

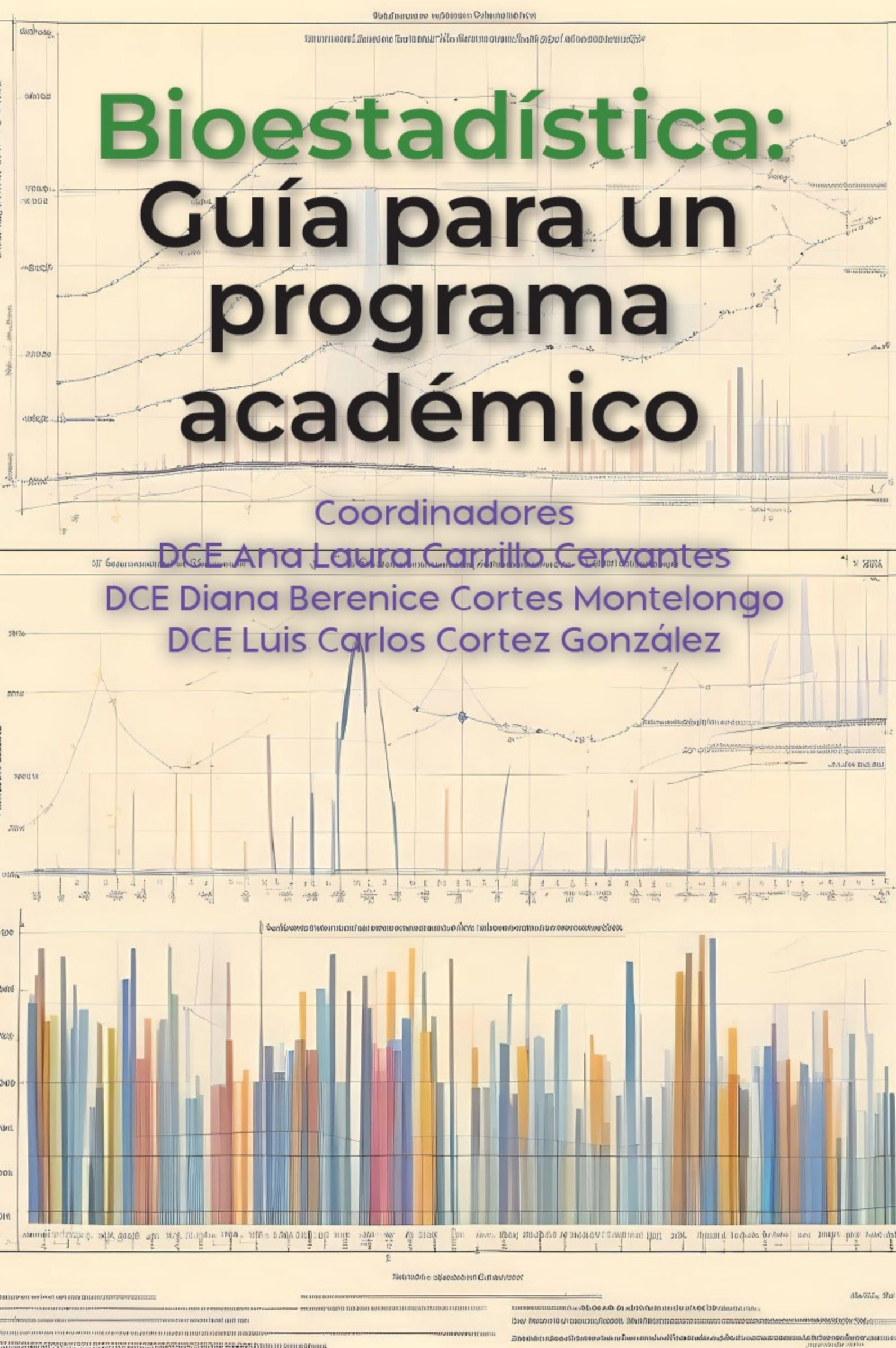
Bioestadística: Guía para un programa académico

Coordinadores

DCE Ana Laura Carrillo Cervantes

DCE Diana Berenice Cortes Montelongo

DCE Luis Carlos Cortez González



**Bioestadística:
Guía para un programa académico**

Bioestadística: Guía para un programa académico

Coordinadores

DCE Ana Laura Carrillo Cervantes

DCE Diana Berenice Cortes Montelongo

DCE Luis Carlos Cortez González



LA BIBLIOTECA

Bioestadística:
Guía para un programa académico

Coordinadores

DCE Ana Laura Carrillo Cervantes

DCE Diana Berenice Cortes Montelongo

DCE Luis Carlos Cortez González

Primera edición: diciembre 2023

D.R. © Grupo Editorial Biblioteca, S.A. de C.V.

Manantiales 29 - 4

Colonia Chapultepec

C.P. 62450, Cuernavaca, Morelos.

Tel. 55-6235-0157 y 55-3233-6910

Email: contacto@labiblioteca.com.mx

ISBN: 978-607-69591-4-5

Diseño tipográfico y de portada: Fernando Bouzas Suárez

Queda prohibida la reproducción parcial o total, directa o indirecta, del contenido de la presente obra, sin contar previamente con la autorización expresa y por escrito de los editores, en términos de lo así previsto por la Ley Federal de Derechos de Autor y, en su caso, por los tratados internacionales aplicables.

Impreso y encuadernado en México

Printed and bound in México

Índice

Presentación	7
Prologo	9
Capítulo 1. Definición y conceptos básicos	13
<i>DCE Diana Berenice Cortes Montelongo</i>	
Capítulo 2. Construcción de bases de datos y el uso de paquete estadístico, definición de conjuntos, cálculo de variables, recodificaciones y transformaciones	17
<i>DCS José Luis Nuncio Domínguez</i>	
Capítulo 3. Análisis descriptivo, frecuencias absolutas y relativas. Medidas de tendencia central, medidas de dispersión y medidas de posición	35
<i>DCEd Tirso Duran Badillo</i>	
<i>DCE Jesús Alejandro Guerra-Ordoñez</i>	
Capítulo 4. Representación gráfica de datos. Gráficas de barra simple, agrupados y segmentados. Gráficos de puntos. Histogramas. Generación de plantillas de gráficos en SPSS	43
<i>DCE Claudia Orozco Gómez</i>	
<i>DCE Marily Daniela Amaro Hinojosa</i>	
Capítulo 5. Tablas de contingencia	53
<i>DCE Daniel Sifuentes Leura</i>	
Capítulo 6. Chi cuadrada.	61
<i>DCE María Magdalena Delabra Salinas</i>	
<i>Dr. Ernesto Rios Willars</i>	
Capítulo 7. Distribución normal	67
<i>DCE Luis Carlos Cortez González</i>	
Capítulo 8. Medidas de asociación para datos numéricos y categóricos Correlación de Pearson. Correlación de Spearman	77
<i>DCE Ana Laura Carrillo Cervantes</i>	

Capítulo 9. Medidas de asociación para datos numéricos y categóricos. Correlación Tau B de Kendall. Tau C de Kendall	89
<i>DCE Claudia Jennifer Domínguez Chávez</i>	
<i>DCE Abraham Isaac Esquivel Rubio</i>	
Capítulo 10. Prueba de hipótesis paramétrica. Prueba T para muestras independientes. Prueba T para muestras relacionadas	99
<i>DSP Josué Arturo Medina Fernández</i>	
<i>DCS Nissa Yaing Torres Soto</i>	
Capítulo 11. Prueba de hipótesis paramétrica. Prueba T para una muestra Análisis de varianza de un factor (ANOVA)	107
<i>DCE Juana Edith Cruz Quevedo</i>	
<i>DCE Marcela Tenorio-Gómez</i>	
Capítulo 12. Prueba de hipótesis no paramétrica. Prueba H de Kruskal Wallis. Prueba U de Mann-Whitney. Rangos de Wilcoxon . . .	117
<i>DSP Isái Arturo Medina Fernández</i>	
<i>DCE Dafne Astrid Gómez Melasio</i>	
Capítulo 13. Análisis de regresión lineal. Regresión lineal simple y múltiple	125
<i>DCE Reyna Torres Obregón</i>	
Capítulo 14. Regresión logística	135
<i>DCE Edna Idalia Paulina Navarro Oliva</i>	
Capítulo 15. Pruebas epidemiológicas.	151
<i>DCE Patricia Enedina Miranda Félix</i>	
<i>DCE Rosario Edith Ortiz Félix</i>	
Capítulo 16. Normas APA para la presentación de datos estadísticos (tablas, figuras y gráficos)	163
<i>DCE Isaías Iván Briceño Rodríguez</i>	
Coordinadores	173
Colaboradores	175

Presentación

DRA. BÁRBARA DE LOS ÁNGELES PÉREZ PEDRAZA

Bioestadística: Guía para un programa académico, es un libro de texto que fue construido para llevar de la mano a estudiantes de licenciatura y maestría de enfermería y otras áreas de las ciencias de la salud en el procesamiento y análisis de datos a través de la estadística.

En estos 16 capítulos, se abordan, desde los conceptos básicos del procesamiento estadístico y la construcción de bases de datos, hasta la elaboración de modelos explicativos y su presentación en formato APA. El estudiante puede encontrar referencias básicas teóricas y prácticas para la elaboración de investigaciones basadas en evidencia, a través de la estadística descriptiva con medidas de tendencia central, de dispersión y de posición, así como de graficación. Estadística inferencial paramétrica y no paramétrica correlacional, pruebas de hipótesis, elaboración de modelos y epidemiología.

En este recorrido, el lector desarrollará habilidades para la toma de decisiones en función de las características de los instrumentos de medición, distribuciones y objetivos de investigación, pudiendo elegir, por ejemplo, entre pruebas de chi cuadrado, t de Student o U de Mann-Whitney para realizar comparación entre grupos; lo que es una función elemental para los profesionales de la salud que contribuyen a la generación del conocimiento.

Así mismo, el lector desarrollará habilidades para el procesamiento de datos en el paquete estadístico SPSS, esto debido a que los autores presentan detalladamente y paso a paso los procedimientos para el cálculo de diversos análisis estadísticos. Y finalmente, lo que más aporta a la generación del conocimiento, la traducción de lo que significan dichos resultados numéricos a la comprensión de los fenómenos de salud.

Es así que, los autores que conformaron este libro, son docentes investigadores con extensa experiencia en la generación del conocimiento y formación de profesionistas en el las área de salud en universidades reconocidas a lo largo del país; Diseñaron este libro para que sea incorporado la bibliografía de programas de grado y posgrado del área, sin embargo, también puede ser utilizado como una guía autodidacta por estudiantes para el procesamiento de datos, dada su presentación sencilla y afable con el lector con poca experiencia en el tema.

Espero disfruten su lectura y los retos que conlleva el ponerlo en práctica, pues es importante recordar que cada resultado numérico se traduce en una característica de salud de una persona o grupo de personas, lo que puede beneficiar en el desarrollo de tratamientos efectivos en el proceso de salud-enfermedad.

Prólogo

JOSÉ GONZÁLEZ TOVAR

Durante los últimos años, con el auge de la generación y acceso libre al conocimiento se ha evidenciado la presencia de una crisis, la llamada crisis de replicabilidad en ciencia, misma que se refiere a la preocupación creciente sobre la capacidad de reproducir los resultados de estudios científicos. En esencia, implica la dificultad o imposibilidad de otros investigadores de replicar o reproducir los estudios originales y obtener resultados consistentes con los informados por los autores originales de los estudios.

Esta crisis ha ganado atención en diversos campos científicos y se ha convertido en un tema importante de discusión en la comunidad académica. Hay dos razones (aunque existen más) clave que contribuyen a la crisis de replicabilidad: la primera es la inadecuada interpretación de las pruebas estadísticas, así como la falta de ajuste para múltiples comparaciones, puede conducir a resultados espurios que no son generalizables o replicables. La otra razón es la tendencia a publicar resultados positivos y significativos, mientras que los resultados negativos o no significativos pueden no ser reportados o publicados. Esto puede llevar a una visión sesgada de la realidad y hacer que los resultados positivos sean difíciles de replicar.

En la era de la información y a causa de esta crisis, el análisis de datos se ha convertido en una herramienta esencial para comprender y dar sentido al vasto conjunto de información que nos rodea para diferentes campos del conocimiento y en campo de la salud no es la excepción. En este contexto, la estadística se erige como una disciplina fundamental que permite extraer conocimientos valiosos a partir de datos aparentemente caóticos, nos ayuda a tomar las mejores de decisiones para intervenir y fundamentar en la evidencia cualquier plan de acción para la mejora de la salud, el bienestar y la calidad de vida.

El presente libro, es una obra colectiva que reúne la experiencia y el conocimiento de destacadas y destacados profesionales en el campo de la investigación del comportamiento saludable, quienes han contribuido con su experiencia y destacada trayectoria como investigadoras e investigadores para abordar de manera integral los diversos aspectos que conforman el análisis estadístico aplicado.

El texto se inicia con el capítulo titulado Definición y Conceptos Básicos, a cargo de Diana Berenice Cortés Montelongo. Este capítulo establece los cimientos necesarios para el lector, proporcionando una comprensión sólida de los principios fundamentales que sustentan el análisis estadístico, ya que para

un mejor entendimiento de los procedimientos se requiere de una breve, pero precisa revisión conceptual.

El segundo capítulo Construcción de Bases de Datos y el Uso de Paquete Estadístico, es texto de José Luis Nuncio Domínguez. En este capítulo, se explora la importancia de la construcción adecuada de bases de datos y el uso eficiente de paquetes estadísticos, abordando temas clave como la definición de conjuntos, el cálculo de variables, así como recodificaciones y transformaciones. En el entendido de que, una preparación correcta de una base de datos de un proyecto de investigación le da una riqueza y valor agregado a los resultados que se analizan mediante cualquier procedimiento estadístico.

Los Capítulos 3 y 4, a cargo de Tirso Durán Badillo, Jesús Alejandro Guerra-Ordoñez, Claudia Orozco Gómez, y Marily Daniela Amaro Hinojosa, se sumergen en el análisis descriptivo y la representación gráfica de datos. Desde frecuencias absolutas y relativas hasta la generación de plantillas de gráficos en el software SPSS, estos capítulos proporcionan una visión completa de cómo presentar y comprender visualmente conjuntos de datos aplicados a la investigación en salud.

El libro avanza hacia aspectos más específicos de la estadística, como el Capítulo 5 sobre la prueba Chi cuadrado, dirigido por Daniel Sifuentes Leura. Le siguen temas como las tablas de contingencia, abordadas por María Magdalena Delabra Salinas, y la distribución normal, presentada por Luis Carlos Cortez González.

Los capítulos subsiguientes, a cargo de Ana Laura Carrillo Cervantes, Claudia Jennifer Domínguez Chávez, Josué Arturo Medina Fernández, y Juana Edith Cruz Quevedo, profundizan en medidas de asociación como son las correlaciones y pruebas de hipótesis tanto desde la perspectiva paramétrica así como las no paramétricas.

Los últimos capítulos, bajo la dirección de destacados expertos como Reyna Torres Obregón, Edna Idalia Paulina Navarro Oliva, Patricia Enedina Miranda Félix, y Rosario Edith Ortiz Félix, exploran análisis de regresión, regresión logística, pruebas epidemiológicas y las normas APA para la presentación de datos estadísticos. Estas herramientas permiten la presentación de resultados con estadísticos orientados en la generación de modelos que permitan predecir y explicar el comportamiento saludable.

Cada capítulo de esta obra se configura como una pieza crucial en el vasto rompecabezas del análisis estadístico, contribuyendo con conocimientos esenciales que son fundamentales para desarrollar una comprensión informada y rigurosa de la disciplina. La estructura del libro se convierte así en un marco integral que guía al lector a través de conceptos clave, métodos y técnicas, proporcionando una base sólida para abordar los desafíos y aplicaciones prácticas de la estadística.

Este libro no solo se limita a ofrecer teorías abstractas, sino que también se enfoca en su aplicabilidad práctica. Cada página se convierte en un recurso valioso, facilitando el entendimiento y la aplicación de la estadística en diversos contextos, desde la investigación académica hasta las demandas del entorno profesional. Los lectores encontrarán en estas páginas una brújula confiable que los orientará a través de la complejidad del análisis estadístico, permitiéndoles tomar decisiones informadas basadas en un fundamento sólido.

Además, la obra no solo destaca por su contenido sustantivo, sino también por su enfoque pedagógico, que busca hacer accesibles los conceptos estadísticos a una amplia audiencia. La claridad en la exposición de ideas y la presentación de ejemplos prácticos convierten a este libro en una herramienta educativa efectiva para estudiantes que buscan consolidar sus conocimientos y para profesionales que desean perfeccionar sus habilidades analíticas.

En síntesis, *Análisis Estadístico: Fundamentos y Aplicaciones* no solo es un libro de referencia, sino también un compañero invaluable en el viaje hacia la comprensión profunda y aplicada de la estadística. Cada página es un paso más hacia la maestría en el análisis estadístico, proporcionando a los lectores las herramientas necesarias para enfrentar los desafíos estadísticos con confianza y competencia.

Capítulo 1

Definición y conceptos básicos

DCE DIANA BERENICE CORTES MONTELONGO

La estadística es un campo de estudio que se ocupa de recopilar, organizar y resumir datos y sacar conclusiones sobre conjuntos de datos. Los conceptos y herramientas estadísticas se utilizan en muchos campos profesionales. Sin embargo, cuando los datos que se analizan provienen de las ciencias biológicas o la medicina, el término bioestadística se utiliza para distinguir esta aplicación particular de la herramienta o concepto. Este alcance es el tema de este capítulo (Daniel, 2016; Peña, 2014).

La unidad funcional de la bioestadística y la investigación son los datos. Los datos son una colección desorganizada de hechos y cifras de diversas fuentes. Por lo tanto, las fuentes de datos pueden variar dependiendo de las necesidades de investigación y análisis que se realicen (Daniel, 2016).

El análisis y la interpretación de datos se basan únicamente en la recopilación de diferentes tipos de datos de las fuentes. Los investigadores o analistas realizan tareas de recolección de datos para recopilar información como registros de rutina (registros clínicos, archivos). Si la información no se incluye en los registros de rutina, se puede realizar una encuesta que pueda proporcionar información a través de una serie de preguntas específicas (Grove & Gray, 2019).

Para obtener datos más específicos sobre un tema o para medir una variable específica de un estudio, se selecciona un instrumento de medición. Una de las características de los instrumentos de medición es que deberán ser fiables y válidos. La fiabilidad se centra en la estabilidad de un método de medición (que se obtengan los mismos resultados en repetidas ocasiones) y la validez de un instrumento es la determinación de lo bien que dicho instrumento mide el concepto que se está explorando. Los datos que se obtienen de censos, datos de encuestas nacionales o banco de datos de investigaciones previas son considerados fuentes externas para la obtención de datos.

La unidad de análisis es el elemento del cual se muestra una propiedad y característica. Las variables es la característica, propiedad o atributo que se predica de la unidad de análisis, son conceptos con diversos niveles de abstracción que se miden, manipulan o controlan en un estudio, toman diferentes valores en cada persona, lugar o situación en donde se esté valorando. Se describen dos tipos de variables, las variables cuantitativas y las variables cualitativas (Figura1) (Grove & Gray, 2019).

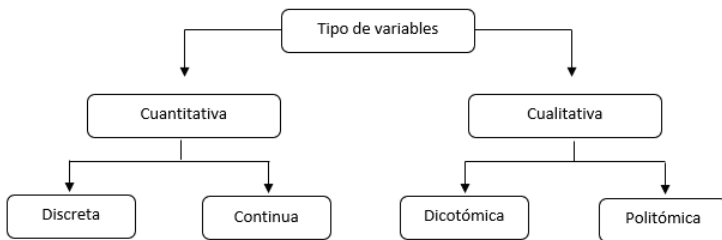
Una variable cuantitativa es aquella que puede medirse y se puede expresar de forma numérica. La característica de la persona u objeto es interpretada con un número. Por ejemplo, la edad, número de hijos, ingresos económicos (Daniel, 2014; Grove & Gray, 2019).

Las **variables cualitativas** son características (cualidades) de un individuo u objeto, que se pueden **expresar con palabras**. Las mediciones hechas sobre este tipo de variables contienen información respecto a los atributos. Las variables cualitativas también son conocidas con el nombre de categóricas. Este tipo de variables puede, a su vez, ser dicotómicas o politómicas. Son **dicotómicas** cuando solamente pueden adquirir dos valores y son **politómicas** cuando la variable puede asumir más de dos valores. **Ejemplo de variables cualitativas, Color de ojos, sexo, estado civil** (Daniel, 2014; Grove & Gray, 2019; Pereira, 2023).

Dependiendo de las características de la variable cuantitativa, esta puede ser discreta o continua. Las variables discretas se caracterizan por separaciones o rupturas en el rango de valores posibles. Estas separaciones o delimitadores indican que no existen valores entre ciertos valores posibles de la variable, y están representados por números enteros. Por ejemplo, el número de camas de hospital podría ser 1, 2, 3, 1,5 o 1,7, pero no se pudo determinar (Daniel, 2014; Grove & Gray, 2019).

