físico de moderado a intenso	MMSE	MMSE T0= 17,0 (6,1) T2= 19,8 (8.4)

Nota. Elaboración propia. GC: Grupo Control; GI: Grupo Intervención, T: Medida realizada; MMSE: Mini-Mental State Examination. ADAS-Cog: Alzheimer disease assessment scale-cognitive.

Metaanálisis

Después de todo este preámbulo es posible abordar el metaanálisis. Una vez realizados todos los procesos previos de forma consciente, organizada y sistematizada, podemos pasar al programa estadístico. En este caso, hablaremos del software Comprehensive Meta-analysis V4, el cual nos permite trabajar con diferentes estadísticos. Sin embargo, es importante aclarar que, si bien el software nos permitirá incluir los datos de diferentes artículos, el éxito del metaanálisis depende del criterio del investigador.

Para que el metaanálisis sea adecuado es importante comprender que no podemos comparar los resultados obtenidos de diferentes artículos si estos emplean métodos distintos de medición. Es decir que no toda revisión sistemática es candidata a metaanálisis, siendo una de las principales limitaciones la heterogeneidad que muchos artículos suelen presentar, tanto en instrumentos como también de los resultados.

Es por esto que, el primer paso para determinar la viabilidad del metaanálisis es la realización de un Funnel Plot, en el que es posible observar la distribución de los resultados

según los tamaños de las muestras, esperando observar una forma de pirámide que de no cumplirse se descarta la opción de realizar un metaanálisis.

Ejemplo 7.4

Una vez extraída la información, es posible evaluar la heterogeneidad de los resultados. Sin embargo, aunque todos los artículos incluidos cumplieran con los criterios de selección planteados, hubo uno que midió la cognición con un instrumento distinto (Autor 3), por lo que no sería posible incluirlo en el metaanálisis. Con los cuatro restantes se realizaría el Funnel Plot a través del software Comprehensive Meta-Analysis, teniendo por resultado algo como esto:

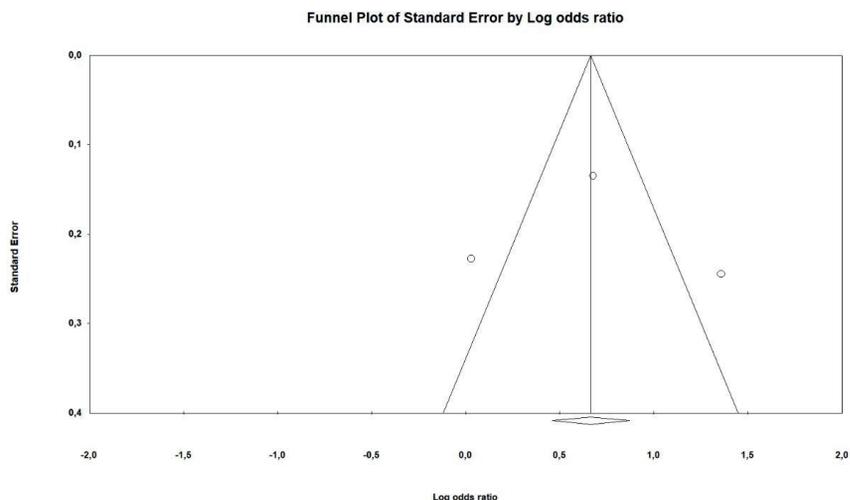


Figura 7.1

Funnel Plot con heterogeneidad baja

Nota. Elaboración propia.

Cuando el Funnel Plot se muestra como el de la Figura 7.1 se entiende que el índice de heterogeneidad es bajo y el metaanálisis es viable, esto se hace observando la distribución de los artículos y estableciendo si se encuentran dentro del rango de la pirámide. Mientras que, cuando la heterogeneidad es alta, los puntos que representan a cada estudio se saldrían de la pirámide, como es el caso de la Figura 7.2.