Las **variables continuas** pueden tomar cualquier valor dentro de un intervalo especificado de valores asumidos por la variable. Diversas mediciones corporales pueden ser variables continuas, por ejemplo, el peso, la estatura y circunferencias corporales (Daniel, 2014; Grove & Gray, 2019).

Figura 1
Tipos de variables



Nota: Elaboración propia (Cortes, 2023).

Diversos instrumentos y escalas de medición dan lugar a datos que están a diferentes niveles. Por tanto, se establece una jerarquía en las mediciones. Los niveles de medición, de menor a mayor, son nominal, ordinal, de intervalo y de razón (Figura 2).

El nivel de **medición nominal** se utiliza cuando los datos pueden organizarse en categorías de una propiedad definida, pero no pueden ordenarse por ran-

go. Las categorías no pueden ordenarse y son excluyentes, es decir, solo un dato encaja en una categoría no es posible establecerse en dos categorías. Ejemplo; Sexo (hombre, mujer), estado civil (soltero, casado, viudo, unión libre), grupo sanguíneo (A, B, O, AB) (Dagnino, 2014; Grove & Gray, 2019).

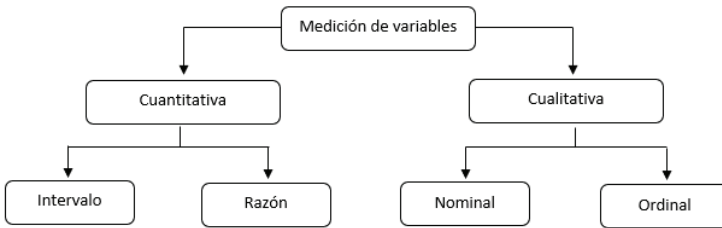
Con el nivel de **medición ordinal**, los datos se asignan a categorías que pueden ordenarse. Para ordenar los datos, se ordena si una categoría es menor a mayor, o mejor o peor que otra. Las reglas y formas de medición e interpretación determinan cómo se ordenan los datos (Dagnino, 2014; Grove & Gray, 2019).

El nivel de **medición de intervalo** usa escalas que tienen la misma distancia numérica entre los intervalos, las categorías deben tener intervalos iguales entre ellas, la magnitud del atributo puede definirse de forma más precisa, la escala de intervalo carece de punto cero (Dagnino, 2014; Grove & Gray, 2019).

Cuando la medida es en **medición de razón** el cero significa ausencia de cantidad, es decir que ese valor tiene un significado real, lo anterior significa que en una escala de razón no pueden existir números negativos. Ejemplo, peso, talla, circunferencias (Dagnino, 2014; Grove & Gray, 2019, Pereira, 2023).

Los datos a nivel de intervalo y de razón pueden sumarse, restarse, multiplicarse y dividirse debido a los intervalos iguales y a la continuidad de valores de estos datos. Por tanto, los datos de intervalo y de razón pueden analizarse con técnicas estadísticas para determinar relaciones y diferencias significativas en los análisis (Grove & Gray, 2019).

Figura 2
Escala de medición de las variables



Nota: Elaboración propia (Cortes, 2023).

El método estadístico que se va a utilizar para realizar un análisis será determinado por el tipo de variable y la escala de medición, de allí la importancia de su identificación (Pereira, 2023).

A continuación, se presentan algunos ejemplos de tipos de variables y escala de medición.

Ejemplo de tipos de variable y escala de medición

Variable	Tipo de variable	Escala de medición
Sexo	Cualitativa dicotómica	Nominal
Color de ojos	Cualitativa politómica	Nominal
Escala de dolor	Cualitativa politómica	Ordinal
Temperatura	Cuantitativa continua	Intervalo
Número de hijos	Cuantitativa discreta	Razón
Circunferencia de cintura	Cuantitativa continua	Razón

Nota: Elaboración propia.

Para el análisis de las variables y los datos la bioestadística se puede clasificar por dos etapas, la bioestadística descriptiva y la bioestadística inferencial. La **bioestadística descriptiva** se ocupa de describir la muestra, y la **bioestadística inferencial** infiere conclusiones a partir de los datos que describen la muestra con respecto a la población (Pereira, 2023).

Una **población** es el conjunto de todos los elementos (individuos, objetos o sustancias) que cumplen ciertos criterios para su inclusión en un estudio, mientras que una muestra es un subconjunto de la población seleccionada para un estudio particular y una muestra es un subconjunto de la población. seleccionados para un estudio en particular, los miembros son participantes (Grove & Gray, 2019).

Tanto la fase de descripción como la de inferencia manejan la adquisición de nuevos datos. La diferencia es que la fase descriptiva continúa resumiendo y organizando estos datos para facilitar el análisis y la interpretación, y la fase inferencial proporciona estimaciones sobre la población basadas en estos datos resumidos de la muestra. Continuaremos formulando y probando hipótesis.

Para poder realizar estos análisis es necesario la elaboración de una **base de datos**, que es la recopilación sistemática de datos de forma electrónica mediante un programa estadístico que ayudará a la organización y procesamiento de los datos y poder realizar la interpretación de estos. Así es como se brinda la base para poder realizar los análisis descriptivos e inferenciales de un estudio en particular.

Referencias

- DAGNINO, S. J. (2014). Tipos de datos y escalas de medida. *Revista Chilena Anestesiología*, 43, 109-111.
- GROVE, S. K., & Gray, J. R. (2019). *Investigación en enfermería: Desarrollo de la práctica enfermera basada en la evidencia*. Elsevier Health Sciences.
- PEÑA, D. (2014). *Fundamentos de estadística*. Alianza editorial.
- PEREIRA-GONZÁLEZ, L. M., & Basantes-Andrade, A. (2023). Probabilidad y estadística.

Capítulo 2.

Construcción de bases de datos y el uso de paquete estadístico, definición de conjuntos, cálculo de variables, recodificaciones y transformaciones

DCS JOSÉ LUIS NUNCIO DOMÍNGUEZ

Introducción

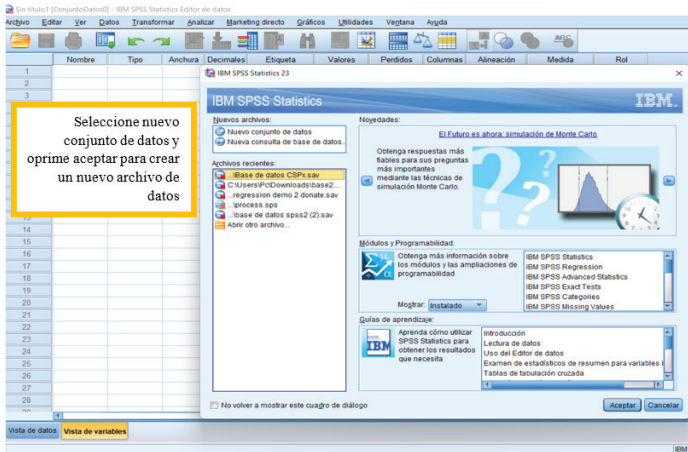
Los profesionales de la salud necesitan de poseer herramientas que le permitan desarrollar el trabajo de forma eficiente mediante los diversos programas estadísticos para la interpretación de los datos para transformarlos en conocimientos y su posterior toma de decisiones.

El paquete estadístico para ciencias sociales (*Statistical Package for the Social Sciences [SPSS]*) de IBM es un programa estadístico informático y es uno de los softwares más utilizado para la realización de las investigaciones relacionadas a las ciencias de la salud, la sociología y la psicología. Cabe mencionar que el SPSS es de fácil uso, con una sencilla interfaz para la mayoría de los análisis, grandes bases de datos, por lo anterior, se convierte en una herramienta de gran utilidad para este tipo de trabajos (William & Wagner, 2019). El programa contiene un módulo de base y módulos anexos que se han ido actualizando con nuevos procedimientos estadísticos. Cada uno de estos módulos se compra por separado.

Construcción de base de datos

Al inicio del software se despliega la información que indica la versión del software y la información del propietario de licencia de este. Cuando se abre el paquete estadístico SPSS se observa una ventana de “Bienvenida” la cual contiene distintas opciones, donde lo más utilizado es “Nuevos archivos:” y “Archivos recientes”, la primera se utiliza cuando se pretende crear una nueva base de datos y la segunda cuando se quiere trabajar sobre una base de datos ya existente (Figura 1). El paquete estadístico SPSS Statistics está compuesto por dos ventanas de trabajo editor de datos y visor de resultados (IMB, 2019).

Figura 1
Ventana de bienvenida SPSS



Editor de datos

El editor de datos es similar a la hoja de cálculo la ventana donde se ingresan los datos con los que se desea trabajar, las columnas (hacia la izquierda) representan las variables con las que se desea trabajar y las filas (hacia abajo) corresponde a los datos o las respuestas que se obtienen de los encuestados o donde sea que se hayan obtenidos los datos para la realización de la base de datos.

En el editor de datos cuenta con dos tipos de vistas o ventanas: “Vista de datos” (Figura 2) y “Vista de variables” (Figura 3). En la primera, el investigador observa los datos ingresados y en la segunda se accede a los atributos de las distintas variables para poder ser editadas (IBM, 2020).

Figura 2
Visor de datos SPSS

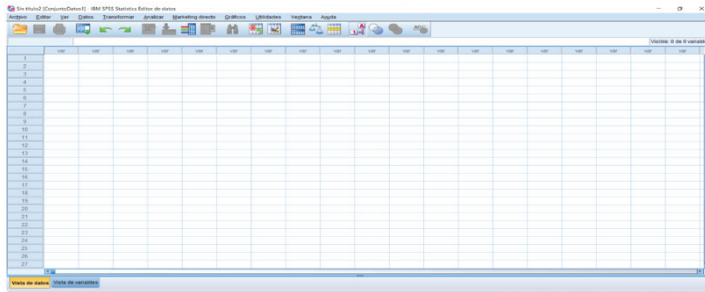
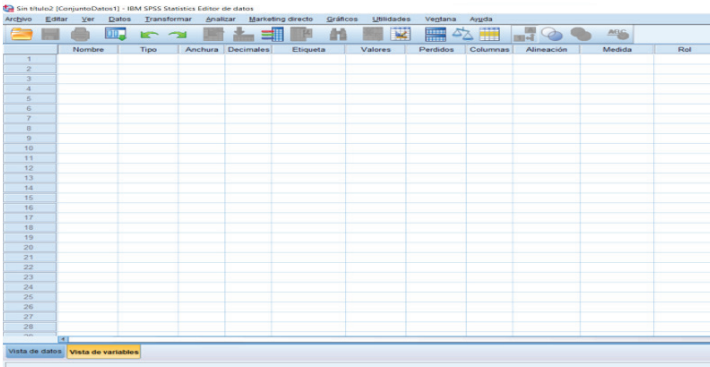


Figura 3
Vista de variables SPSS



Una vez que se ha trabajado con al menos una variable o un dato ingresado se podrá guardar en archivo > guardar como: > “Buscar en” (lugar en el ordenador donde se desea guardar el visor de datos) > “nombre de archivo” > guardar, el tipo de archivo guardado será SPSS Statistics Data Document (.sav).

La vista de datos es similar a las hojas de cálculo y se describen a continuación:

- En las filas se encuentran los casos y estas corresponden a un caso u observación, en este sentido, cada sujeto de estudio que responde una entrevista o cuestionario es un caso.
- Las columnas corresponden a las variables, cada columna representa a una variable o característica que se mide, cada una de las columnas representan a una pregunta o una característica que se mide en un cuestionario y que corresponde a una variable.
- Las casillas contienen valores y el valor es único de una variable para cada caso o sujeto de estudio, una de las diferencias en las casillas del editor de datos con los programas de hojas de cálculo es que en las casillas no puede contener formulas. En el caso de las variables numéricas, cada casilla vacía se convierte en valor perdido en el sistema y para las variables de cadena, una casilla vacía es un valor valido.

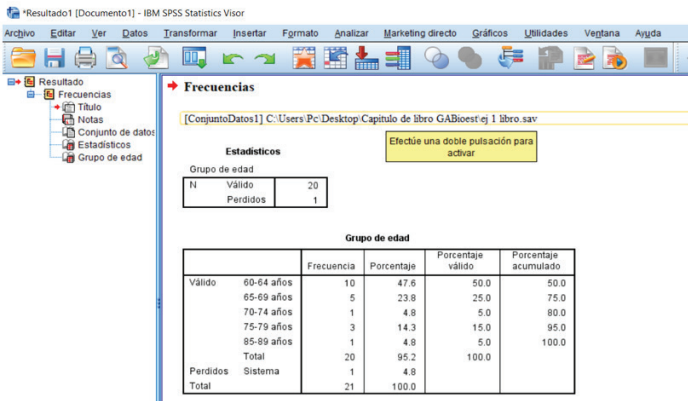
Visor de resultados

El Visor de resultados de SPSS (Figura 4) corresponde a una ventana en la que los paquetes estadísticos muestran los resultados de varios análisis realizados en el editor de datos. A la izquierda se pueden ver los títulos de los distintos análisis y se muestran como un dendrograma. En el lado derecho se muestran resultados como tablas, gráficos y resultados estadísticos que se pueden extraer

a otro software como Microsoft Office Word o Excel para editar o ejecutar el informe o realizar cualquier otra acción que el investigador considere relevante.

La ventana del visor se abre automáticamente la primera vez que se ejecuta un procedimiento desde el visor de datos que genere resultados y se crea automáticamente. Una vez que se ha trabajado con al menos un análisis de resultados se podrá guardar en archivo, guardar como > “Buscar en” (lugar en el ordenador donde se desea guardar el visor de resultados) > “nombre de archivo” > guardar, el tipo de archivo guardado será SPSS Statistics Output Document (.spv).

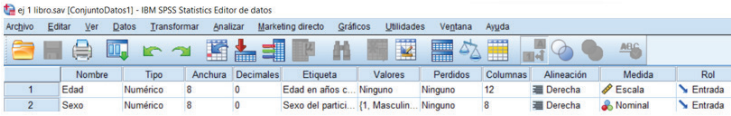
Figura 4
Visor resultados SPSS



Base de datos

Cuando se ha abierto el paquete estadístico SPSS y se ha ingresado a una base de datos, en primer lugar, se tiene que dirigir a al editor de “vista de variables” y en este apartado se crearan las variables que el investigador considere necesarias para su estudio. Se encuentra una serie de parámetros o atributos para cada una de las variables, se describen en el orden de aparición: Nombre, Tipo, Anchura, Decimales, Etiqueta, Valores, Perdidos, Columnas, Alineación, Desconocido, Rol (Figura 5).

Figura 5
Atributos de variable SPSS

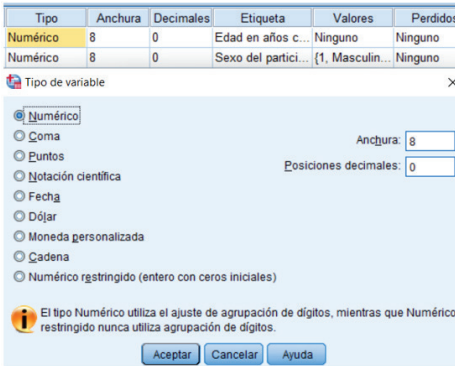


	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Edad	Numérico	8	0	Edad en años c...	Ninguno	Ninguno	12	Derecha	Escala	Entrada
2	Sexo	Numérico	8	0	Sexo del partici...	{1, Masculin...	Ninguno	8	Derecha	Nominal	Entrada

En la sección del “Nombre” se debe de asignar un nombre para la variable, el cual debe contener un máximo de 8 letras y o números, se debe de evitar uso de símbolos, caracteres especiales como ñ o ü y tildes, además no se deberá de comenzar el nombre de la variable con un numero pues aparecerá una leyenda “El primer carácter del nombre de la variable no es válido”.

Los tipos de variables incluyen la capacidad de asignar características a las variables para facilitar la manipulación y dar a los paquetes estadísticos la interpretación correcta de la variable. De forma predeterminada, se supone que todas las variables nuevas son numéricas. El cuadro de diálogo Tipo de variable proporciona las siguientes opciones para cambiar el tipo de datos que ingresa: Este contenido incluye: números, comas, puntos, notación científica, fechas, dólares, monedas personalizadas, cadenas y números restringidos (enteros). (con un cero a la izquierda), también se muestra una sección que contiene el ancho y el punto decimal en la parte superior derecha (Figura 6).

Figura 6
Características del tipo de variable SPSS



Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos
Numérico	8	0	Edad en años c...	Ninguno	Ninguno
Numérico	8	0	Sexo del partici...	{1, Masculin...	Ninguno

Tipo de variable

Numérico

Coma

Puntos

Notación científica

Fecha

Dólar

Moneda personalizada

Cadena

Numérico restringido (entero con ceros iniciales)

Anchura:

Posiciones decimales:

El tipo Numérico utiliza el ajuste de agrupación de dígitos, mientras que Numérico restringido nunca utiliza agrupación de dígitos.

Aceptar Cancelar Ayuda

- Numérico: es una variable cuyo valor son números y los valores se muestran en formato numérico estándar o en notación científica.
- Coma: es una variable numérica donde se muestran con comas que delimitan cada 3 posiciones y con el punto como delimitador decimal, este tipo de variables son valores numéricos con o sin comas o en notaciones científicas.

Los valores no pueden contener comas a la derecha después del indicador decimal.

- Punto: es una variable numérica donde los valores se muestran con puntos que delimitan cada tres posiciones y con la coma como delimitador decimal, este tipo de variables acepta valores numéricos con o sin puntos o en notación científica. Los valores no pueden contener puntos a la derecha después del indicador decimal.
- Notación científica: es una variable numérica donde el valor se muestra con una E intercalada y un exponente con signo que representan una potencia de base de 10, este tipo de variable acepta valores numéricos con o sin exponente, donde el exponente puede aparecer precedido por una E o una D con un signo opcional o bien solo por el signo (ejemplo 1E+002, 5E+000, 1D+002).
- Fecha: es una variable numérica la cual muestra uno de los diversos formatos de fecha-calendario y se puede introducir utilizando como delimitadores las barras inclinadas “/”, guiones “-”, puntos “.”, comas “,” o espacios “ ”. también se encuentra el formato hora-reloj.
- Dólar: es una variable numérica que se muestra con el signo “\$”, comas que delimitan cada tres posiciones y punto como delimitador decimal, se puede introducir valores de datos con o sin el signo de dólar inicial. Para este tipo de variables la puntuación decimal no debe ser superior a dos.
- Moneda personalizada: es una variable numérica la cuál contienen valores que se haya elegido en uno de los dos formatos de moneda personalizada en el cuadro de diálogo de opciones.
- Cadena: es una variable la cual sus valores no son números, por lo que no podrán ser utilizados cálculos, estos valores pueden contener cualquier carácter y el numero de estos debe de ser definidos para limitar su longitud (ejemplo: si introducimos un límite de caracteres de 8 como viene predeterminado, en este caso no podremos introducir palabras o valores superiores a este límite, no se podría introducir “Hipertensión” o “Diabetes mellitus” pero si HTA o DM, para poder introducir la palabra completa se debe de ampliar a más de 18 caracteres en caso de la diabetes mellitus). Las palabras mayúsculas y las minúsculas serán consideradas como diferentes, los valores se pueden introducir siempre y cuando no exceda la longitud definida previamente.

El “Ancho de columna” se puede especificar un número de caracteres para el ancho de la columna de la variable elegida que pueden cambiarse o quedarse como aparece por defecto que es 8, esto no altera la variable que se desea introducir.

Los “Decimales” aparecen por defecto 2, no obstante, se puede cambiar según sea la necesidad del valor a introducir, por ejemplo, si la variable del investigador es peso se puede modificar a 3 decimales y podrá introducir valores

como 75.400, en cambio, cuando se habla de la edad se puede modificar a 0 para eliminar las decimales y al introducir la variable aparecerá como 45 y no como 45.00 si no se modifica las decimales.

La “Etiqueta” es para introducir correctamente los nombres de las variables sin límite de caracteres o letras como lo limita el “nombre” de la variable (ya que solo se limita a 8 letras o caracteres), se puede establecer una etiqueta para cada una de las variables de creación según considere el investigador.

Los “Valores” pueden asignar etiquetas de valor descriptivas a cada valor de una variable, los valores son muy útiles si el archivo de datos utiliza códigos numéricos para representar una categoría que no es numérica como el sexo o el estado civil. Por ejemplo, si la variable es sexo se puede asignar un valor para cada tipo (se puede asignar 1 para masculino y 2 para femenino) (Figura 7), también se puede utilizar en escalas de tipo Likert (1 = totalmente en desacuerdo, 2 = en desacuerdo, 3 = de acuerdo, 4 = totalmente de acuerdo). En la casilla de valores que se desee agregar etiqueta de valores se debe hacer clic en los tres puntos de la casilla del valor y aparecerá un cuadro con la etiqueta de valores, posterior se agrega el valor deseado, la etiqueta para el valor y se pulsa añadir para continuar con otro valor. Una vez que se han introducido los valores se pulsa aceptar y esta variable quedará lista con los valores agregados.

Figura 7
Etiqueta de valores SPSS

Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
Edad en años c...	Ninguno	Ninguno	12	Derecha	Escala
Sexo del partici...	1, Masculin...	Ninguno	8	Derecha	Nominal
Escolaridad del...	1, Masculin...	Ninguno	8	Derecha	Nominal

Etiquetas de valores

Valor: Ortografía...

Etiqueta:

1 = "Masculino"

Los “Valores perdidos” se define como los valores de datos definidos como perdidos del usuario. Un ejemplo, es probable que una persona encuestada se niegue a contestar una o varias preguntas o variables, por que dicha(s) variable(s) no afectan o no aplican al entrevistado por lo que estos datos no son porque el investigador ha capturado mal u omitido un valor sino porque el entrevistado no cuenta con los datos de la variable. Cuando un valor de datos se especifica

como perdido aparece señalado como excluido y esto implica ajustar el cálculo de la variable y que éste no afecte a los resultados. Para definir un valor perdido se debe pulsar en la casilla de perdido de la variable que se quiere definir (Figura 8), se introduce los valores o el rango de valores que representen los datos perdidos como por ejemplo si una persona contesta que no fuma y la siguiente variable es cuantos cigarrillos fuma al día, se selecciona la variable Valores perdidos y se selecciona valores perdidos discretos y puede agregar un código como 777 o 999, se finaliza con aceptar y el código o códigos identificaran que estos sujetos de estudio no aplica esta variable.

Figura 8
Valores perdidos SPSS

Selección de valores perdidos discretos en el diálogo 'Valores perdidos':

- No hay valores perdidos
- Valores perdidos discretos
 - 777 999
- Rango más un valor perdido discreto opcional

Botones: Aceptar, Cancelar, Ayuda

Tabla de Frecuencia para 'fuma':

Valor	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Si	10	70.0	70.0	70.0
No	3	30.0	30.0	100.0
Total	13	100.0	100.0	

Tabla de Frecuencia para 'Número de cigarrillos que fuma al día':

Valor	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
1	1	10.0	14.3	14.3
8	1	10.0	14.3	28.6
10	1	10.0	14.3	42.9
12	1	10.0	14.3	57.1
15	2	20.0	28.6	85.7
20	1	10.0	14.3	100.0
Perdidos	7	70.0	100.0	
Total	10	100.0		

Valores perdidos (777 y 999) circunscritos en rojo en la tabla de frecuencias.

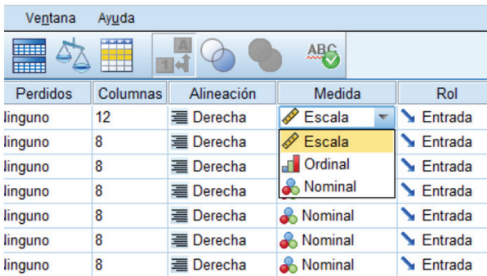
Las “Columnas” se puede especificar un número de caracteres (8 por defecto) para el ancho de la columna de la variable elegida, y estos también pueden cambiarse en la Vista de datos, pulsando y arrastrando los bordes de las columnas. El ancho será según los caracteres y depende de las características de los valores que se desea introducir y dichos valores se mostraran más o menos caracteres

con el ancho determinado, sin embargo, aunque no se visualice el valor introducido este no altera el valor introducido.

La ubicación de las variables afecta la visualización de valores de datos y/o etiquetas de valores en la vista de datos. De forma predeterminada, aparece a la derecha para variables numéricas y a la izquierda para variables de cadena, y solo afecta la configuración de datos. Audiencia. La orientación variable puede ser izquierda, derecha o centro, según la preferencia del investigador.

La “Medida” hace referencia a las características de las unidades de medida de cada variable con el fin de que el paquete estadístico realice correctamente los análisis que desea obtener, por lo que las variables pueden ser de tres opciones: Escala, Ordinal y Nominal (Figura 9). La variable será modificada automáticamente y no por defecto, según los valores obtenidos en la captura de datos (esto dependerá de la versión de IBM SPSS que se esté utilizando).

Figura 9
Medida de la variable SPSS



Perdidos	Columnas	Alineación	Medida	Rol
linguno	12	Derecha	Escala	Entrada
linguno	8	Derecha	Escala	Entrada
linguno	8	Derecha	Ordinal	Entrada
linguno	8	Derecha	Nominal	Entrada
linguno	8	Derecha	Nominal	Entrada
linguno	8	Derecha	Nominal	Entrada
linguno	8	Derecha	Nominal	Entrada

Las variables de “Escala” son datos numéricos de una escala de intervalo o de razón, este tipo de variables, los valores representan categorías ordenadas, con una métrica con significado por lo que son adecuadas las comparaciones de distancia entre cada valor. Un ejemplo puede ser la edad, el ingreso económico, años de escolaridad. Su valor será numérico siempre

La variable “Nominal” es cuando sus valores representan categorías y no representa un orden específico. Por ejemplo, el departamento o servicio de una enfermera, religión, sexo, estado civil u ocupación. Su valor puede ser numérico o alfanumérico.

La variable “Ordinal” es cuando su valor representa categorías con algún orden determinado por ejemplo la satisfacción del usuario en el servicio de urgencias que puede ir desde muy insatisfecho hasta muy satisfecho, o que tan de acuerdo se encuentra con la atención brindada por el personal de salud de una institución determinada, también pueden ser las puntuaciones de evaluación de las preferencias sobre algún tema en particular. Su valor puede ser numérico o alfanumérico.

Libro de códigos

Un libro de códigos (Tabla 1) es una ficha en la cual se incluyen todos los valores numéricos asignados a cada una de las variables, de tal forma que al capturar los datos al SPSS el investigador sepa que rango de valores debe de usar para cada una de las variables de la base de datos. Se muestra un ejemplo para la creación de un libro de códigos con el instrumento Short Portable Mental State Questionnaire (SPMSQ) de Pfeiffer (Figura 10).

Figura 10

Instrumento de medición Short Portable Mental Status Questionnaire de Pfeiffer

Short Portable Mental Status (SPMSG) de Pfeiffer

Instrucciones: Si es aplicado directamente para responder por el adulto mayor, colocar las respuestas en el espacio, después el aplicador convertirá a clave.

De lectura textualmente a cada pregunta, mediante entrevista al adulto mayor y conteste 1 (si es falso) y 0 "cero" (en caso de que sea verdadero)

Edad: _____

Sexo: _____ Hombre _____ Mujer

¿Hasta qué año estudio, Usted?

1.- Ninguno

4.- Primaria completa

7.- Profesional

2.- Sabe leer y/o escribir

5.- Secundaria

8.- Otra

3.- Primaria incompleta
6.- Bachillerato / técnico

		Respuesta del AM	Clave
1	¿En qué día estamos? ¿Día, mes y año?		
2	¿En qué día de la semana estamos?		
3	¿En qué lugar nos encontramos?		
4	¿Cuál es el número de teléfono que tiene? ¿Cuál es su dirección? (en caso de que no tenga teléfono)		
5	¿Qué edad tiene?		
6	¿Cuál es la fecha de su nacimiento?		
7	¿Cómo se llama el actual presidente de la república?		
8	¿Cómo se llama el anterior presidente?		
9	¿Cómo se llama o llamaba su mamá?		
10	¿Si de 30, restamos 3 cuantos quedan? ¿Y si quitamos 3 más cuantos quedan? ¿Y si quitamos 3 más cuantos quedan?		
		Total	
	Nivel de funcionamiento Intelectual		

Tabla 1
Libro de códigos

# variable	Nombre de variable	Etiqueta de la variable	Valores
1	Edad	Edad en años cumplidos	No aplica
2	Sexo	Sexo del participante	1= Masculino 2= Femenino
3	Escolaridad	Escolaridad del participante	1= Ninguna 2= Sabe leer y escribir 3= Primaria incompleta 4= Primaria completa 5= Secundaria 6= Bachillerato/Preparatoria 7= Profesional 8= Otra
4	SPMSQ1	¿En qué día estamos? ¿Día, mes y año?	0= Verdadero 1= Falso
5	SPMSQ2	¿En qué día de la semana estamos?	0= Verdadero 1= Falso
6	SPMSQ3	¿En qué lugar nos encontramos?	0= Verdadero 1= Falso
7	SPMSQ4	¿Cuál es el número de teléfono que tiene? ¿Cuál es su dirección? (en caso de que no tenga teléfono)	0= Verdadero 1= Falso
8	SPMSQ5	¿Qué edad tiene?	0= Verdadero 1= Falso
9	SPMSQ6	¿Cuál es la fecha de su nacimiento?	0= Verdadero 1= Falso
10	SPMSQ7	¿Cómo se llama el actual presidente de la república?	0= Verdadero 1= Falso
11	SPMSQ8	¿Cómo se llama el anterior presidente?	0= Verdadero 1= Falso
12	SPMSQ9	¿Cómo se llama o llamaba su mamá?	0= Verdadero 1= Falso
13	SPMSQ10	¿Si de 30, restamos 3 cuantos quedan? ¿Y si quitamos 3 más cuantos quedan? ¿Y si quitamos 3 más cuantos quedan?	0= Verdadero 1= Falso

Definición de conjuntos de variables

Un conjunto de variables es un conjunto de subconjuntos de variables que aparecen en el editor de datos y en la lista de variables de los cuadros de diálogo. El conjunto de variables definido se guarda con el archivo de datos de IBM SPSS Statistics. Los nombres de los registros pueden tener hasta 64 bytes de longitud. Se permiten todos los caracteres, incluidos los espacios. Un conjunto de variables puede constar de cualquier combinación de variables numéricas y de cadena. El orden de las variables en un conjunto no afecta el orden en que aparecen en el editor de datos o en la lista de variables en un cuadro de diálogo. Las variables pueden pertenecer a múltiples conjuntos (IBM, 2023).

Transformaciones

A partir de una o varias variables se pueden crear otras variables para el mejor análisis de los datos obtenidos. Los datos capturados no siempre pueden ser analizados directamente como fueron capturados, el análisis preliminar puede requerir de esquemas de codificación o transformación de las variables de estudio para poder ser procesadas correctamente. Se puede realizar agrupaciones de todo tipo de datos, desde variables sencillas, así como la agrupación de categorías para su análisis o la creación de nuevas variables que requieren ecuaciones más complejas (Pina Romero, 2020).

Cálculo de variables

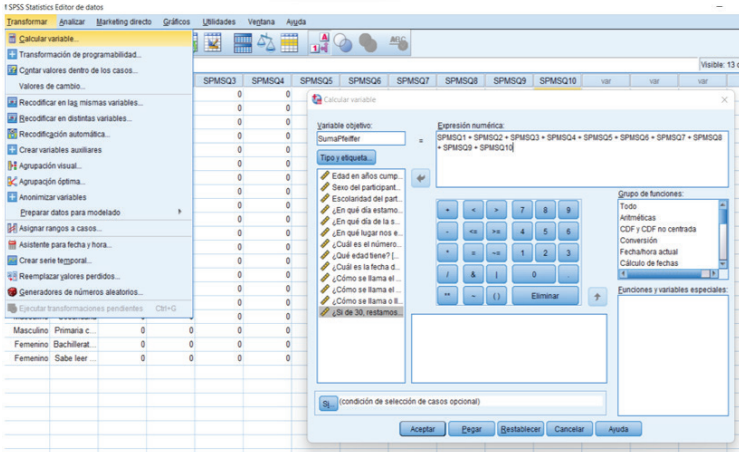
Se puede calcular variables numéricas o de cadena (alfanuméricas), se puede crear nuevas variables o bien remplazar los valores de las variables existentes, se pueden calcular los valores de forma selectiva para un subconjunto de datos, se pueden realizar operaciones aritméticas, funciones estadísticas, funciones de distribución y funciones de cadena de un conjunto de variables.

Para realizar un cálculo de variable se dirige al menú > Transformar > Calcular variable. Al realizar esta acción aparece un cuadro de dialogo y se escribe el nombre de una sola variable objetivo (puede ser una variable existente o una nueva variable que se desea añadir al conjunto de datos. En el cuadro de diálogo aparece la expresión numérica donde se ingresan las variables del conjunto de datos y lo que se desea hacer con ellas en la variable objetivo ya sea nueva o existente (operaciones sumas, restas, multiplicaciones, divisiones funciones estadísticas, etc.). al finalizar se oprime aceptar y será modificada la variable o se creará una nueva según sea el caso. Para crear una expresión puede pegar los componentes en el campo Expresión o escribir directamente en el campo. En el grupo de funciones con la etiqueta *Todo* contiene una lista de todas las funciones y variables de sistema disponibles.

Tomemos el ejemplo del instrumento de medición de Pfeiffer, una vez capturados los datos se tiene que realizar una sumatoria de todos los ítems del cuestionario para identificar el número de errores y así poder identificar el deterioro cognitivo. Para realizar la suma de las variables del instrumento se ingresa al menú > Transformar > Calcular variable y se le coloca el nombre "SumaPfeiffer" en la variable objetivo y en el cuadro de expresión numérica se ingresa la variable SPMSQ1 se le coloca el signo de "+" (porque se desea realizar una suma de un conjunto de variables) y se ingresa la segunda variable SPMSQ2 hasta completar el total de las variables a calcular, en este ejemplo se ingresan las 10 variables del cuestionario (Figura 11). En el cuadro de la expresión numérica se observa de la siguiente manera: SPMSQ1+ SPMSQ2+ SPMSQ3+ SPMSQ4+

SPMSQ5+ SPMSQ6+ SPMSQ7+ SPMSQ8+ SPMSQ9+ SPMSQ10 (Se recuerda que el SPMSQ es el nombre que se le dio a cada una de las variables del cuestionario de Pfeiffer y se puede observar en el ejemplo del libro de códigos).

Figura 11
Cálculo de variables SPSS



Cálculo de índices

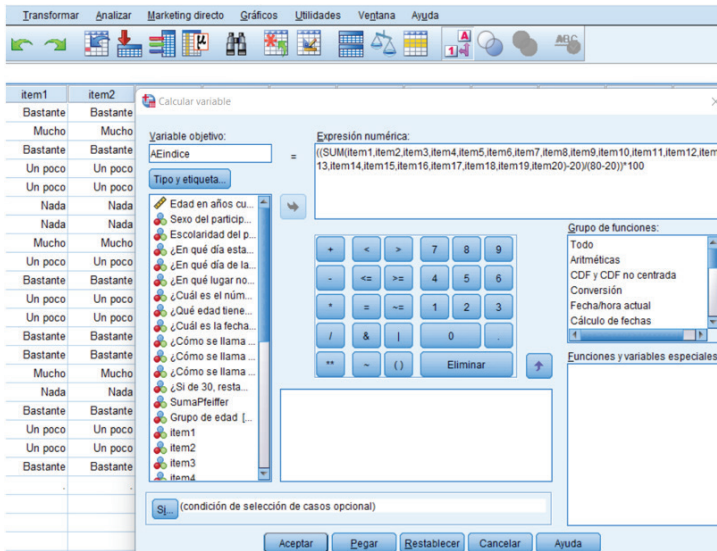
Un índice es una ponderación que puede de 0 a 100, se realiza con la finalidad de facilitar análisis estadísticos de instrumentos con gran número de ítems y que cuentan con una escala de respuesta de tipo Likert. Para calcular el índice de un instrumento se utiliza una operación aritmética, y se debe de seguir la misma ruta que la sumatoria. Se ingresa al menú > transformar > Calcular variable y aparece el cuadro de diálogo calcular variable, en la variable objetivo se ingresa el nombre de la nueva variable índice, en la expresión numérica se escribe ((sum(total de ítems del instrumento una por una separadas por una coma, por ejemplo, var1,var2,var3, etc.) - Valor mínimo del instrumento) / (Valor Máximo del instrumento-Valor Mínimo del instrumento))*100, oprimir aceptar y se habrá creado una variable de índice de un instrumento.

Para dar un ejemplo se tomará el Inventario de Ansiedad Rasgo-Estado (IDARE), cuyos autores son Charles D. Spielberger y su traducción al español es por Rogelio Díaz-Guerrero. El instrumento cuenta con 20 preguntas para medir la ansiedad estado y 20 para ansiedad rasgo, para este ejemplo solo utilizaremos ansiedad estado. Los índices se evalúan por tipo de ansiedad y refieren que, a mayor puntaje, mayor será la ansiedad que presente el entrevistado. La escala de medición es de tipo Likert donde 1= nada, 2= un poco, 3= bastante

y 4= mucho. Para poder transformar el instrumento a índices, se debe de identificar los valores mínimos y máximos del instrumento, para el inventario de ansiedad estado el valor mínimo es de 20 puntos (porque si un sujeto de estudio contesta 1 a todos los ítems) y el valor máximo es de 80 puntos (porque si un sujeto contesta 4 a todos los ítems).

Se ingresa al menú > transformar > Calcular variable y aparece el cuadro de diálogo calcular variable, en la variable objetivo se ingresa el nombre de la nueva en este caso se utilizó “AEindice” (al finalizar la operación se podrá ingresar en la vista de variables en el apartado de Etiqueta el nombre correspondiente a esta variable que es Índices de la Ansiedad Estado), en la expresión numérica se escribe ((sum(total de ítems del instrumento una por una separadas por una coma,) - Valor mínimo del instrumento) / (Valor Máximo del instrumento-Valor Mínimo del instrumento))*100, entonces se tiene que sustituir a los valores del instrumento y quedaría de la siguiente manera ((SUM(item1, item2, item3, item4, item5, item6, item7, item8, item9, item10, item11, item12, item13, item14, item15, item16, item17, item18, item19, item20)-20)/(80-20))*100 se oprime aceptar y aparecerá la nueva variable de índice de 0-100 de ansiedad estado (Figura 12).

Figura 12
Cálculo de variables SPSS



Cabe señalar que si algún valor no se encuentra bien ingresado por ejemplo, que falte una coma entre dos ítems o variables (Figura 13) o falta alguna expresión aritmética como los paréntesis (Figura 14).

Figura 13
Error en variables

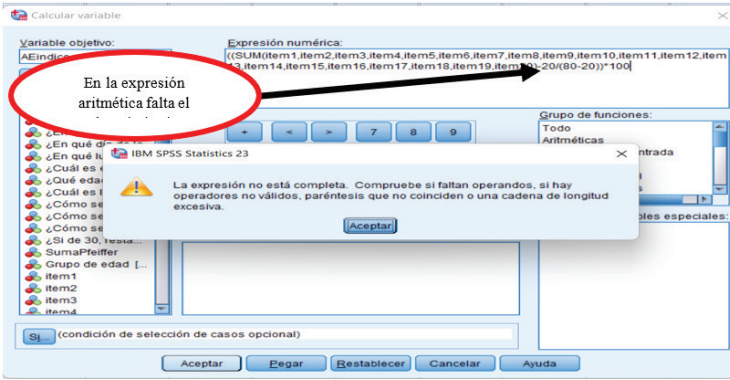
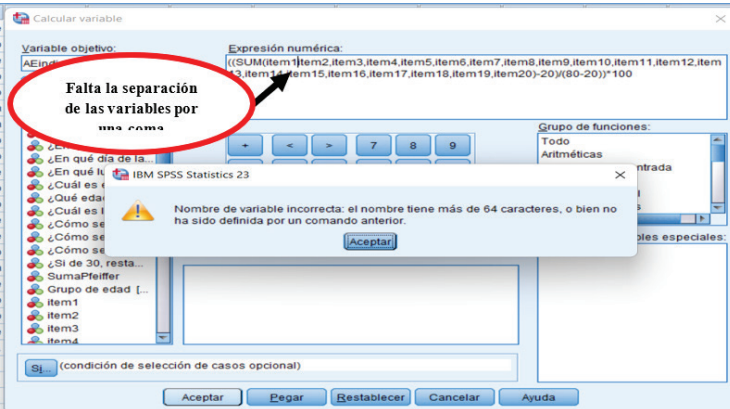


Figura 14
Error en la expresión aritmética



Recodificaciones

Puede cambiar los valores de los datos recodificando los datos. Esto se utiliza para agrupar o combinar categorías. Puede recodificar valores dentro de la misma variable existente o crear nuevas variables basadas en valores recodificados a partir de variables existentes. Si vuelve a codificar la misma variable, el valor de la variable existente se puede reasignar, pero el valor de origen se pierde y no se puede restaurar. Puede recodificar variables numéricas y de cadena. Si seleccio-

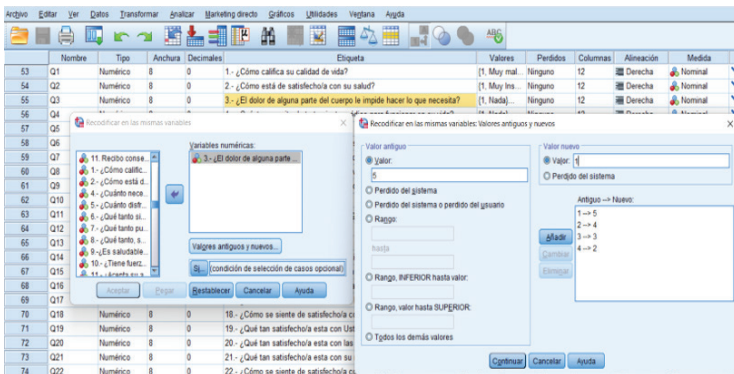
na varias variables, todas deben ser del mismo tipo. No se pueden recodificar variables numéricas y de cadena al mismo tiempo.