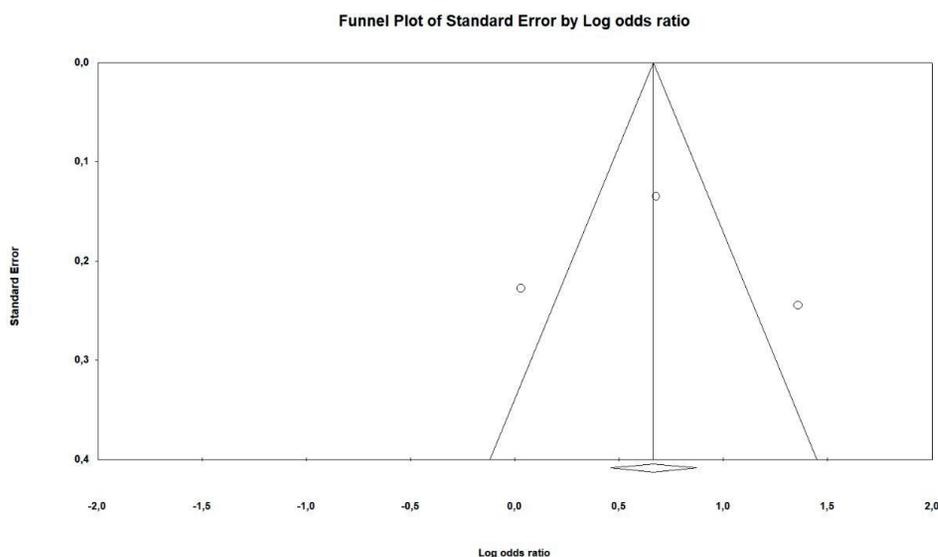


Figura 7.2

Funnel Plot con heterogeneidad alta

Nota. Elaboración propia.

A propósito de estos temas, ¿qué es el riesgo de sesgo de publicación? Pues bien, en la cultura actual de la investigación es común encontrarse con revistas y autores que optan por publicar únicamente cuando las intervenciones presentan resultados favorables sobre las variables medidas. Es por esto que el estado del arte se llena de artículos que afirman la efectividad de ciertas intervenciones que cuando se replican en entornos reales no alcanzan los resultados esperados. Esto se debe a que no siempre los editores o investigadores consideran que el no evidenciar cambios en las variables posterior a una intervención u observar efectos negativos sea importante, sesgando el estado del arte. A este sesgo se le denomina sesgo de publicación.

Después de evaluado el sesgo de publicación, se procede entonces a planear el metaanálisis, aquí es importante establecer los análisis por subgrupos que se van a realizar. Entiéndase por análisis de subgrupos las divisiones que se realizarán de los datos seleccionados, siendo uno de los más comunes el análisis por sexo, pero también es posible hacerlo a partir de la calidad metodológica de los estudios incluidos, el tipo de intervención, el tipo de control (placebo, no intervención), la dosificación, entre otros. Este tipo de análisis nuevamente depende de la necesidad de los investigadores.

Después de esto se realizará el metaanálisis, escogiendo la ruta adecuada según sean los datos extraídos, ya sean medidas de tendencia central y dispersión (Medía + Desviación Estándar, Mediana + Rango Intercuartílico, Media + Intervalo de Confianza) o medidas de

probabilidad (OR, RR, HR). Es importante aclarar que no se pueden combinar estos dos grandes grupos, pero sí los estadísticos asociados entre ellos, es decir, pueden tener estudios que reporten media + desviación estándar y otros con mediana + rango intercuartílico, pero no analizarlos junto con estudios que reporten el OR. Tampoco sería adecuado realizar un metaanálisis donde se comparen los datos de estudios con naturalezas distintas, es decir, no se pueden comparar estudios observacionales y experimentales, aunque sí es posible analizar juntos estudios transversales y longitudinales, ya que comparten la misma naturaleza.

Siguiendo con los posibles análisis, también es posible unificar diferentes instrumentos de medición siempre y cuando se enfoquen al mismo resultado; por ejemplo, la cognición global puede ser medida con diferentes instrumentos como es el caso del Mini Mental State Examination, el Alzheimer's Disease Assessment Scale-Cognitive Subscale o el Montreal Cognitive Assessment, cada uno con puntuaciones distintas y formas de evaluar distintas, pero que son posibles de unificar a través de tamaños de efecto, los cuales ya han sido abordados en capítulos previos.