Recodificar en la misma variable

Para recodificar los valores de una variable se elige menú > Transformar > recodificar en las mismas variables y aparecerá un cuadro de diálogo de recodificar en la misma variable. En este apartado se encuentran del lado izquierdo las variables existentes en la base de datos y del derecho las variables numéricas, se selecciona la variable que se desea recodificar y se pasa al cuadro de variables numéricas, posterior se selecciona valores antiguos y nuevos y aparece un nuevo cuadro de diálogo Recodificar en las mismas variables: Valores antiguos y nuevos, en este apartado se deberá de ingresar el valor antiguo y el valor nuevo que desee modificar, oprime añadir si desea seguir agregando modificaciones a los valores en la variable a recodificar. Al finalizar del ingreso de cada uno de los valores (antiguos y nuevos) se oprime “Continuar” y regresará al cuadro de diálogo de recodificar en la misma variable y se pulsa aceptar, con esto ya se habrá modificado la variable que se ha elegido (Quezada Lucio, 2017).

Existen cuestionarios cuyas preguntas son evaluadas en orden inverso y tienen que recodificarse en orden contrapuesto para su tratamiento estadístico. Por ejemplo, una pregunta sobre la calidad de vida del instrumento Whoqol-Bref es ¿El dolor de alguna parte del cuerpo le impide hacer lo que necesita? Y las opciones de respuesta son: 1= nada, 2= un poco, 3= lo normal, 4= bastante y 5= extremadamente, el mismo instrumento en sus indicaciones menciona que esta pregunta se evalúa en orden inverso, por lo que cuando la respuesta sea 1(nada) su valor deberá de ser 5 (Extremadamente), por lo que esta variable puede ser recodificada para su análisis tanto individual como en conjunto (Figura 15).

Figura 15
Recodificar en la misma variable SPSS



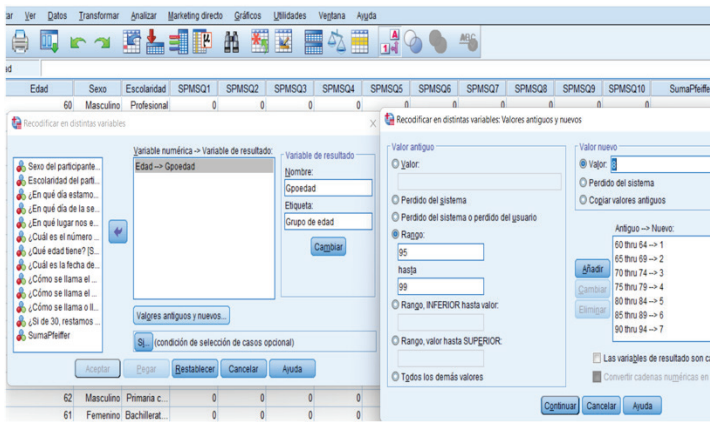
Recodificar en distinta variable

El cuadro de dialogo de recodificar en distinta variable le permite reasignar los valores de las variables existentes o agrupar un rango de valores existentes en nuevos valores para poder crear una variable nueva. Se pueden recodificar variables numéricas y de cadena, se pueden recodificar variables numéricas a variables de cadena y/o de cadena a numéricas, cuando se selecciona múltiples variables todas deberán ser del mismo tipo por lo que no se pueden recodificar conjuntamente las variables numéricas y de cadena.

Para recodificar los valores de una variable se elige menú > Transformar > recodificar en distinta variable y aparecerá un cuadro de diálogo de recodificar en distintas variables. Aparecerá del lado izquierdo las variables ya existentes, en el centro Variable numérica -> valor de resultado y en la parte derecha se encontrará variable de resultado y debajo las casillas de Nombre y etiqueta. Introduce el nombre de la nueva variable y la etiqueta correspondiente. Posterior selecciona valores antiguos y nuevos y aparecerá un nuevo cuadro de dialogo de Recodificar en distintas variables: Valores antiguos y nuevos, en este apartado se deberá ingresar el valor antiguo y el valor nuevo que se desea crear, pulsa añadir y continúa agregando valores que se deseen modificar en nueva variable. Al finalizar cada uno de los valores (antiguos y nuevos) oprime “Continuar” y regresará al cuadro de dialogo de recodificar en distinta variable, se pulsa “Cambiar” y luego pulsa aceptar, con esto ya se habrá creado una nueva variable con los datos correspondientes (IBM, 2021).

Un ejemplo de esta función es cuando tenemos la variable edad que es una variable cuantitativa numérica y deseamos identificar a los participantes por grupo de edad, seleccionamos menú > Transformar > recodificar en distinta variable y colocamos en la variable numérica -> variable de resultado la variable edad, en el nombre pondremos “Gpoedad” y en la casilla de etiqueta “Grupo de edad”, posterior se oprime valores antiguos y nuevos, en el cuadro de diálogo seleccionamos en valores antiguos > rangos y se comienza a añadir cada rango de edad por quinquenio empezando por 60 hasta 64 y en la casilla de valor nuevo colocamos 1 y se pulsa añadir, se repite la operación hasta completar los quinquenios de edad que se desean obtener en la nueva variable, en este ejemplo, se continuó el valor antiguo hasta 95-99, colocando en el valor nuevo 8. Se oprime en continuar, se pulsa cambiar y aceptar. Con esto ya se ha creado una nueva variable, ya solo faltaría modificar las etiquetas de valores para la nueva variable donde 1= 60-64 años, 2= 65-69 años, 3= 70-74 años, 4= 75-79 años, 5= 80-84 años, 6=85-89 años 7= 90-94 años y 8= 95-99 años. De esta manera se recodifica una variable numérica cuantitativa a una variable categórica sin perder ningún valor en la base de datos (Figura 16).

Figura 16
Recodificación en distinta variable SPSS



Referencias

- IBM (2019) *Guía del usuario de IBM SPSS Statistics 26 Core System*. IBM Corporation
- IBM (2020) *Guía del usuario del sistema básico de IBM SPSS Statistics 28*. IBM Corporation
- IBM (2021) *IBM SPSS Statistics 29 Core System User's Guide*
- IBM, *Statistical Package for the Social Sciences* (2023) *Defining variable sets*. disponible en: <https://www.ibm.com/docs/en/spss-statistics/29.0.0?topic=utilities-defining-variable-sets>.
- JUAN Antonio González González, J. A. (2009). *Manual Básico SPSS, Manual de introducción a SPSS*. Universidad de Talca, Centro de Inserción Laboral, Programa Jóvenes Profesionales.
- PINA Romero, J. A. (2020). *Manual Básico SPSS*. IBM SPSS. Departamento de Enfermería Comunitaria, Universidad de Alicante.
- QUEZADA Lucio, N. (2017). *Estadística con SPSS 24*. Ed. **MACRO**
- WILLIAM, E. & Wagner, III. (2019), *Using IBM SPSS statistics for research methods and social science statistics*. Ed. SAGE Publications, Inc Seventh Edition. California State. University Channel Islands.

Capítulo 3.

Análisis descriptivo, frecuencias absolutas y relativas. Medidas de tendencia central, medidas de dispersión y medidas de posición

DCEd TIRSO DURAN BADILLO

DCE JESÚS ALEJANDRO GUERRA-ORDOÑEZ

Introducción

La estadística descriptiva es una rama de la estadística que se centra en la recopilación, organización, resumen y presentación de datos de manera significativa. Su objetivo principal es describir y resumir las características principales de un conjunto de datos, proporcionando herramientas y técnicas para analizar y entender la información de manera más clara. La estadística descriptiva es esencial porque condensa datos complejos en medidas comprensibles, facilitando la identificación de patrones, la comparación de conjuntos, respaldando decisiones informadas y comunicando eficazmente la información a través de gráficos y visualizaciones (Rendón-Macías, Villasís-Keeve, & Miranda-Novales, 2016).

SPSS dispone de varios métodos estadísticos para analizar la información de forma descriptiva, algunos de los cuales son más útiles según el nivel de medición (cualitativa o cuantitativa) de cada variable. Es importante recalcar que el programa debe estar ejecutándose y el archivo debe estar abierto para realizar los pasos del análisis. Todos los procedimientos estadísticos de SPSS se pueden encontrar en el menú Analizar. Al hacer clic en él, aparecerá una lista desplegable con varios pasos para el programa, como por ejemplo: Ejemplos: informes, estadísticas descriptivas, tablas, comparaciones de promedios, etc. El submenú Estadísticas descriptivas incluye pasos para frecuencias, tablas descriptivas, de búsqueda, de contingencia y razones. Cada uno tiene un propósito específico y su uso depende del criterio y requerimientos analíticos del investigador.

Frecuencias Absolutas y Relativas

La frecuencia se define como el número de ocurrencias de un valor dentro de un conjunto de datos. En otras palabras, representa cuántas veces se repite un valor en una muestra estadística específica. Por ejemplo, si en una encuesta

cinco personas indican que tienen presión alta, entonces la frecuencia de hipertensión sería 5.

En estadística, comúnmente se utiliza la letra f con el subíndice i para denotar la frecuencia del valor i , y el símbolo de la frecuencia se representa como f_i . La suma de todas las frecuencias equivale al número total de datos en la muestra (Sánchez-Barajas, 2007). Por lo tanto, en cualquier estudio estadístico, siempre se cumple la siguiente fórmula:

$$\sum_{i=1}^N f_i = f_1 + f_2 + f_3 + \dots + f_N = N$$

Donde f_i es la frecuencia del valor i , y N es el número total de observaciones.

En estadística, los diferentes tipos de frecuencias son los siguientes, de las cuales haremos referencia a:

Frecuencia absoluta: consiste en el número de veces que aparece un valor en una muestra estadística.

Frecuencia relativa: es la frecuencia absoluta partido por el número total de datos.

Por ejemplo, en un estudio sobre los niveles de glucosa en sangre en una muestra de pacientes con diabetes, se examinaron y registraron los valores de glucosa en sangre de 50 pacientes. Los valores específicos se clasificaron en rangos, y los números de pacientes que cayeron en cada rango se consideran frecuencias absolutas observadas. Por otro lado, las frecuencias relativas se obtuvieron dividiendo el número de pacientes en cada rango por el total de pacientes observados (50).

Supongamos que los resultados fueron los siguientes:

10 pacientes tenían un nivel de glucosa en sangre entre 80 y 100 mg/dL.

20 pacientes tenían un nivel de glucosa en sangre entre 101 y 120 mg/dL.

15 pacientes tenían un nivel de glucosa en sangre entre 121 y 140 mg/dL.

5 pacientes tenían un nivel de glucosa en sangre superior a 140 mg/dL.

Estos valores representarían las frecuencias absolutas observadas. Las frecuencias relativas se calcularían dividiendo cada una de estas cifras por el total de pacientes observados (50). La suma de todas las frecuencias relativas sería igual a 1, lo que indica la totalidad de la muestra.

Ruta en SPSS: Analizar >>> estadísticos descriptivos >>> frecuencias >>> colocar la variable a analizar (preferentemente una variable categórica) en el recuadro para análisis >>> pulsar el botón de aceptar.

Estado civil

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Soltero	6	5.3	5.3	5.3
	Casado	72	63.2	63.2	68.4
	Divorciado	1	.9	.9	69.3
	Viudo	31	27.2	27.2	96.5
	Unión libre	4	3.5	3.5	100.0
	Total	114	100.0	100.0	

Nota. El sistema SPSS no incluye una opción para obtener frecuencias relativas, únicamente absolutas. Dado la poca utilidad de estas en ciencias sociales y de la salud. Sin embargo, pueden calcularse de manera independiente.

Medidas de Tendencia Central: Media, Moda y Mediana

Las medidas de tendencia central son indicadores que revelan el punto central alrededor del cual se agrupan los datos. Se emplean principalmente para resumir la información estadística. Estas medidas son esenciales en el análisis de datos, ya que ayudan a identificar el valor que representa el centro de un conjunto de datos, especialmente cuando se trata de variables con numerosos datos diversos. En tales casos, se prefiere calcular cantidades que resuman la información, como las medidas de tendencia central, también conocidas como medidas de centralización.

Estas medidas, como la media, mediana y moda, son parámetros estadísticos que señalan el centro de un conjunto de datos o muestras. Son herramientas ampliamente utilizadas, ya que condensar un conjunto de datos en un solo valor simplifica el análisis y proporciona una visión generalizada.

La media es la suma total de los valores dividida por el número de puntos de datos, representando el punto de equilibrio de la distribución, aunque puede ser influenciada por valores extremos.

La mediana es el valor central cuando los datos están ordenados, siendo el dato en el medio para conjuntos impares o el promedio de los dos valores centrales para conjuntos pares.

La moda, por otro lado, es la variable más frecuente en un conjunto de datos, pudiendo haber más de una moda en una muestra (Hernández, Fernández, & Baptista, 2010).

Ruta en SPSS: Analizar >>> estadísticos descriptivos >>> descriptivos >>> colocar la variable a analizar (debe ser una variable numérica) en el recuadro para análisis >>> pulsar el botón de aceptar.

Estadísticos descriptivos

	N	Mínimo	Máximo	Media	Desviación estándar
Edad del adulto mayor	114	60	93	72.22	8.119
N válido (por lista)	114				

Con la ruta anterior podemos obtener la media y los valores mínimos y máximos de la variable. Si se requiere obtener la mediana existe una ruta alterna:

Ruta en SPSS: Analizar >>> estadísticos descriptivos >>> explorar >>> colocar la variable a analizar (debe ser una variable numérica) en el recuadro para análisis “lista de dependientes” >>> pulsar el botón de aceptar.

Descriptivos

		Estadístico	Error estándar	
Edad del adulto mayor	Media	72.22	.760	
	95% de intervalo de confianza para la media	Límite inferior	70.71	
		Límite superior	73.73	
	Media recortada al 5%	71.82		
	Mediana	70.00		
	Varianza	65.925		
	Desviación estándar	8.119		
	Mínimo	60		
	Máximo	93		
	Rango	33		
	Rango intercuartil	12		
	Asimetría	.781	.226	
	Curtosis	-.252	.449	

Con esta ruta se puede obtener la media y la mediana, además de otros estadísticos. Para el caso particular de la moda, esta puede obtenerse por un análisis de frecuencias y verificar que valor se repite más veces.

Medidas de variabilidad

Las medidas de variabilidad señalan la dispersión de los datos en la escala de medición y abordan la cuestión de dónde se encuentran distribuidas las puntuaciones o los valores obtenidos. Mientras que las medidas de tendencia central son valores dentro de una distribución, las medidas de variabilidad representan intervalos que indican distancias o un número de unidades en la escala de medición. Entre las medidas de variabilidad más comúnmente empleadas se encuentran el rango, la desviación estándar y la varianza.

El rango, también conocido como recorrido, representa la diferencia entre la puntuación más alta y la más baja, indicando el número de unidades en la escala de medición necesarias para abarcar los valores máximo y mínimo. Su cálculo se realiza restando la puntuación menor de la mayor ($X_M - X_m$). Por ejemplo, si tenemos los valores: 17, 18, 20, 20, 24, 28, 28, 30, 33, el rango sería $33 - 17 = 16$. A mayor rango, mayor es la dispersión de los datos en la distribución.

La desviación estándar o desviación típica representa la desviación promedio de las puntuaciones de la media. Esta medida se expresa en las unidades originales de la distribución y se interpreta en relación con la media. Cuanto mayor es la dispersión de los datos alrededor de la media, mayor es la desviación estándar (también abreviada como SD), denotada por “s” o la letra griega sigma minúscula (σ). Por ejemplo, si un investigador obtiene una media de ingreso familiar anual de \$6,000 y una desviación estándar de \$1,000, esto significa que, en promedio, los ingresos familiares de la muestra difieren en mil unidades monetarias respecto a la media. Cabe destacar que la desviación estándar se aplica exclusivamente a variables medidas por intervalos o de razón.

La varianza, representada como s al cuadrado (s^2), es la desviación estándar elevada al cuadrado y constituye un concepto estadístico crucial, ya que muchas pruebas cuantitativas se basan en ella. Aunque con fines descriptivos se prefiere utilizar la desviación estándar, la varianza desempeña un papel importante en diversos métodos estadísticos, como la descomposición de la varianza (Rustom, 2012).

Ruta en SPSS: Analizar >>> estadísticos descriptivos >>> descriptivos >>> colocar la variable a analizar (debe ser una variable numérica) en el recuadro para análisis >>> seleccionar el botón de opciones >>> seleccionar las casillas que se desea que sean analizadas, en este caso las casillas de “Rango, Mínimo, Máximo, Varianza y Desviación Estándar” >>> pulsar el botón de continuar >>> pulsar el botón de aceptar.

Estadísticos descriptivos

	N	Rango	Mínimo	Máximo	Desviación estándar	Varianza
Funcionalidad Total 2	114	7	1	8	1.286	1.655
N válido (por lista)	114					

Medidas de Posición

Un cuartil es una medida de posición estadística que divide un conjunto de números en cuatro grupos de términos iguales. De manera similar, los deciles y los percentiles dividen una serie en partes iguales de 10 y 100, respectivamente.

Así como la mediana divide la serie en dos partes iguales, existen tres cuartiles, el noveno decil y el percentil 99, que dividen la distribución en 4, 10 y 100 partes iguales. Entre estas medidas de posición, los cuartiles son particularmente relevantes y se utilizan comúnmente en la identificación de estratos o grupos relacionados con fenómenos socioeconómicos, monetarios o teóricos.

Estos tres cuartiles se designan generalmente con los símbolos:

Q1 = primer cuartil

Q2 = segundo cuartil

Q3 = tercer cuartil

En cuanto a los deciles, se denotan como D1, D2, D3,..., D9, y los percentiles como P1, P2, P3, ..., P99. En cualquier caso, la medida posicional seleccionada toma el valor de uno de los términos o el punto medio de los dos términos. Para calcular estas medidas de posición, los términos deben ordenarse en orden ascendente o descendente. En el ordenamiento simple, el siguiente paso es determinar el “número de orden” del cuartil, decil o percentil, indicando qué posición ocupan en la distribución. Para los cuartiles, el número ordinal del primer cuartil es igual al número total de términos de la distribución más 1 dividido por 4. Para el segundo cuartil, el número ordinal se calcula sumando 1 al número total de términos y dividiendo por 2 (Posadas-Hernández, 2016).

Ruta de SPSS: analizar >>> estadísticos descriptivos >>> frecuencias >>> colocar la variable a analizar (debe ser una variable numérica) en el recuadro para análisis >>> seleccionar el botón de estadísticos >>> seleccionar las casillas que se desea que sean analizadas, en este caso las casillas de pueden ser la de “cuartiles” o puede ser la de “puntos de corte para” y dejar la cifra que viene por default (10) para los deciles, o la casilla de percentiles y colocar el número de percentil o percentiles deseados, estos pueden ser del 1 al 100, dependiendo de las necesidades del investigador >>> pulsar el botón de continuar >>> pulsar el botón de aceptar.

Cuartiles		
Estadísticos		
Funcionalidad Total 2		
N	Válido	114
	Perdidos	0
Percentiles	25	7.00
	50	8.00
	75	8.00

Deciles		
Estadísticos		
Funcionalidad Total 2		
N	Válido	114
	Perdidos	0
Percentiles	10	6.00
	20	7.00
	30	7.00
	40	8.00
	50	8.00
	60	8.00
	70	8.00
	80	8.00
	90	8.00

Percentiles
(1, 50 y 100)

Estadísticos

Funcionalidad Total 2

N	Válido	114
	Perdidos	0
Percentiles	1	1.15
	50	8.00
	100	8.00

Referencias

- HERNANDEZ Sampieri, R., Fernandez Collado, C., & Baptista Lucio, M. d. (2010). *Metodología de la Investigación*. México DF: Mc Grae Hill.
- POSDAS Hernandez, G. J. (2016). *Elementos básicos de estadística descriptiva para el análisis de datos*. Medellín: Fundación Universitaria Luis Amigo.
- RENDÓN-MACÍAS, M. E., Villasis-Keeve, M. Á., & Miranda-NOVALES, M. G. (2016). Estadística descriptiva. *Revista Alergia México*, 63(4), 397-407.
- RUSTOM J, A. (2012). *Estadística descriptiva, probabilidad e inferencia. Una visión conceptual y aplicada*. Santiago de Chile: Universidad de Chile.
- SÁNCHEZ Barajas, G. (Abril de 2007). *La estadística aplicada al análisis económico*. Obtenido de Sitio web de la Facultad de Economía de la Universidad Nacional Autónoma de México: <http://herzog.economia.unam.mx/profesor/barajas/estadis/indice.pdf>

Capítulo 4.

Representación gráfica de datos. Gráficas de barra simple, agrupados y segmentados. Gráficos de puntos Histogramas. Generación de plantillas de gráficos en SPSS

DCE CLAUDIA OROZCO GÓMEZ

DCE MARILY DANIELA AMARO HINOJOSA

Introducción

Las gráficas o gráficos permiten la representación de la información estadística a través de líneas o figuras que brindan una visión general de los hallazgos relacionados con las variables de estudio y resumen o desglosan la información que se encuentra en tablas estadísticas (Pérez et al., 2011). La mayoría de las gráficas tienen ejes (uno para los valores y otro para la variable) coordinados que usan líneas rectas ortogonales o perpendiculares que se interseccionan. El eje de valores puede mostrarse en forma horizontal o vertical y no necesariamente incluye el cero, ya que ello depende de la escala utilizada (Centro de Investigación y Desarrollo [CID], 2009). También, se pueden agrupar las columnas, barras, líneas y áreas de un gráfico, por lo que se pueden encontrar gráficos apilados (IBM Corporation, 2023).

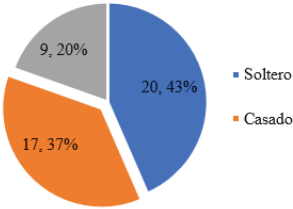
Al presentar una gráfica en un informe escrito la Asociación Americana de Psicología (2020) en la cuarta edición en español recomienda contemplar los siguientes elementos: colocar el número de la gráfica arriba de la tabla (en negrita), debajo de este el título (escrito en cursiva), este deberá describir claramente los valores numéricos representados en el gráfico. Enseguida, colocar la gráfica, incluir la clave o leyenda, la cual deberá colocarse dentro de los límites del gráfico. Debajo del gráfico escribir la nota (escrita en cursiva) donde se describirá parte del contenido que no pueda ser comprendido a partir del título y/o la leyenda solamente (Figura 1). También, puede explicar las unidades de medida, los símbolos o abreviaturas empleados.

Existen diferentes tipos de gráficos que permiten presentar los datos analizados de manera clara y precisa, para seleccionarlo adecuadamente se debe tomar en cuenta el número de variables que se quieren mostrar, cuántos puntos de datos serán visibles y como desea escalar el eje (Infogram, 2023). También es importante considerar que el gráfico que se seleccione debe ser acorde al tipo

de variable, es decir si es cuantitativa en las que se pueden tener datos discretos o continuos; si son variables cualitativas los cuales pueden ser datos ordinales y nominales (Ortega et al., 2019). Otro aspecto que tomar en cuenta es el objetivo que tiene la gráfica, el cual puede ser mostrar la distribución de las frecuencias ya sea de forma general, cambios en un periodo de tiempo, compararlas y relacionar las variables (CID, 2009; Infogram, 2023). Algunos de los gráficos más utilizados se describen a continuación:

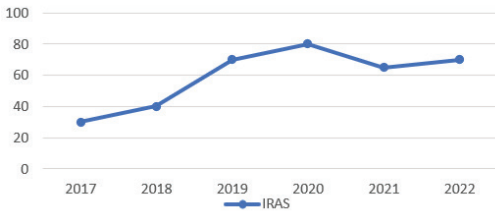
Gráficos circulares: Los gráficos circulares también llamados de sectores o pastel (Figura 1) resumen datos únicamente nominales; son útiles para resaltar proporciones y representar las partes en el conjunto o el todo dividido en sectores correspondientes a los diferentes componentes, en este gráfico se traza solo una serie de datos (CID, 2009; IBM Corporation, 2023; Quirós et al. 2019). En los gráficos circulares se recomienda disponer los sectores en orden según su tamaño, comenzando en la posición correspondiente a las 12 y en el sentido de las agujas del reloj (Bonita et al., 2008). Asimismo, dividir los sectores hasta en cinco o seis partes, si son más de estos, es mejor utilizar el gráfico de columnas apiladas.

Figura 1
Distribución del estado civil de los participantes



Gráficos de líneas: Los gráficos de líneas (ver Figura 2) son útiles para mostrar tendencias a lo largo del tiempo u observar la evolución de una variable en el tiempo relativas de un total y comparar varias series de datos en el tiempo. Generalmente se usan cuando hay series de datos con muchos puntos (IMB Corporation, 2023).

Figura 2
Número de casos de IRAS del 2017 al 2022



Gráficos de barras: Los gráficos de barras son útiles para resumir variables categóricas, por ejemplo, se puede utilizar un gráfico de barras para mostrar las frecuencias obtenidas de las variables: nivel socioeconómico, escolaridad, entre otras. También pueden emplearse para presentar variables cuantitativas discretas u ordinales, dichas variables pueden seguir un orden o jerarquía (Quirós, 2019), por lo que son útiles para clasificar valores de manera descendente o ascendente. También permiten mostrar la totalidad de un conjunto, tendencias a lo largo del tiempo, comparar la frecuencia entre las categorías y trazar varias series de datos (IBM Corporation, 2023).

En el gráfico de barras las categorías se pueden situar de manera vertical u horizontal, esto depende del número de categorías, se prefiere que se configure de manera horizontal cuando hay muchas categorías o los nombres son demasiado largos (CDI, 2009). La altura del rectángulo es proporcional al número de observaciones en cada categoría, estos se encuentran separados entre sí (Martínez, et al., 2020).

Gráfico de barras superpuestas: Los gráficos de barras superpuestas o apilados se usan cuando lo que se busca es comparar la magnitud proporcional dentro de una categoría, permiten estudiar de forma independiente varias series de datos y compararlas mediante barras superpuestas (IBM Corporation, 2023). Los gráficos rectangulares, permiten analizar la participación porcentual de los elementos respecto a un total. Pueden presentarse de forma vertical u horizontal mostrando cifras absolutas o relativas (CDI, 2009).

Gráfica de barras agrupadas: Las gráficas de barra agrupadas se se utiliza para representar valores individuales para más de un elemento que comparten la misma categoría. Se encuentran en el mismo eje y se comparan las variables o categorías agrupadas a otros grupos de esas mismas variables o tipos de categoría (IBM Corporation, 2021).

Figura 3
Gráfica de barras simple
Clasificación del nivel del dolor

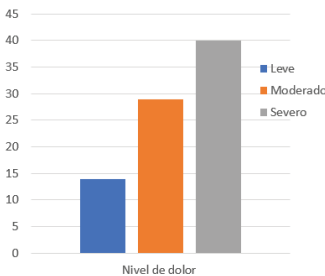


Figura 4
Gráfica de barras superpuesta
Patologías por sexo

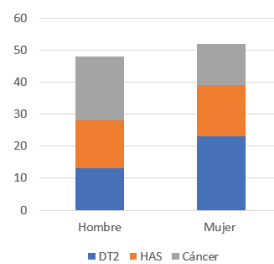
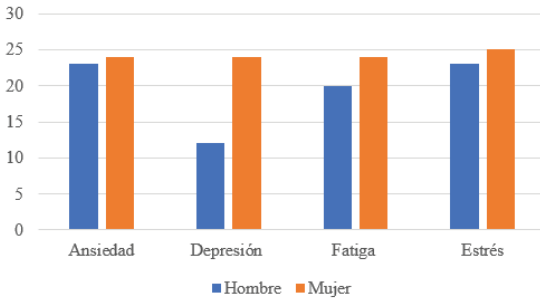
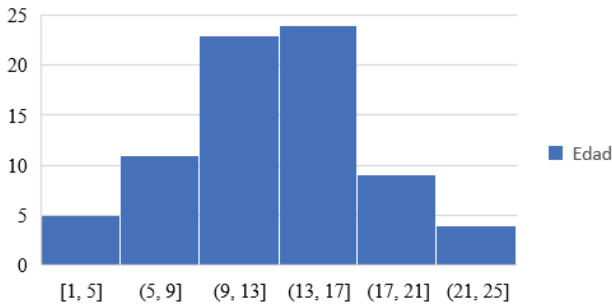


Figura 5
Gráfica de barras agrupada/Trastornos mentales por sexo

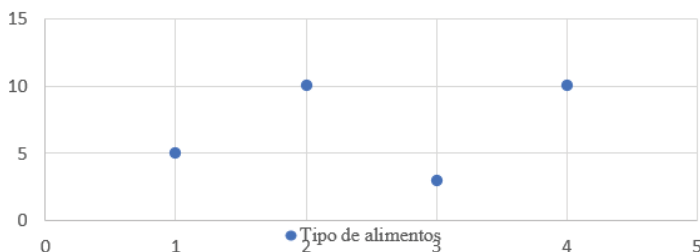


Histogramas: El histograma es un gráfico que se utiliza con variable continua o discreta, representa la distribución de las frecuencias u observaciones agrupadas en una escala o eje horizontal (los intervalos o las marcas de clase que son los puntos medios de cada intervalo) y las frecuencias de clase en una escala o eje vertical y trazando rectángulos. A diferencia del gráfico de barras en este no existe separación entre las barras (Cornejo et al., 2019; IBM Corporation, 2023).

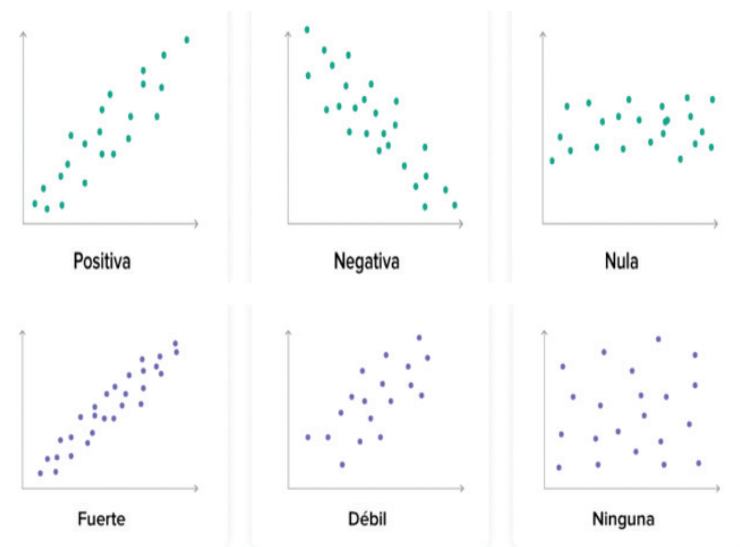
Figura 6
Edad de las mujeres



Gráficos de puntos: Los gráficos de puntos son útiles para mostrar datos cuantitativos de una forma organizada. En este tipo de gráficos se usan varios puntos para trazar datos a lo largo de un eje ordinal (eje X). En este gráfico solo se dibuja o presenta punto, sin unirlos con una línea, son útiles para mostrar la distribución de una única variable de escala. Los datos se agrupan, pero, en vez de un valor para cada intervalo (como un recuento), se muestran y apilan todos los puntos en cada intervalo (IBM Corporation, 2023).

Figura 7*Gráfico de puntos: Consumo de alimentos a la semana*

Gráficos de dispersión: Los gráficos de dispersión utilizan puntos de datos para trazar dos medidas en cualquier lugar de una escala, no sólo con marcas de escala regulares. Se utiliza cuando hay muchos puntos de datos diferentes y se desea destacar las similitudes en el conjunto de datos. Esto es útil cuando se buscan valores atípicos o para entender la distribución de los datos. Además, permite encontrar la existencia de una correlación o relación entre variables cuantitativas. En la figura 8 se puede observar que la relación puede ser positiva, negativa o nula, esta se expresa de acuerdo con la dirección de los puntos y la fuerza de la correlación se puede determinar de acuerdo a la dispersión entre los puntos.

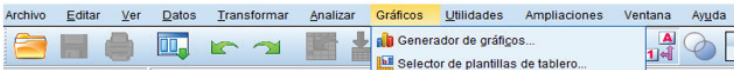
Figura 8*Gráfico de dispersión*

Generación de plantillas y gráficos en SPSS

Para crear un gráfico, en SPSS de acuerdo con el Manual del usuario del sistema básico de IBM SPSS Statistics (2011), es necesario seguir las instrucciones que aparecen a continuación.

Pulse el icono en la barra del menú **Gráficos** y posteriormente **Generador de gráficos** (ver Figura 9).

Figura 9
Opciones de Menú: Gráficos



Se abrirá el cuadro de diálogo **Generador de gráficos** (ver Figura 10).

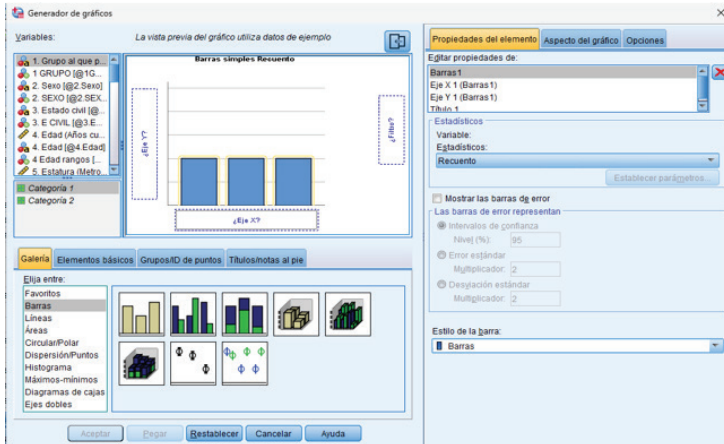
Figura 10
Cuadro de diálogo: generador de gráficos



La estrategia más sencilla para generar gráficos es a través de la galería. Enseñada, se describen los pasos generales que se recomienda seguir para generar el gráfico por este medio.

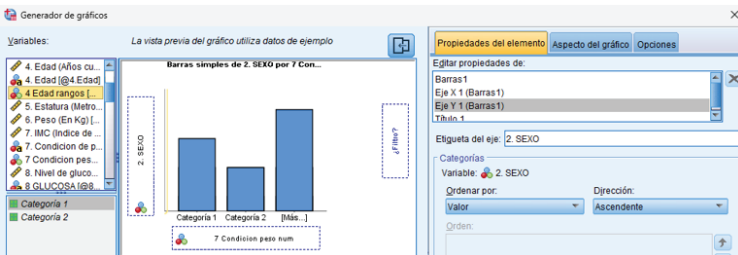
- Si no, haga clic en la pestaña **Galería**
- Seleccione una categoría de gráficos de la lista **Seleccionar**. Cada categoría tiene varios tipos, y elegiremos **Barras**
- Arrastre la imagen del gráfico que desee al lienzo (consulte la Figura 11). También puedes hacer doble clic en la imagen.
- Si el gráfico ya se muestra en el lienzo, el gráfico de galería reemplaza los ejes del gráfico y el conjunto de elementos gráficos.

Figura 11
Cuadro de diálogo, después de arrastrar la imagen de barra simple



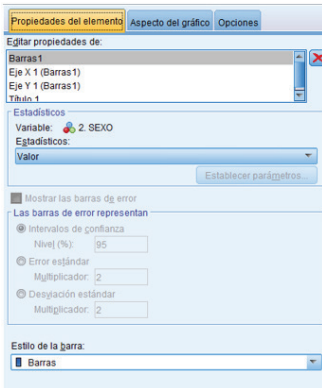
Arrastre las variables desde la lista de variables y suéltelas en la zona de colocación del eje y, si está disponible, en la zona de colocación de agrupación (consulte la Figura 12). Si la estadística que desea utilizar ya se muestra en la zona de colocación del eje, no necesita arrastrar la variable a la zona de colocación. Agregue variables a una zona solo si el texto de la zona es azul. Si el texto es negro, la zona ya contiene variables o estadísticas.

Figura 12
Cuadro de diálogo con zonas de colocación rellenas



Es importante tener en cuenta que el generador de gráficos establece configuraciones predeterminadas según el nivel de medición durante la creación del gráfico. Además, los gráficos resultantes pueden verse diferentes en diferentes niveles de medición. Puede cambiar temporalmente el nivel de medición de una variable haciendo clic derecho en la variable y seleccionando Opciones. Si necesita cambiar estadísticas o cambiar algún eje o atributo de leyenda (como el ancho de escala), haga clic en Propiedades del elemento (ver Figura 13).

Figura 13
Cuadro de diálogo: Propiedades del elemento



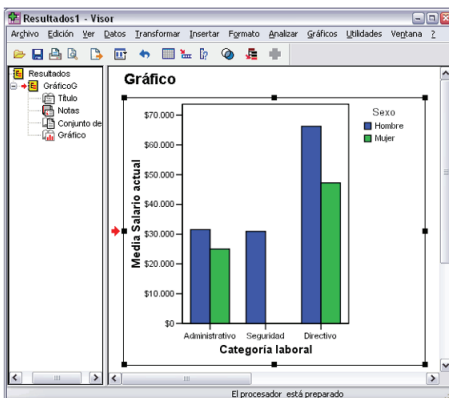
En la lista **Editar** propiedades de, seleccione el elemento que desea cambiar. (Si desea obtener información acerca de propiedades específicas, pulse en Ayuda.)

Una vez realizados los cambios, pulse en Aplicar.

Si necesita agregar más variables al gráfico (por ejemplo, para la agrupación o la adición de paneles), pulse en la pestaña Grupos/ID de puntos del cuadro de diálogo Generador de gráficos y seleccione una o más opciones. A continuación, arrastre las variables categóricas a las nuevas zonas de colocación que aparecen en el lienzo.

Si desea transponer el gráfico (por ejemplo, para que las barras sean horizontales), pulse en la pestaña Elementos básicos y, a continuación, pulse en Transponer. Pulse en Aceptar para crear el gráfico. Aparecerá el gráfico en el Visor (ver Figura 14).

Figura 14
Gráfico de barras en la ventana de Resultados



Enseguida se presenta una ruta rápida para elaborar otros gráficos:

Histograma

1. Haga clic en Gráficos > Cuadro de diálogo antiguos > Histograma
2. Arrastre la variable que desea trazar como histograma desde la izquierda al cuadro de texto variable
3. Seleccione Mostrar curva normal (recomendado)
4. Haga clic en aceptar
5. El histograma aparecerá en el visor de salida de SPSS

Gráfico de dispersión

1. Hacer clic en gráficos >cuadros de diálogo antiguos>dispersión/puntos>Dispersión simple>definir
2. Arrastre la variable de la escala al eje X>arrastre variable o variables de escala a la zona de colocación del eje Y>
3. Hacer clic en aceptar
4. El gráfico de dispersión aparecerá en el visor de salida de SPSS

Gráfica circular

1. Hacer clic en gráficos >cuadros de diálogo antiguos>circular>resúmenes para grupos de casos>definir
2. Arrastre variable a definir porciones por
3. Hacer clic en aceptar
4. El gráfico circular aparecerá en el visor de salida de SPSS

Referencias

- ASOCIACIÓN Americana de Psicología. (2020). Manual de publicaciones de la American Psychological Association (4ª ed.) (T. F. Oriol y R., Gallegos). Manual Moderno.
- BONITA, R., Beaglehole, R., & Kjellström, T. (2008). Epidemiología básica. Organización Panamericana de la Salud.
- CENTRO de Investigación y Desarrollo. (2009). *Guía para la prestación de gráficos estadísticos*. Instituto Nacional de Estadística e Informática. <https://www.inei.gov.pe/media/MenuRecursivo/metodologias/libro.pdf>
- CORNEJO, A. M., García, M. A., Quirós, G. V., Rodríguez, L. L., & Ordaz, J. G. (2019). *Metodología básica de la investigación clínica*. Grupo CTO.
- IBM Corporation. (2023). *Gráficos*. <https://www.ibm.com/docs/es/cognos-analytics/11.1.0?topic=reporting-charts>

- IBM Corporation. (2011). *Manual del usuario del sistema básico de IBM SPSS Statistics 20*. http://www.tese.edu.mx/documentos2004/7821_XOEQIKG.pdf
- IBM Corporation. (2021). Valores de los datos de los gráficos de barras agrupadas. ibm.com/docs/es/jfsm/1.1.2.1?topic=SSEKCU_1.1.2.1/com.ibm.psc.doc/tip_original/dash_r_dv_clustered_bar_charts.htm
- INFOGRAM. (2023). Cómo elegir el gráfico correcto para tus datos. <https://info-gram.com/es/pagina/elige-el-grafico-correcto-visualizacion-datos>
- ORTEGA, P. E. Ochoa, S. C. Molina, A. M. (2019). Representación gráfica de variables. *Evidencias en Pediatría*, 15(13), 1-7. <https://evidenciasenpediatria.es/articulo/7392/representacion-grafica-de-variables->
- TUDASHBOARD. (2021). Gráfica de dispersión. Qué es y cuáles son sus características.

Capítulo 5.

Tablas de contingencia

DCE DANIEL SIFUENTES LEURA

Introducción

También conocida como tablas dinámicas, de control o *crosstabs*, tablas cruzadas, tablas de 2×2 se utiliza en estadística para representar datos categóricos en términos de conteo de frecuencias donde se crean al menos 2 filas y 2 columnas, en análisis de datos es posiblemente en estadística la técnica más utilizada.