Los resultados obtenidos del metaanálisis serán de dos tipos principalmente, cuando lo que se incluyó fueron medidas de tendencia central y dispersión, el metaanálisis arrojará el tamaño del efecto medio, siendo posible calcular a través de diferentes estadísticos como es el caso del Cohen's d o el Hedges' g . En ambos casos, cuanto más cercano esté el valor obtenido de 1, más grande será el tamaño del efecto medio. Por otro lado, cuando se analizan medidas de probabilidad, lo que obtendremos será la misma medida unificada con lo que ella represente. Adicionalmente, el software arrojará un valor de p que permitirá identificar la significancia estadística de los resultados. Adicionalmente, Estos resultados pueden ser representados gráficamente a través de un Forest Plot, siendo este de gran utilidad para la interpretación rápida de los resultados.

Ejemplo 7.5

Por último, después de verificar el riesgo de sesgo de publicación de los datos sería posible correr el metaanálisis. Como los estudios incluidos eran de carácter experimental, lo que buscamos calcular es un tamaño medio del efecto. Una vez tabulados los datos en el software estadístico y corrido el metaanálisis, nos quedaría algo así:

Tabla 7.2

Tamaño del efecto de cada estudio y tamaño del efecto medio

	Estadística de cada estudio			
	Hedges's g	Límit Inf	Límit Sup	p-Valor
Autor 1, 2020	-0.334	-0.521	-0.147	<0.001
Autor 2, 2021	-0.996	-1.251	-0.741	<0.001
Autor 3, 2014	-0.777	-1.049	-0.505	<0.001
Autor 4, 2019	-0.996	-1.251	-0.741	<0.001
Autor 5, 2018	-1.744	-2.237	-1.251	<0.001
Media	-0.775	-0.887	-0.663	<0.001

Nota. Elaboración propia.

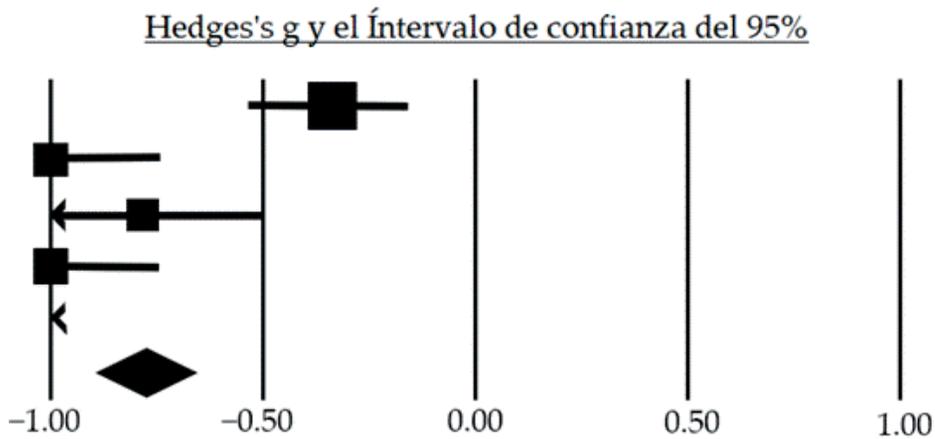


Figura 7.3

Forest Plot resultado del metaanálisis

Nota. Elaboración propia.

En este punto podría darse por finalizado el metaanálisis, sin embargo, es posible seguir analizando los datos si se desea realizar más subgrupos.

Hasta este momento hemos aprendido cómo se realiza el metaanálisis; sin embargo, el software nos arroja dos modelos distintos que debemos comprender para escoger el que más se adecúe a nuestras necesidades. Estos modelos llevan el nombre de modelo ajustado (*Fixed model*) y modelo aleatorio (*Random model*). Para entender la diferencia entre uno y otro es importante conocer el rol que juega la heterogeneidad y la variabilidad dentro de un metaanálisis. La heterogeneidad, por su parte, se calcula a través del I², estadístico que es presentado a través de un porcentaje y que indica qué tan heterogéneos son los datos que fueron analizados. Por otro lado, la variabilidad se calcula a través del valor Q de Cochran, el cual reporta el porcentaje de variabilidad asociado a la heterogeneidad de los datos.