Cómo se analizan tablas de contingencia

Una tabla cruzada se basa en cálculo de porcentajes y la forman 2 variables. El objetivo de la tabla de contingencia es checar si hay relación entre estas 2 variables y la forma de detectarlo es viendo la distribución de porcentajes.

En concreto, el objetivo es analizar si la distribución porcentual de una variable se repite uniformemente dentro de categorías de otras variables. Dado que una variable se estudia en función de otra variable, el investigador debe tener en cuenta tanto las variables dependientes como las independientes (variables dependientes y explicativas). Esto se debe a que las variables independientes están organizadas en columnas y las variables dependientes en filas.

Las condiciones para hacer un análisis de tablas de contingencia

Como estrategia para analizar los datos hay dos condiciones:

- Solo se pueden relacionar dos o más variables.
- Las variables para analizar deben ser nominales u ordinales. Las nominales no tienen orden interno establecido, y las ordinales sí tienen un orden interno establecido y el paso de una categoría a otra no es igual (nivel educativo alto, medio, bajo). En estas tablas no se recomienda usar variables escalares como la edad por tener múltiples categorías sería muy extensa e ilegible. Si se deseara usar la edad se deberá recodificarla por rangos, al realizar esto se convierte en variable ordinal por lo tanto se puede analizar en dichas tablas.

Tabla de contingencia interés en la política*grupos de edad

	Grupo edad			
	-35	36 - 64	+ 65	Total
Interés en la política mucho	53 10%	225 21.9 %	148 30.5%	426 20.8 %
Algo	218 41.1%	445 43.3%	188 38.7%	851 41.6%
Poco	160 30.2%	278 27%	130 26.7%	568 27.8%
Nada	99 18.7%	80 7,8%	20 4.1%	199 9.7%
Total	530 100%	1028 100%	486 100%	2044 100%

Debemos considerar la hipótesis de partida para analizar tablas de contingencia, para comprobar la hipótesis donde se presenta relación entre 2 variables.

Veamos un ejemplo, quiero ver si la “creencia en Dios” explica el interés en la política. La hipótesis de partida es que las personas que tienen “creencia en Dios” tienen más “interés en la política”. Por tanto, tengo dos variables a relacionar: “interés en la política” y “la creencia en Dios”.

- La variable “interés en la política” respuesta: (1-mucho, 2-algo, 3-poco, 4-nada).
- La variable “creencia en Dios” dicotómica en dos categorías: (1-sí, 2-no).

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Interés en la política mucho	426	20.8	20.8	20.8
Algo	852	41.6	41.6	62.5
Poco	568	27.8	27.8	90.2
Nada	199	9.7	9.8	100
Total	2045	99.9	100	

Primeramente, se presentan las frecuencias de las variables por separado sin realizar la cruz ni la relación.

Tabla de frecuencias interés en la política

Los datos mencionan que el 41,6% de los encuestados en España refieren estar algo interesados en la política. El 27,8 dicen estar poco interesados. El 20,8 están muy interesados. Solo un 9,8% dicen no estar nada interesados en la política.

Tabla creencia en Dios

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Si	1286	62.9	65.1	65.1
No	690	33.7	34.9	100
Total	1976	96.6	100	

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
No contestaron	32	.2		
No saben	34	1.5		
Datos perdidos	4	1.7		
Total	70	3.4		
Total	2046	100		

De la mayoría de los encuestados en España, el 65,1%, sí “creen en Dios”. Mientras que el 34,9% menciona no creen en Dios.

Posteriormente se relacionar estas dos variables.

Analizando la situación la variable independiente sería la “creencia en Dios” y la variable dependiente el “interés en la política” y checando la hipótesis inicial es que la “creencia en Dios” lo que explica el “interés en la política”

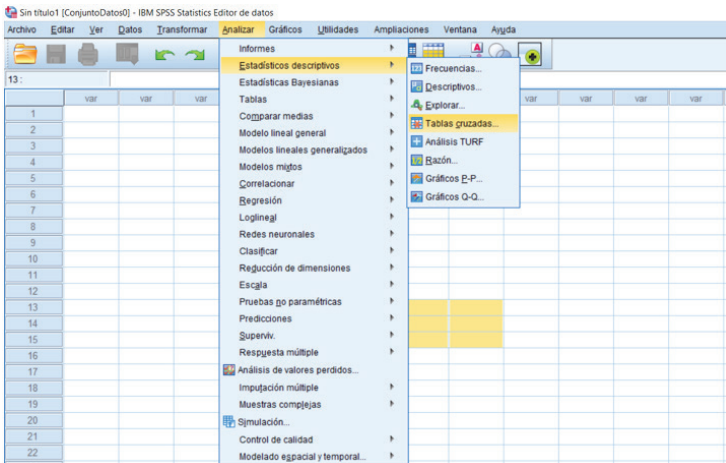
Tomando en cuenta lo anterior la variable “creencia en Dios” se presentará en columnas y la variable “interés en la política” en filas

Depende del investigador el acomodo de las variables en filas y en las columnas tomado en cuenta su hipótesis.

Construcción de las tablas de contingencia

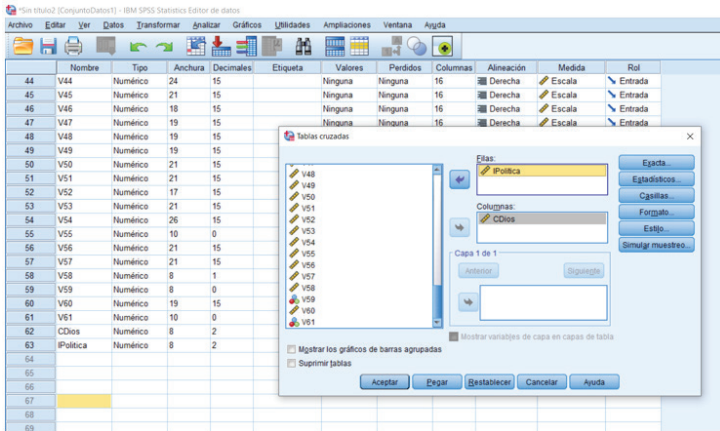
Analizamos en SPSS el procedimiento para realizar una tabla de contingencia con la captura de pantalla.

Seguimos este orden: Analizar > Estadísticos Descriptivos > Tablas de contingencia.

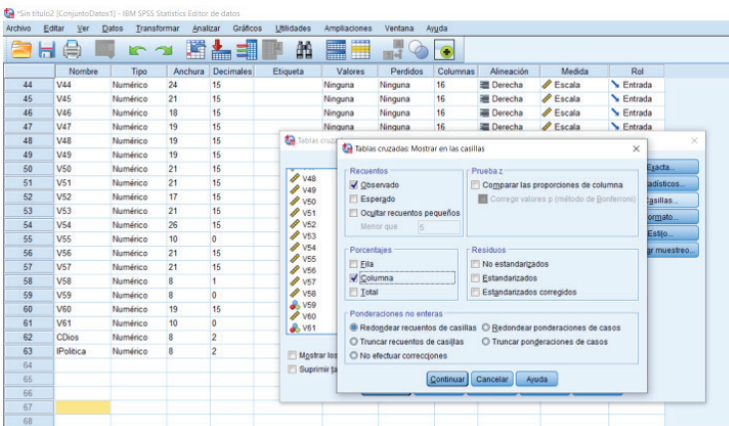


Después se acomodan las variables en filas (variable dependiente) y columnas (variable independiente)

Tablas de contingencia



Posteriormente, para sacar los “porcentajes por columnas” nos vamos donde pone “Casilla” y activamos la opción de “Porcentajes por columnas”. activamos, Continuar y Aceptar



Este es el resultado. La tabla de contingencia:

Tabla de contingencia interés en la Política * Creencia en Dios

		Cree en Dios		
Interés en la política		Si	No	Total
mucho	% dentro Cree en Dios	283	122	405
		22%	17.7%	20.5%
Algo	% dentro Cree en Dios	528	306	834
		41.1%	44.3%	42.2%

		Cree en Dios		
Interés en la política		Si	No	Total
Poco	% dentro Cree en Dios	356 27.7%	190 27.5%	546 27.6%
Nada	% dentro Cree en Dios	119 9.3%	73 10.6%	192 9.7%
Total	% dentro Cree en Dios	1286 100%	691 100%	1977 100%

Al relacionar las variables los porcentos se calculan por columnas es importante mencionar que el programa donde estas procesando tus tablas te procesa estos porcentajes.

Para mejor comprensión del proceso y el resultado se explica cómo se realiza el cálculo de dichos porcentos.

- En el primer recuadro tenemos que 283 que resulta del cruce de las variables “si ser creyentes en Dios” además de referir “mucho interés en la política” es la suma de las 2 variables.
- Analizando por columnas se calcula el porciento dividiendo los casos que, si son creyentes y tienen mucho interés en la política entre el total de casos que, si son creyentes, y lo multiplicamos por 100, obteniendo un porcentaje $(283 / 1286) * 100 = 22,0\%$
- Posteriormente se promedian las demás columnas el número de casos que no son creyentes y refieren mucho interés en la política entre el total de casos que no son creyentes $(122 / 691) * 100 = 17,7\%$. Después se procesan las demás columnas.

Una vez expresada en porcentaje la tabla de contingencia que relaciona las dos variables, se puede continuar leyendo. Si no se calculan los porcentajes, la tabla no se puede leer correctamente porque el número de casos en cada columna no es el mismo. Para analizar una tabla de contingencia, es necesario mostrar porcentajes. Una tabla de contingencia sin porcentajes está incompleta.

Análisis de una tabla de contingencia

Estas tablas se interpretan en porcentajes se leen de derecha a izquierda y fila por fila.

Leemos la primera fila y vemos si varían los porcentos entre las categorías de la variable independiente (columnas). Entre más diferencia en porcentos exista en cada fila la relación es más significativa entre las variables.

En el caso de que los porcentos sean semejantes o no haya diferencia significativa en la misma fila, no existe relación entre variables.

Si los porcentos son diferentes hay una significancia en la relación entre las variables.

Una variable explica a la otra si estas se relacionan entre sí, si presenta diferencia significativa de las respuestas entre las categorías que presenta una variable.

Analicemos como leer la tabla con el presente ejemplo: tenemos que el 20,5% de los sujetos encuestados en España refieren estar muy interesados en la política. Este porcentaje es más grande en los sujetos creyentes en Dios (22% comparado con el 17,7% de los que no son creyentes). Mencionan estar algo interesados el 42,2% de los españoles y el valor del porcentaje es más grande en los no creyentes en Dios con un 44,3%. Con relación a los que refieren poco interés en la política el 27,6 %, comparación semejante para los creyentes y no creyentes en Dios (27,7% y 27,5 %). Por último, el 9,7% de los sujetos participantes refirieron no estar nada interesados en la política considerando este porcentaje con un 9,3% creyentes en Dios resultados más grande en los no creyentes en Dios con el 10,6%. Analizando los resultados de la tabla el creer en Dios si explica el interés en la política presentando una relación débil ya que en la variable de los muy interesados en la política son medianamente considerables viendo las diferencias entre creyentes y no creyentes en Dios. Checando las demás categorías en relación con el interés en la política los resultados son muy semejantes o inexistentes entre los que creen o no creen en Dios. Analizando los resultados nos damos cuenta de que las cuestiones de religión no dependen el interés en la política y que el debate político no influye en las cuestiones religiosas. En futuras mediciones se deberán analizar otras variables como la educación o el ingreso económico para analizar el impacto.

Se recomienda analizar con otros estadísticos para ver la relación y con que fuerza como lo son la Chi-cuadrada, V de Cramer etc.

Las tablas de contingencia permiten la relación entre variables ordinales y nominales.

No permite relacionar variables escalares al menos que se recodifiquen por rangos.

Se tienen que ordenar en filas (variable dependiente la que quiero explicar) y columnas (variable independiente explicativa).

La lectura se realiza de derecha a izquierda y por filas.

Son una técnica de análisis bivariado relaciona 2 variables trata de averiguar si una explica a la otra.

Se calcula el porcentaje por columna para que se pueda leer.

La importancia radica en averiguar si los porcentajes de las filas presentan diferencias significativas (poco, Mucho o nada) con relación a las columnas, si presentan mucha diferencia en el porcentaje las 2 variables están relacionadas quiere decir que una variable explica a la otra. Si no presentan diferencias o son semejantes no existe relación. Cuando la diferencia es mínima o se presenta en solo algunas categorías la relación es débil.

Referencias

- ARRIAZA, B.M. (s.f.). Tipos de variables y alternativas de análisis. En IFAPA (Ed.), *Guía práctica de análisis de datos* (pp. 63-65). https://scholar.google.es/scholar?hl=es&cas_sdt=0,5&q=guía+práctica+de+análisis+de+datos+Manuel+Arraiza+Balmón
- HAIR, F. J., Black, C. W., Babin, J.B. & Anderson, E.R. (2013). Multidimensional scaling. Analysing Nominal Data with correspondence Analysis. En *Pearson Education Limited* (Ed.). *Multivariate Data-Analysis* (7ª. ed., p. 519). Impreso https://books.google.com.mx/books/about/Multivariate_Data_Analysis.html?id=VvXZnQEACAAJ&redir_esc=y
- TABLA de contingencia (10 mayo 2023). En *Wikipedia*. https://es.wikipedia.org/wiki/Tabla_de_contingencia
- IBM, 2017. Tablas cruzadas. Actualización 28 junio 2022. <https://www.ibm.com/docs/es/spss-statistics/saas?topic=tests-chi-square-test>
- SMARTLINE (21 julio 2021). Prueba Chi-cuadrado, V-Cramer, y tablas cruzadas en SPSS. [Video]. YouTube. https://www.youtube.com/watch?v=eo_YbULtWdQ&t=329s

Capítulo 6.

Chi cuadrada

DCE MARÍA MAGDALENA DELABRA SALINAS
DR. ERNESTO RIOS WILLARS

Introducción

La prueba Chi-Cuadrado (X^2) es una prueba no paramétrica o de libre distribución, se usa generalmente para buscar la relación entre dos variables categóricas o nominales calcula el estadístico Chi (Arriaza, M., s.f.). Su creación y desarrollo se debe a Karl Pearson en 1900 y sus trabajos son considerados pilares fundamentales de la estadística moderna (Cochan, William, 1952).

El valor de chi-cuadrada es una medida de asociación entre las categorías de filas y columnas; se basa en las frecuencias esperadas y observadas en una tabla de contingencia. Las frecuencias esperadas están basadas en las probabilidades marginales de sus filas y columnas (probabilidad de una fila y columna entre todas las filas y columnas) (Hair, Black, Babin, Anderson, 2013).

Usos de la prueba Chi-cuadrada

La prueba Chi-Cuadrada, al ser la suma de todos los valores elevados al cuadrado no puede obtener valores negativos por o que va del cero al infinito (Mendivelso, Rodriguez, 2018; McHugh, 2013). Tiene tres usos principales: a) Prueba de bondad de ajuste (una variable); b) Prueba de independencia (dos variables); c) Prueba de homogeneidad (dos variables) (Betensky, Rabiowitz, 1999; Mendivelso, Rodriguez, 2018).

- a. Prueba de bondad de ajuste (una variable): es un contraste de significación para cuando se quiere saber si los datos de la muestra se apegan a una ley de distribución teórica que se considera como correcta (Saldaña, 2011).
- b. Prueba de independencia (dos variables): Es un contraste para datos cualitativos o categóricos. Para saber si hay dependencia o independencia de los valores Es una prueba de tablas de contingencia (Saldaña, 2011).
- c. Prueba de homogeneidad (dos variables): se usa para contrastar la igualdad de procedencia de un conjunto de muestras de tipo cualitativo o categóricos (Saldaña, 2011).

Para efectos de este capítulo, se prestará atención a la prueba de independencia (opción b). En el que se abordarán ejemplos para determinar si existe o no relación entre dos variables cualitativas.

Supuestos de la prueba Chi-cuadrada

El cálculo de la prueba χ^2 se realiza a partir del uso de datos de variables categóricas numéricas ordenadas o no, cuyos niveles de medición pueden ser ordinal o nominal (IBM, 2017)

La prueba χ^2 es no paramétrica, por lo que no requiere supuestos para su aplicación. Sin embargo, es necesario considerar:

1. Se asume que los datos son una muestra aleatoria.
2. Las frecuencias esperadas por cada categoría deberán ser 1 como mínimo.
3. No más de un 20% de las categorías deberán tener frecuencias esperadas menores que 5

Ejemplo prueba Chi-cuadrada en SPSS

Supongamos que queremos saber si existe relación o asociación entre el nivel de depresión y el sexo en 405 adultos mayores de comunidades rurales. Se procede a plantear las hipótesis de asociación:

H0: Las variables nivel de depresión y el sexo son independientes.

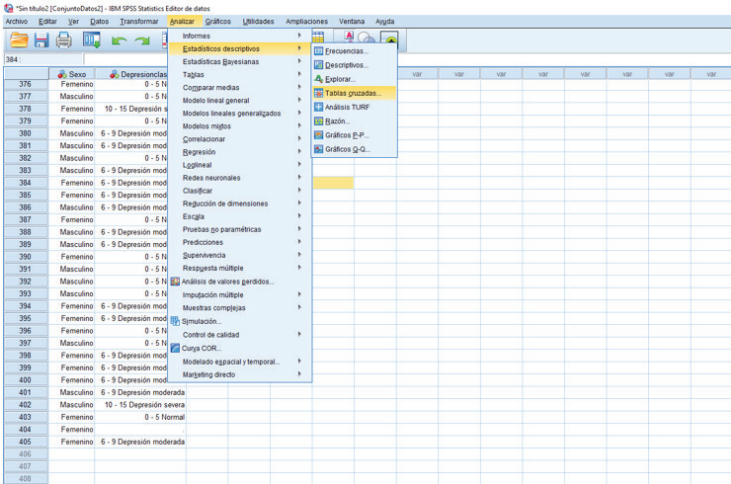
H1: Las variables nivel de depresión y el sexo son relacionadas (asociadas o dependientes)

El nivel de significancia que se tomará es 0.05, con un margen de error del 0.05.

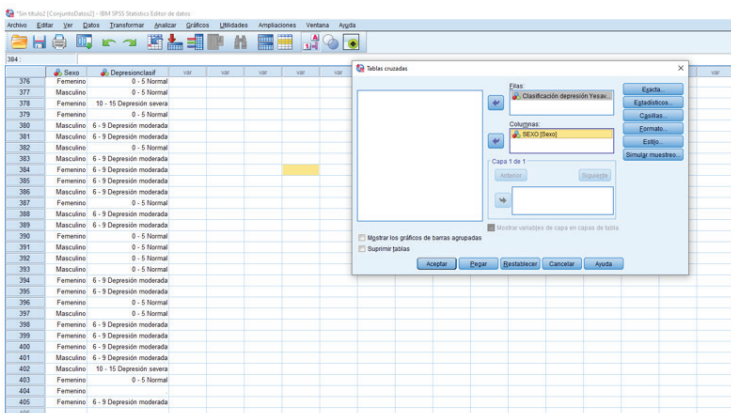
Pasos de procesamiento en SPSS

(puede auxiliarse del video de SmartLine (2021))

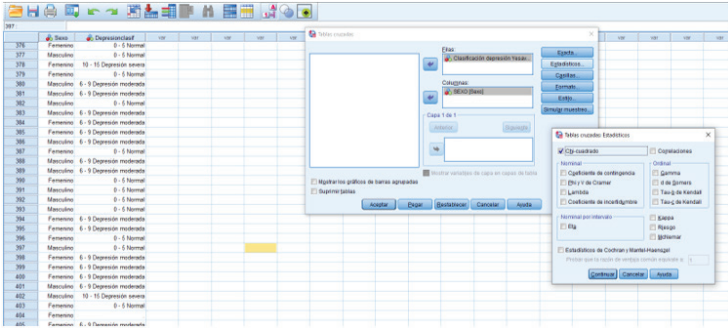
1. Una vez identificadas las variables a asociar, que son sexo y nivel de depresión, se pasa al apartado de “Análizar”, se selecciona “Estadísticos descriptivos”, “tablas cruzadas”, se hace click.



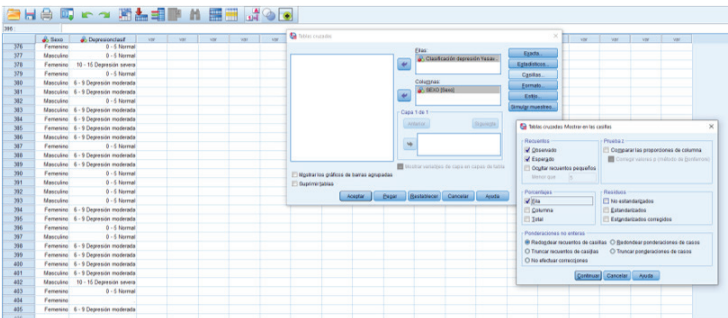
- Se seleccionan las variables para acomodarlas en filas y columnas. Algunos criterios de decisión para acomodar las variables son: a) acomodar en columnas la variable ordinal y en las filas, la variable nominal; o b) la variable de interés, la que se pretende analizar va en columnas.