Para tomar una decisión sobre cuál modelo usar, nos fijaremos principalmente en el valor del I² tomando como referencia los siguientes valores: 0% al 40%: sin importancia; 30% al 60%: podría representar una heterogeneidad moderada; 50% al 90%: podría representar una heterogeneidad importante; 75% al 100%: alto grado de heterogeneidad. Cuando el valor del I² supera el 50% se recomienda usar el modelo ajustado, ya que en este se realiza una normalización de los datos empleados y se asume una heterogeneidad del 0%, pero si el valor del I² es $\leq 50\%$ el modelo aleatorio es el que debemos usar.

Para dar claridad a lo visto anteriormente se invita a que revisen el siguiente video en YouTube.



Escanea el
código QR

REFERENCIAS



- Agresti, A. (2018). *An Introduction to Categorical Data Analysis*. Wiley.
- Agresti, A., & Finlay, B. (2018). *Statistical Methods for the Social Sciences*. Pearson.
- Aguinis, H. (2004). *Regression Analysis for Categorical Moderators*. Guilford Press.
- Argyrous, G. (2008). *Statistics for Research: With a Guide to SPSS*. SAGE Publications.
- Bausell, R. B. (2015). *Beyond Null Hypothesis Significance Testing: Reconceptualizing Statistical Inference*. SAGE Publications.
- Bluman, A. G. (2017). *Elementary Statistics: A Step-by-Step Approach*. McGraw-Hill Education.
- Briega, R. E. L. (2016, 29 junio). Distribuciones de probabilidad con Python. <https://relopezbriega.github.io/blog/2016/06/29/distribuciones-de-probabilidad-con-python/>
- Borenstein, M. (2022). *Comprehensive meta-analysis software. Systematic reviews in health research: meta-analysis in context* (pp. 535-548). Wiley Online Books.
- Bulmer, M. G. (1979). *Principles of Statistics*. Dover Publications.
- Carcelén-Fraile, M. D. C., Lorenzo-Nocino, M. F., Afanador-Restrepo, D. F., Rodríguez-López, C., Aibar-Almazán, A., Hita-Contreras, F., . . . Castellote-Caballero, Y. (2022). Effects of different intervention combined with resistance training on musculoskeletal health in older male adults with sarcopenia: A systematic review. *Front Public Health*, (10), 1037464. doi:10.3389/fpubh.2022.1037464
- Christensen, L. B., Johnson, R. B., & Turner, L. A. (2019). *Research Methods, Design and Analysis*. Pearson.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge.
- Cohen, J. (1994). The Earth Is Round ($p < .05$). *American Psychologist*, 49(12), 997-1003.
- Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. Routledge.
- Conover, W. J. (1999). *Practical nonparametric statistics* (3rd ed.). Wiley.
- Córdova Zamora, M. (2005). *Introducción a la Inferencia Estadística*. Ecoe Ediciones.
- Cumming, G. (2012). *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. Routledge.
- Cumming, G. (2014). Hypothesis Testing in Statistics: A Reconceptualization of Statistical Inference. *Psychological Science*, 25(1), 7-29.
- Dalgaard, P. (2008). *Introductory Statistics with R*. Springer.
- Diez, D. M., Barr, C. D., & Çetinkaya-Rundel, M. (2015). *Introductory Statistics with Randomization and Simulation*. OpenIntro.
- Diez, D. M., Barr, C. D., & Çetinkaya-Rundel, M. (2017). *OpenIntro Statistics*. OpenIntro.

- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. SAGE Publications.
- Field, A. (2017). *Discovering Statistics Using IBM SPSS Statistics*. SAGE Publications.
- Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics*. SAGE Publications.
- Field, A., & Hole, G. (2018). *How to Design and Report Experiments*. SAGE Publications.
- Fox, J. (2015). *Applied Regression Analysis and Generalized Linear Models*. SAGE Publications.
- Gibbons, J. D. (2011). *Nonparametric statistics: An introduction* (2nd ed.). Sage.
- Gravetter, F. J., & Wallnau, L. B. (2017). *Essentials of Statistics for the Behavioral Sciences*. Cengage Learning.
- Gravetter, F. J., & Wallnau, L. B. (2017). *Statistics for the Behavioral Sciences*. Cengage Learning.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis*. Pearson.
- Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo Results in Methodology Articles: Some Rejoinders to Thompson and Pierce. *Journal of Experimental Education*, 61(2), 163-170.
- Hays, W. L. (2018). *Statistics for the Social Sciences*. Cengage Learning.
- Heiman, G. W. (2017). *Basic Statistics for the Behavioral Sciences*. Cengage Learning.
- Hollander, M., & Wolfe, D. A. (1999). *Nonparametric statistical methods* (2nd ed.). Wiley.
- Howell, D. C. (2013). *Estadística para las Ciencias del Comportamiento*. Cengage Learning.
- Howell, D. C. (2013). *Fundamental Statistics for the Behavioral Sciences*. Cengage Learning.
- Howell, D. C. (2013). *Statistical Methods for Psychology*. Cengage Learning.
- Jaén-Extremera, J., Afanador-Restrepo, D. F., Rivas-Campo, Y., Gómez-Rodas, A., Aibar-Almazán, A., Hita-Contreras, F., . . . Ortiz-Quesada, R. (2023). Effectiveness of Telemedicine for Reducing Cardiovascular Risk: A Systematic Review and Meta-Analysis. *J Clin Med*, 12(3). doi:10.3390/jcm12030841
- Jason, F. (2021). The correlation coefficient: What it is, what it tells investors. Investopedia. <https://www.investopedia.com/terms/c/correlationcoefficient.asp>
- Johnson, R. A., & Wichern, D. W. (2007). *Métodos Estadísticos*. Pearson Educación.
- Keppel, G., & Wickens, T. D. (2004). *Design and Analysis: A Researcher's Handbook*. Pearson.
- Kruskal, W., & Tanur, J. (2016). *The American Statistician*. Taylor & Francis.
- Lehmann, E. L. (2006). *Nonparametrics: Statistical methods based on ranks* (1st ed.). Springer.
- Levin, J., & Fox, J. A. (2014). *Elementary Statistics in Social Research: The Essentials*. Pearson.

- Levin, J., & Fox, J. A. (2016). *Elementary Statistics in Social Research: The Essentials*. Pearson.
- Lix, L. M., Keselman, H. J., & Keselman, J. C. (1996). Consequences of Assumption Violations Revisited: A Quantitative Review of Alternatives to the One-Way Analysis of Variance F Test. *Review of Educational Research*, 66(4), 579-619.
- Massé, E. L. (2005). *Estadística Aplicada a la Investigación Científica*. Limusa Wiley.
- Maxwell, S. E., & Delaney, H. D. (2004). *Designing Experiments and Analyzing Data: A Model Comparison Perspective*. Psychology Press.
- McElreath, R. (2016). *Inferential Statistics: A Bayesian Approach*. CRC Press.
- Missau, E., Teixeira, A. D. O., Franco, O. S., Martins, C. N., Paulitsch, F. da S., Perez, W., Silva, A. M. V. da, & Signori, L. U. (2015). Estadística aplicada a las ciencias de la salud. En *Estadística aplicada a las ciencias de la salud*. Elsevier España. <https://doi.org/10.1016/c2013-0-18792-2>
- Navarro, D. J. (2018). *Learning Statistics with R: A Tutorial for Psychology Students and Other Beginners*. University of Adelaide Press.
- Navarro, J. I. (2015). *Estadística para psicólogos*. Editorial Médica Panamericana.
- Newbold, P., Carlson, W. L., & Thorne, B. (2013). *Estadística para Administración y Economía*. Pearson Educación.
- Norušis, M. J. (2018). *IBM SPSS Statistics 25 Guide to Data Analysis*. Pearson.
- Olive, D. J. (2015). *Statistical Inference: A Short Course*. Routledge.
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), 210. doi:10.1186/s13643-016-0384-4
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., . . . Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj*, 372, n71. doi:10.1136/bmj.n71
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., . . . Moher, D. (2021a). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, 88, 105906. doi:<https://doi.org/10.1016/j.ijssu.2021.105906>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., . . . Moher, D. (2021b). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews*, 10(1), 89. doi:10.1186/s13643-021-01626-4
- Pallant, J. (2016). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS*. McGraw-Hill Education.