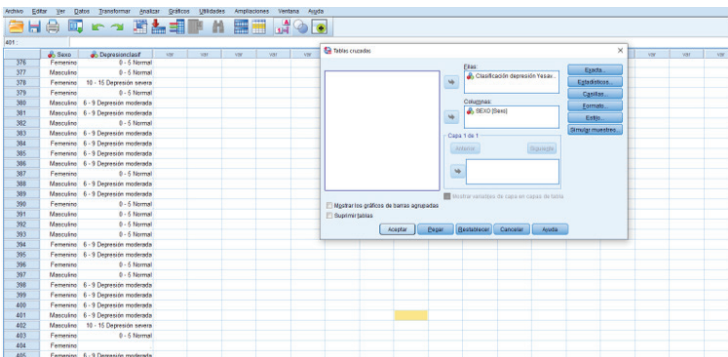
- Se abre la ventana “Estadísticos”, seleccione la opción “Chi Cuadrado”, “continuar”.



4. Abrir la pestaña “Casillas”, una vez dentro de ese apartado, seleccionar “recuentos observados”, “recuentos esperados”, “porcentajes filas” y “continuar”.



5. Una vez seleccionados los comandos en las pestañas “estadísticos” y “casillas”, se oprime la opción “Aceptar”, lo que abrirá la hoja de resultados.



6. En la hoja de resultados se observa una tabla de resumen de casos con valores válidos y perdidos. Por el momento, se omitirá ese resultado. La atención se centrará en la tabla cruzada. En la tabla cruzada se observan los resultados observados y los esperados. Se puede observar que el total de casos es de 393 casos. Hay que observar que todas las casillas esperadas, los valores son mayores a 1 (supuesto 2). Además, ninguna casilla presenta valores esperados menores a 5 (supuesto 3). Por lo que la Prueba Chi-cuadrada es apta para calcular la relación entre la variable “sexo” y “nivel de depresión”. Este dato también se corrobora en la nota al final de la tabla del estadístico de prueba “Pruebas de chi-cuadrado”.

Resumen de procesamiento de casos

	Válido		Perdido		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
Clasificación depresión Yesavage* SEXO	393	97.0%	12	3.0%	405	100.0%

Tabla cruzada Clasificación depresión Yesavage* SEXO

Clasificación depresión Yesavage	Nivel de Depresión	Recuento	SEXO		Total
			Femenino	Masculino	
0 - 5 Normal	Recuento	55	50	105	
	Recuento esperado	56.1	48.9	105.0	
	% dentro de Clasificación depresión Yesavage	52.4%	47.6%	100.0%	
6 - 9 Depresión moderada	Recuento	128	114	242	
	Recuento esperado	129.3	112.7	242.0	
	% dentro de Clasificación depresión Yesavage	52.9%	47.1%	100.0%	
10 - 15 Depresión severa	Recuento	27	19	46	
	Recuento esperado	24.6	21.4	46.0	
	% dentro de Clasificación depresión Yesavage	58.7%	41.3%	100.0%	
Total		Recuento	210	183	393
		Recuento esperado	210.0	183.0	393.0
		% dentro de Clasificación depresión Yesavage	53.4%	46.6%	100.0%

Pruebas de chi-cuadrado

	Valor	df	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	.587 ^a	2	.746
Razón de verosimilitud	.591	2	.744
Asociación lineal por lineal	.351	1	.554
N de casos válidos	393		

^a 0 casillas (0.0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 21.42.

7. Se interpreta el estadístico de prueba:
 Si p -valor < 0.05 se rechaza la H_0
 Si p -valor ≥ 0.05 se acepta la H_0 y se rechaza la H_1
 En este ejemplo, el valor Chi-cuadrado es de .587 y el valor de p .746, lo que significa que se rechaza la hipótesis alterna y se acepta la hipótesis nula.
8. Resultado: Los resultados mostraron que el sexo es independiente del nivel de depresión en adultos mayores de comunidades rurales.

Referencias

- ARRIAZA, B.M. (s.f.). Prueba Chi-cuadrado. En IFAPA (Ed.), *Guía práctica de análisis de datos* (pp. 63-65). https://scholar.google.es/scholar?hl=es&as_sdt=0,5&q=guía+práctica+de+análisis+de+datos+Manuel+Arraiza+Balmón
- BETENSKY A.R. & Rabiowitz, D. (1999). Maximally selected chi2 statistics for kx2 tables. *Biometrics*, 55(1), 317-320. <https://doi.org/10.1111/j.0006-341x.1999.00317.x>
- COCHRAN, W.G. (1952). The Chi-square Test of Goodness of Fit. En Prueba X^2 (10 agosto 2023). En *Wikipedia*. https://es.wikipedia.org/wiki/Prueba_X^2
- HAIR, F. J., Black, C. W., Babin, J.B. & Anderson, E.R. (2013). Key Terms. En *Pearson Education Limited* (Ed.). *Multivariate Data-Analysis* (7ª. ed., p. 519). Impreso https://books.google.com.mx/books/about/Multivariate_Data_Analysis.html?id=VvXZnQEACAAJ&redir_esc=y
- IBM, 2017. Prueba Chi-cuadrado. Actualización 22 marzo 2021. <https://www.ibm.com/docs/es/spss-statistics/saas?topic=tests-chi-square-test>
- McHUGH, I.M. (2013). The Chi-square test of Independence. *Biochemic Medica*, 23(2); 143-9. <http://dx.doi.org/10.11613/BM.2013.018>
- MENDIVELSO, F. & Rodriguez, M. (2018). Prueba Chi-cuadrado de independencia aplicada a tablas 2xN. *Rev Med Sanitas*. 21(2); 92-95. DOI: 10.26852/01234250.6
- SALDAÑA, R.M. (2011). La prueba chi-cuadrado o ji-cuadrado (X^2). *Enfermería del Trabajo*. 1, 31-38. <https://dialnet.unirioja.es/servlet/articulo?codigo=3995561>
- SMARTLINE (21 julio 2021). Prueba Chi-cuadrado, V-Cramer, y tablas cruzadas en SPSS. [Video]. YouTube. https://www.youtube.com/watch?v=eo_YbULtWdQ&t=329s

Capítulo 7.

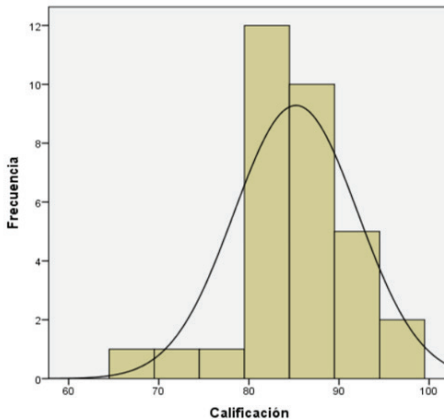
Distribución normal

DCE LUIS CARLOS CORTEZ GONZÁLEZ

Introducción

La distribución normal es una de las distribuciones teóricas mejor estudiadas, así como una de las más importantes y utilizadas en la práctica de la bioestadística. Su importancia radica principalmente en su simplicidad y en la frecuencia con la que algunas variables que se estudian en fenómenos comunes, naturales y de manera cotidiana presentan esta distribución (Dagnino, 2014; Wayne, 2002). A menudo en la naturaleza se encuentran datos graficados mediante histogramas que forman una distribución en forma de campana. A esto se le conoce como distribución normal, curva de distribución normal, campana de Gauss, distribución gaussiana o curva normal (Jackson, 2009; Celis y Labrada, 2014; Pértegas, S. & Pita, S. 2001 y Polit, D. 2013).

Gráfica 1
Campana de Gauss



Diferentes autores mencionan que esta distribución tiene ciertas características especiales (Celis y Labrada, 2014, Cordoba, 2003; Dagnino, 2014; Wayne, 2002) las cuales se enumeran a continuación:

1. Tiene forma de campana y es simétrica.
2. Su distribución es simétrica y la media la divide en dos partes iguales: 50% del área está a la derecha y 50%, a la izquierda.
3. Cuando la variable es discreta, la media, la mediana y la moda son iguales.
4. La mayoría de las observaciones se agrupan alrededor del centro de la distribución, con muchas menos observaciones en los extremos o “colas” de la distribución.
5. Cuando las desviaciones estándar se trazan en el eje x, el porcentaje de puntuaciones que se encuentran entre la media y cualquier punto del eje x es el mismo para todas las curvas normales.

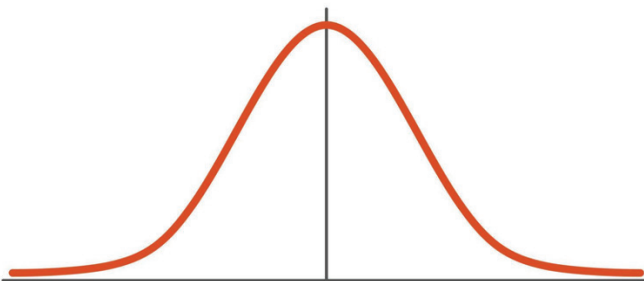
La simetría que se menciona que tiene como principal propiedad la distribución normal indica que la distribución es exactamente igual hacia ambos lados de la media, lo que significa que es unimodal y simétrica, la mediana y la moda coinciden en el mismo valor, que es el centro de la distribución.

Distribuciones normales

Cuando una distribución de puntajes es bastante grande ($N > 30$), a menudo tiende a aproximarse a un patrón llamado distribución normal. Cuando se grafica como un polígono de frecuencia, una distribución normal forma un patrón simétrico en forma de campana, a menudo llamado curva normal.

Decimos que el patrón se aproxima a una distribución normal porque una verdadera distribución normal es una construcción teórica que en realidad no se observa en el mundo real.

Figura 1
Distribución normal



Aplicaciones de la distribución normal

La distribución normal sirve para conocer la probabilidad de encontrar un valor de una variable que sea igual o inferior a un cierto valor. Por ejemplo, se puede aplicar en pruebas de laboratorio de análisis clínicos, Esta se aplica en los niveles de química sanguínea como la glicemia, el ácido úrico, el colesterol y la hemoglobina; en mediciones como la temperatura, la tensión arterial sistólica y diastólica.

Test de normalidad

Las pruebas de normalidad se aplican a un conjunto de datos para determinar distribución normal. Una de las suposiciones más comunes de los procedimientos de prueba estadística es que los datos utilizados deben tener una distribución normal. Por ejemplo, si desea calcular una prueba t o ANOVA (capítulo 10 y 11), primero debe verificar si sus datos o variables se distribuyen normalmente.

Si no existe una distribución normal de los datos, no se pueden utilizar los métodos anteriores y se debe recurrir a pruebas no paramétricas que no requieren una distribución normal de los datos.

La distribución normal se puede comprobar analítica o gráficamente. Las pruebas de gráficos utilizan histogramas y las pruebas analíticas más habituales para comprobar la normalidad de los datos son:

- Prueba de Kolmogorov-Smirnov
- Prueba de Shapiro-Wilk

Todas estas pruebas se utilizan para probar la hipótesis nula de que los datos se distribuyen de manera normal. Por tanto, la hipótesis nula es que la distribución de frecuencia de los datos se distribuye normalmente. Para rechazar o no rechazar la hipótesis nula, obtenga valores p de todas estas pruebas. La gran pregunta aquí es si este valor p es menor o mayor que 0,05.

Si el valor p es inferior a 0,05, esto se interpreta como una desviación significativa de una distribución normal y los datos pueden considerarse no distribuidos normalmente. Si su valor p es mayor que 0,05 y quiere estar completamente limpio estadísticamente, no necesariamente puede afirmar que la distribución de frecuencia sigue una distribución normal y no puede refutar la hipótesis nula. Aunque en realidad, no está completamente limpio, aún puede asumir valores superiores a 0,05 como distribución normal.

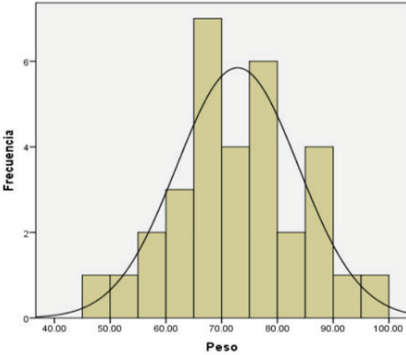
Desafortunadamente, los métodos analíticos tienen importantes inconvenientes, por lo que cada vez se utilizan más métodos gráficos.

El problema es que el valor p calculado se ve afectado por el tamaño de la muestra. Por lo tanto, si tiene una muestra muy pequeña, el valor p puede ser mayor que 0,05, pero si tiene una muestra muy grande de la misma población, el valor p puede ser menor que 0,05.

Supongamos que la distribución de la población se desvía ligeramente de la distribución normal. En segundo lugar, asumimos que los datos se distribuyen normalmente ya que obtenemos valores p muy grandes para muestras muy pequeñas. Sin embargo, a medida que las muestras crecen, el valor p se hace cada vez más pequeño, incluso si las muestras pertenecen a la misma población con la misma distribución. Si la muestra es muy grande, es posible que obtenga un valor p inferior a 0,05, rechazando la hipótesis nula de distribución normal. Razón por la que los métodos gráficos se utilizan cada vez más para resolver este problema.

Si quieres realizar un análisis gráfico se revisa el histograma. Se traza la distribución normal sobre el histograma de sus datos y vea si la curva de la distribución normal coincide estrechamente con la curva de la distribución normal.

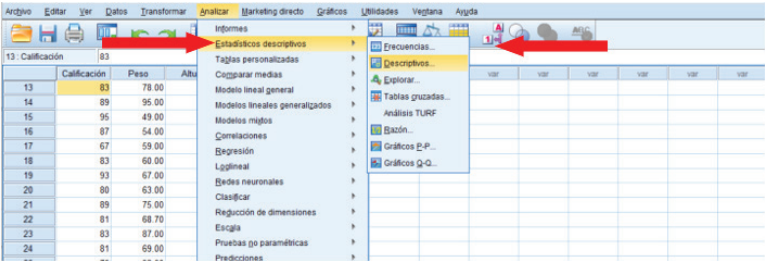
Gráfica 2
Comparación con curva de distribución normal



Ruta SPSS Histogramas

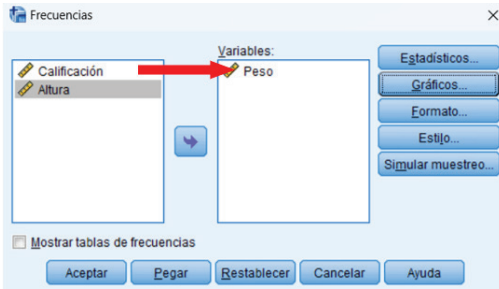
Analizar >>> Estadísticos descriptivos >>> Frecuencias ... (ver figura 2).

Figura 2
Frecuencias/gráficos



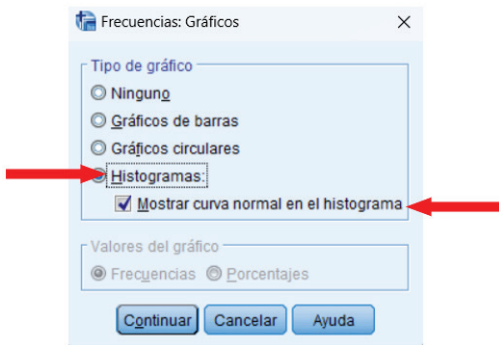
Obtendrás un recuadro denominado Frecuencias donde seleccionaras la variable que quieres graficar en histograma (ver Figura 3).

Figura 3
Recuadro frecuencias



Obtendrás un segundo recuadro denominado Frecuencias/gráficos, ahí seleccionaras histograma y mostrar curva normal en el histograma. Después das continuar (ver Figura 4).

Figura 4
Recuadro histogramas



Kolmogorov-Smirnov

La prueba de Kolmogorov-Smirnov pertenece a la estadística inferencial, la cual tiene como propósito extraer información sobre una población.

Esta es una prueba de bondad de ajuste. Se utiliza para comprobar si los valores determinados a partir de la muestra siguen una distribución normal. En otras palabras, es posible medir el grado de concordancia entre la distribución de un conjunto de datos y una distribución teórica particular. Su objetivo es

indicar si los datos provienen de una población con una distribución teórica específica. Esto representa una comparación de si el valor observado se puede derivar razonablemente de la distribución especificada.

Hipótesis

Para aplicar correctamente la prueba de Kolmogorov-Smirnov, es necesario hacer una serie de suposiciones. Primero, la prueba supone que los parámetros de la distribución de la prueba están predeterminados. Este paso estima los parámetros de la muestra. Por otro lado, la media muestral y la desviación estándar son los parámetros de una distribución normal, el mínimo y máximo muestral definen el área de una distribución uniforme, la media muestral es un parámetro de una distribución de Poisson y la media muestral es un parámetro de una distribución exponencial. La capacidad de la prueba de Kolmogorov-Smirnov para detectar desviaciones de la distribución hipotética puede verse gravemente comprometida. Para comparar esto con la distribución normal a partir de la cual estimamos los parámetros, deberíamos considerar la posibilidad de utilizar la prueba de K-S Lilliefors.

FORMULACIÓN DE HIPÓTESIS

H_0 : hipótesis nula o de trabajo > 0.05 HOMOGENEIDAD (la distribución es normal)

H_1 : hipótesis alterna o del investigador < 0.05 DE DIFERENCIAS (la distribución no es normal)

Supuestos de la prueba

Para aplicar correctamente la prueba de Kolmogorov-Smirnov, es necesario hacer una serie de suposiciones. Primero, la prueba supone que los parámetros de la distribución de la prueba están predeterminados. Este paso estima los parámetros de la muestra. Por otro lado, la media muestral y la desviación estándar son los parámetros de una distribución normal, el mínimo y máximo muestral definen el área de una distribución uniforme, la media muestral es un parámetro de una distribución de Poisson y la media muestral es un parámetro de una distribución exponencial.

La capacidad de la prueba de Kolmogorov-Smirnov para detectar desviaciones de la distribución hipotética puede verse gravemente comprometida. Para comparar esto con la distribución normal a partir de la cual estimamos los parámetros, deberíamos considerar la posibilidad de utilizar la prueba de K-S Lilliefors.

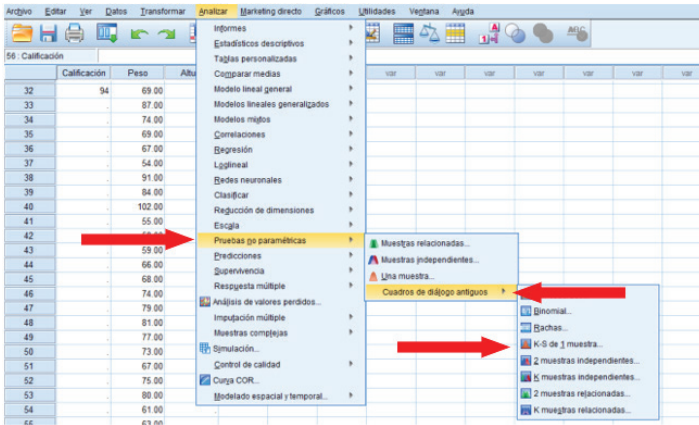
La principal diferencia con esto es la cantidad de muestras. La prueba de Shapiro-Wilk se puede utilizar para hasta 50 puntos de datos, mientras que la prueba de Kolmogorov-Smirnov se recomienda para más de 50 observaciones.

Antes de ejecutar la prueba de Kolmogorov-Smirnov (con corrección de Lilliefors) en R, necesita saber qué prueba de hipótesis desea realizar.

Ruta SPSS prueba de Kolmogorov-Smirnov

Analizar >>> Pruebas no paramétricas >>> Cuadro de diálogos antiguos >>> K-S de 1 muestra (ver figura 5).

Figura 5
Prueba de Kolmogorov-Smirnov



Obtendrás un recuadro denominado prueba de Kolmogórov-Smirnov para una muestra, seleccionas la variable a analizar y das Aceptar (ver Figura 6).

Figura 6
Cuadro de dialogo Kolmogorov-Smirnov para una muestra

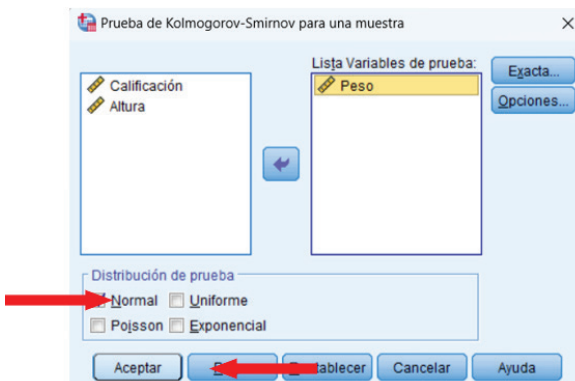



Figura 7

Resultados de Prueba de Kolmogórov-Smirnov para una muestra

Prueba de Kolmogorov-Smirnov para una muestra

		Peso
N		55
Parámetros normales ^{a,b}	Media	72.6400
	Desviación estándar	11.26253
Máximas diferencias extremas	Absoluta	.064
	Positivo	.064
	Negativo	-.054
Estadístico de prueba		.064
Sig. asintótica (bilateral)		.200 ^{c,d}



Se acepta la hipótesis nula o de trabajo y se interpreta como que la variable presenta una distribución normal.