- Peck, R., Olsen, C., & Devore, J. (2018). *Statistics: The Exploration and Analysis of Data*. Cengage Learning.
- Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, Design, and Analysis: An Integrated Approach*. Psychology Press.
- Peña, D., Torgo, L., & Vallejo, C. (2016). *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*. Morgan Kaufmann.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. SAGE Publications.
- Rivas-Campo, Y., García-Garro, P. A., Aibar-Almazán, A., Martínez-Amat, A., Vega-Ávila, G. C., Afanador-Restrepo, D. F., . . . Hita-Contreras, F. (2022). The Effects of High-Intensity Functional Training on Cognition in Older Adults with Cognitive Impairment: A Systematic Review. *Healthcare (Basel)*, *10*(4). doi:10.3390/healthcare10040670
- Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of Behavioral Research: Methods and Data Analysis* (3rd Ed.). McGraw-Hill Education.
- Rosenthal, R., & Rosnow, R. L. (2008). *Artifacts in Behavioral Research: Robert Rosenthal and Ralph L. Rosnow's Classic Books*. Oxford University Press.
- Rouder, J. N., & Morey, R. D. (2012). A Bayesian Approach to Hypothesis Testing: The Savage-Dickey Method. *Psychological Science*, *23*(11), 1385-1393.
- Runyon, R. P., & Haber, A. (2018). *Fundamentals of Behavioral Statistics*. Cengage Learning.
- Salkind, N. J. (2017). *Statistics for People Who (Think They) Hate Statistics*. SAGE Publications.
- Sheskin, D. J. (2011). *Handbook of Parametric and Nonparametric Statistical Procedures*. CRC Press.
- Siegel, S., & Castellan Jr, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). McGraw-Hill.
- Stevens, J. (2012). *Applied Multivariate Statistics for the Social Sciences*. Routledge.
- Stuart, E. A., & Rubin, D. B. (2005). Introduction to the Theory of Hypothesis Testing. *The American Journal of Public Health*, *95*(3), 400-406.
- Sullivan, L. M. (2017). *Essentials of Biostatistics in Public Health*. Jones & Bartlett Learning.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics*. Pearson.
- Trochim, W. M. K., & Donnelly, J. P. (2007). *The Research Methods Knowledge Base*. Atomic Dog.
- Wackerly, D., Mendenhall, W., & Scheaffer, R. L. (2008). *Estadística Matemática con Aplicaciones*. Cengage Learning.

- Wasserman, L. (2006). *All of nonparametric statistics* (1st ed.). Springer.
- Wasserstein, R. L., & Lazar, N. A. (2016). The Role of Hypothesis Testing in Science. *The American Statistician*, 70(2), 104-109.
- Weakliem, D. L. (2010). Hypothesis Testing and Model Selection in the Social Sciences. *Sociological Methods & Research*, 39(4), 570-603.
- Webb, N. L. (2018). *Statistics for Business and Economics*. Cengage Learning.
- Wilcox, R. R. (2009). *Basic Statistics: Understanding Conventional Methods and Modern Insights*. Oxford University Press.
- Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing* (3rd ed.). Academic Press.
- Zar, J. H. (2013). *Biostatistical Analysis*. Pearson.
- Zimmerman, D. W. (2004). A Note on the Influence of Study Size on Statistical Significance. *Educational and Psychological Measurement*, 64(3), 391-402.