Prueba de Shapiro-Wilk

La prueba de Shapiro-Wilk se utiliza para evaluar si un conjunto de datos sigue una distribución normal. Es una de las más utilizadas y fiables para evaluar la normalidad. Esta prueba compara los datos recopilados con una distribución normal teórica y determina si existe una diferencia significativa entre los dos y se emplea cuando la muestra es menor a 50 observaciones.

Este hecho es muy importante porque muchos otros análisis estadísticos requieren la normalidad de los datos.

FORMULACIÓN DE HIPÓTESIS

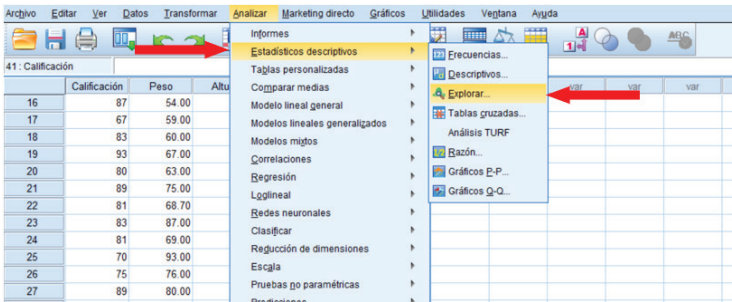
H_0 : hipótesis nula o de trabajo > 0.05 (proviene de una distribución normal)

H_1 : hipótesis alterna o del investigador < 0.05 (no proviene de una distribución normal)

Ruta SPSS Prueba de Shapiro-Wilk

Analizar >>> Estadísticos descriptivos >>> Explorar (ver figura 8).

Figura 8
Prueba de Shapiro-Wilk



Obtendrás un recuadro denominado Explorar, seleccionas la variable a analizar, seleccionas ambos y das Aceptar (ver Figura 9).

Figura 9
Cuadro de dialogo Prueba de Shapiro-Wilk

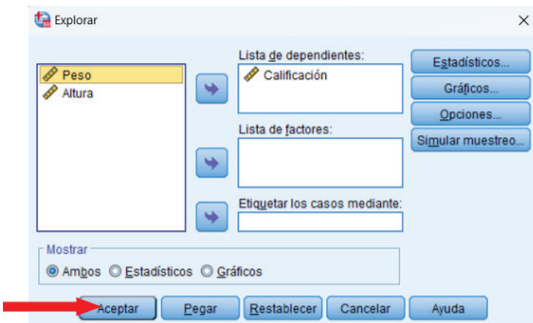


Figura 10
Resultado Prueba de Shapiro-Wilk

Shapiro-Wilk		
Estadístico	gl	Sig.
.975	40	.507

Se acepta la hipótesis nula o de trabajo y se interpreta como que la variable proviene de una distribución normal.

Referencias

- BEYER, W. (2002). CRC Standard Mathematical Tables, 31^a ed. Boca Raton, FL: CRC Press, págs. 536 y 571.
- CELIS, A. y Labrada, V. (2014). Bioestadística. Manual Moderno Ed. Tercera.
- CORDOBA, M. (2003). Estadística: descriptiva e inferencial. Aplicaciones. Distribuidora, Imprenta, Editorial, Librería Ed. Quinta.
- DAGNINO, J. (2014). La distribución normal. Rev Chil Anest, 43, 116-121.
- GARCÍA, R. González, J. y Jornet, J. (2010). SPSS: Pruebas No Paramétricas.
- JACJSON, S. (2009). Research methods and statistics: a critical thinking approach. Wadsworth Cengage Learning. Ed Tercera.
- PÉRTEGAS Díaz, S., & Pita Fernández, S. (2001). La distribución normal. Cad Aten Primaria, 8, 268-274.
- POLIT, D. (2013). Statistics and Data Analysis for Nursing Research: Pearson New International Edition. Pearson Higher Ed.
- TAPIA, C. & Cevallos, K. (2021). Pruebas para comprobar la normalidad de datos en procesos productivos: anderson-darling, ryan-joiner, shapiro-wilk y kolmogórov-smirnov. *Societas*, 23(2), 83-106.
- WAYNE, D. (2002). Bioestadística base para el análisis en ciencias de la salud (John Wiley & Sons, Ed.; Cuarta, Vol. 1).

Capítulo 8.

Medidas de asociación para datos numéricos y categóricos. Correlación de Pearson. Correlación de Spearman

DCE ANA LAURA CARRILLO CERVANTES

Definición de la prueba

Cuando se realiza una investigación de dos o más variables, se puede plantear un objetivo de correlación y se puede describir si existe o no correlación y la fuerza de asociación entre ellas. Cuando las variables son cuantitativas, el modo de determinar si están relacionadas y la fuerza con la que lo están, es el coeficiente de correlación lineal. En este capítulo se describe las pruebas para datos paramétrico y no paramétricos o de libre distribución, a través de las pruebas de coeficiente de correlación de Pearson y el de Spearman, según corresponda.

Las pruebas de correlación pertenecen al nivel investigativo relacional, en el cual se realiza el análisis de dos variables (X y Y) y se describe la fortaleza y dirección de la relación entre ellas. Dos variables están asociadas cuando una variable nos da información acerca de la otra. Por el contrario, cuando no existe relación, el aumento o disminución de una variable (X) no nos dice nada sobre el comportamiento de la otra variable (Y).

La correlación mide el valor en el que cambian las variables estudiadas en un mismo momento (esto es, cambia X y se modifica Y). Es una prueba que se usa de manera común para describir relaciones simples sin hacer afirmaciones causales.

El coeficiente de Pearson es elegido para variables en el nivel de intervalo (continuas). También puede ser utilizada si tiene una variable continua (ej. Puntuaciones en la medición de ansiedad) y una variable dicotómica (ej. Sexo: hombres/mujeres).

El Coeficiente de correlación de rango de Spearman se utiliza para calcular la correlación entre dos variables por sus rangos. Es una medida no paramétrica de correlación. El coeficiente de correlación de Spearman, que se denota con el símbolo r y también con r_s , mide la fuerza de la asociación entre dos variables clasificadas para datos ordinales, de intervalo y de razón. Puede tomar valores de $+1$ a -1 . Un valor de r_s de $+1$ indica una asociación perfecta de rangos, cero indica que no hay asociación entre rangos y -1 indica una asociación negativa

perfecta de rangos. Cuando r_s está muy cerca de cero, indica una asociación más débil entre los rangos.

Un diagrama de dispersión, puede ser un apoyo visual para ver si existe correlación entre las variables. En el eje de las X (horizontal), se coloca la variable que se mide primero y en el eje de las Y (vertical) la segunda variable. Una inspección al diagrama provee información sobre la dirección de la relación (positiva o negativa) tanto como de la fortaleza de la relación. Un diagrama de dispersión de una correlación perfecta ($r = 1$ o -1) mostraría una línea recta. Cuando $r = 0$, el diagrama de dispersión mostraría un conjunto de puntos sin un patrón evidente.

Supuestos de las pruebas

A continuación, se muestran los supuestos de las pruebas de correlación Pearson y Rho de Spearman.

Nivel de Medición

Para las pruebas de correlación, el nivel de medición de las variables debe ser de intervalo o continuas. Una excepción a esto es si tienes una variable independiente dicotómica (con solo dos valores; ej. Sexo (masculino y femenino) y una variable dependiente continúa.

La Rho de Spearman, el cual es un coeficiente de correlación, se puede usar para datos ordinales o de rango.

Pares Relacionados

Lo cual se refiere a que debe haber un valor para la variable X y Y, además de provenir de un mismo sujeto.

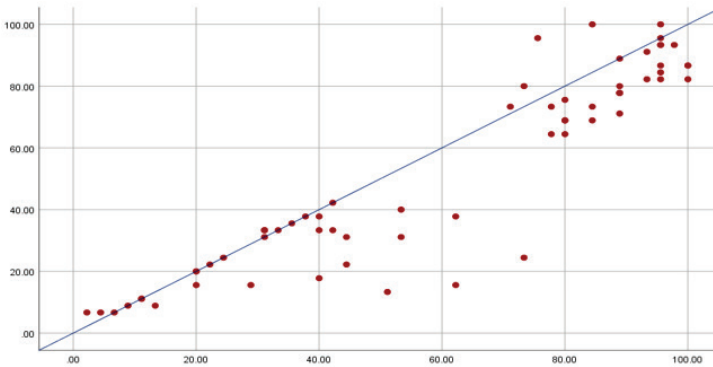
Normalidad

Las variables deben mostrar una distribución normal, para usar Pearson y una distribución diferente a la normal para Rho de Spearman.

Linealidad

La relación entre las dos variables debe ser lineal. Esto significa que cuando ves tu diagrama de dispersión, debes ver una línea estrictamente recta, no una curva (Figura 1). Cuando se obtiene valores de r muy bajos se debe revisar el diagrama de dispersión.

Figura 1
Gráfica de dispersión

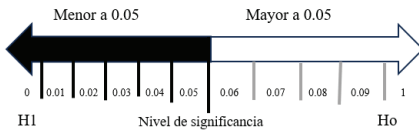


Para una prueba de correlación de Pearson, las observaciones deben provenir de una variable numérica, categóricas y con distribución normal. Para la prueba de correlación de Spearman pueden ser variables categóricas, ordinales o numéricas y su distribución puede ser no normal (Supo, 2022).

El planteamiento de la hipótesis consiste en escribir una hipótesis nula (H_0) y una alterna o del investigador (H_1), primero se aconseja escribir la alterna y después oponiéndose a esta la nula. La H_1 , se obtiene afirmando el enunciado del estudio, por ejemplo, si el enunciado del estudio dice que se relaciona el tiempo desempeñando el rol de cuidador con la funcionalidad familiar. La H_1 diría que Existe relación entre el tiempo desempeñando del rol de cuidador con la funcionalidad familiar; mientras que la H_0 diría que No Existe relación entre el tiempo desempeñando el rol de cuidador con la funcionalidad familiar

La elección de la hipótesis de correlación es de acuerdo con el valor p obtenido, si el valor p es menor a 0.05 se rechaza la hipótesis nula, por lo que se acepta la hipótesis alterna. Si el valor p es mayor a 0.05 se acepta la hipótesis nula (Diagrama 1).

Diagrama 1
Elección de hipótesis



Aquí algunos ejemplos de hipótesis:

H_0 : Las madres con bajos niveles de hemoglobina **no tienen** recién nacidos con peso disminuido

H1: Las madres con bajos niveles de hemoglobina **tienen** recién nacidos con peso disminuido

Ho “**no existe** correlación entre la satisfacción laboral y el desempeño de las enfermeras de un hospital público del noreste de México”.

H1 “**existe** correlación entre la satisfacción laboral y el desempeño de las enfermeras de un hospital público del noreste de México”.

Ho “**no existe** correlación de la sarcopenia con el componente físico de la calidad de vida relaciona con la salud, capacidad funcional, composición corporal, velocidad de la marcha, masa músculo esquelético apendicular y fuerza de empuñadora”.

H1 “**existe** correlación de la sarcopenia con el componente físico de la calidad de vida relaciona con la salud, capacidad funcional, composición corporal, velocidad de la marcha, masa músculo esquelético apendicular y fuerza de empuñadora”.

Interpretación

El valor de los coeficientes de correlación puede oscilar de -1 hasta $+1$. El signo indica si es que existe una correlación positiva (mientras una variable se incrementa, la otra también) o negativa (mientras una variable se incrementa, la otra disminuye). Por su parte el número (sin el signo) provee información sobre la fuerza de la relación. Si el valor es de -1 a $+1$, la correlación es perfecta, significa que el valor de X puede ser determinada al conocer exactamente el valor de Y . Si el valor es de 0 , significa que no hay relación.

El coeficiente de correlación de Pearson tiene el objetivo de indicar cuán asociadas se encuentran dos variables entre sí por lo que:

Correlación menor a cero: Si la correlación es negativa, significa que las variables se relacionan inversamente. Es decir, a mayor de X , menor de Y . Mientras más próximo se encuentre a -1 , más clara será la correlación. Si el coeficiente es igual a -1 , nos referimos a una correlación negativa perfecta.

Correlación mayor a cero: La correlación es positiva, es decir a mayor de X , mayor de Y , mientras más cercana está a 1 mayor es la fuerza de correlación, y si está cercana a 0 la correlación es débil.

Correlación igual a cero: Cuando la correlación es igual a cero significa que no es posible determinar algún sentido de covariación. Sin embargo, no significa que no exista una relación no lineal entre las variables. Cuando las variables son independientes significa que estas se encuentran correlacionadas, pero esto no significa que el resultado sea verdadero.

La interpretación de los valores es diversa, en la Tabla 1 se pueden observar diferentes autores.

Tabla 1
Interpretación de la correlación

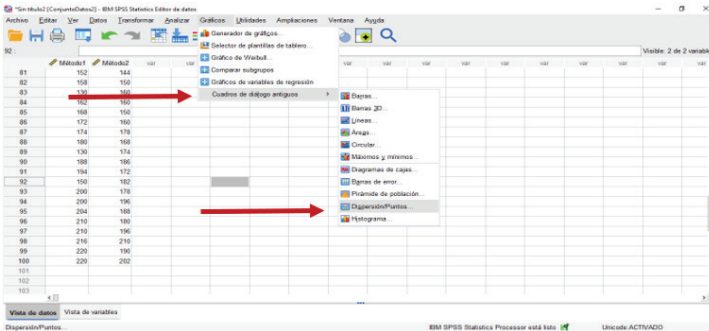
Valor de coeficiente de correlación	Cohen, 1988	Supo, 2023	CIMEC, 2023	Chan YH	
+1	-1	Grande	Excelente	Fuerte	Perfecta
+0.9	-0.9	Grande	Excelente	Fuerte	Muy fuerte
+0.8	-0.8	Grande	Bueno	Fuerte	Muy fuerte
+0.7	-0.7	Grande	Bueno	Fuerte	Moderadamente fuerte
+0.6	-0.6	Grande	Regular	Fuerte	Moderadamente fuerte
+0.5	-0.5	Grande	Regular	Fuerte	Justa
+0.4	-0.4	Mediano	Bajo	Moderada	Justa
+0.3	-0.3	Mediano	Bajo	Moderada	Justa
+0.2	-0.2	Pequeño	Muy Bajo	Débil	Pobre
+0.1	-0.1	Pequeño	Muy Bajo	Inexistente	Pobre
0	0	Ninguna	Ninguna	Ninguna	Ninguna

Interpretación 1= Cohen (1988, p.79-81); Interpretación 2= Supo (2023); Interpretación 3= CIMEC (2023).

Ruta SPSS Gráfica de Dispersión

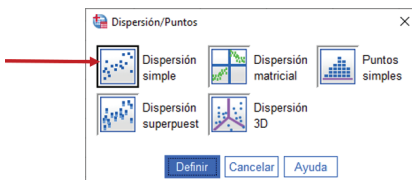
Gráficos >>> Cuadro de diálogos antiguos >>> Dispersión/puntos... (Ver Figura 2).

Figura 2
Dispersión/puntos



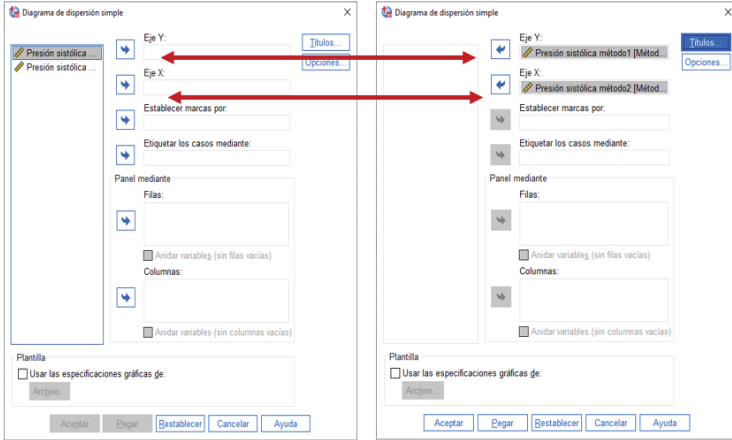
Obtendrás un segundo recuadro denominado Dispersión/puntos donde seleccionarás: Dispersión simple (Ver Figura 3).

Figura 3
Cuadro de diálogo dispersión simple



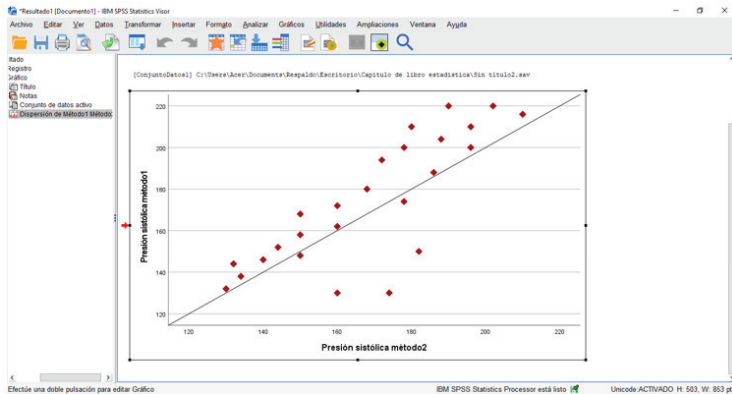
Posteriormente se activará el recuadro de Diagrama de dispersión simple. Donde colocarás la primera variable en el Eje Y, y la segunda variable en el Eje X (Ver Figura 4).

Figura 4
Diagrama de dispersión simple



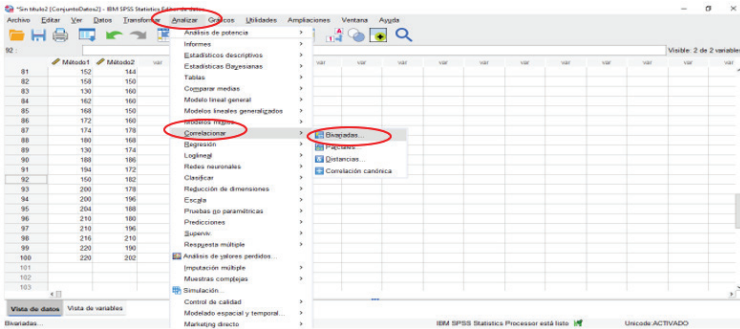
Una vez colocadas las variables, da aceptar. Te moverá al visor de resultados, ya con la gráfica (Ver Figura 5).

Figura 5
Visor de resultado de gráfica de dispersión



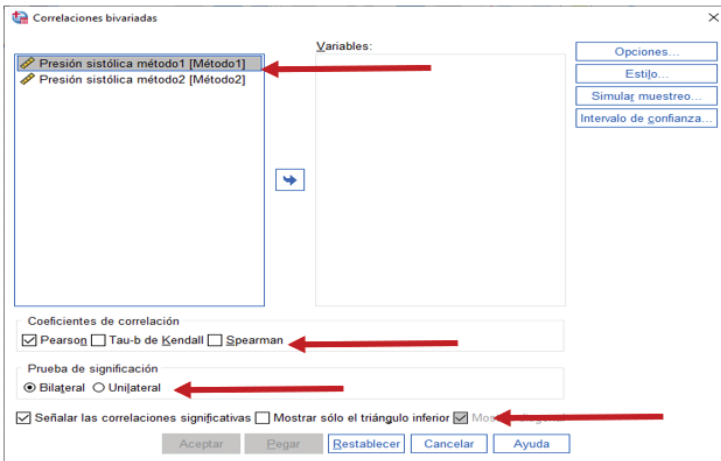
El procedimiento para el cálculo del coeficiente de correlación entre dos variables es Correlación bivariada, que se realiza mediante: Analizar >>> Correlacionar >>> Bivariadas...(Ver Figura 6).

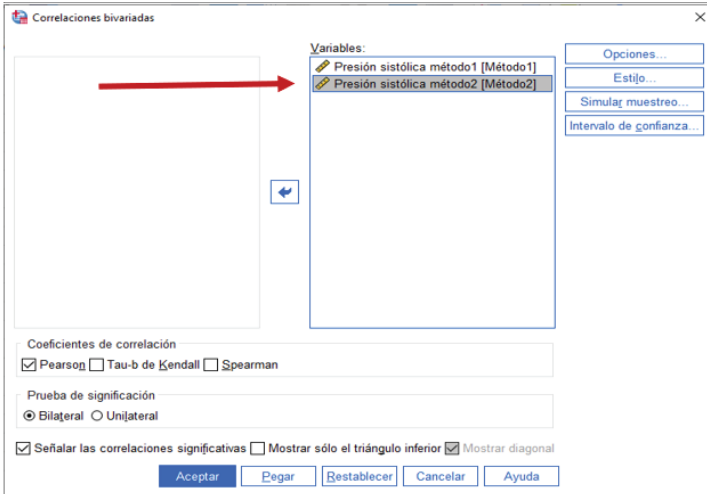
Figura 6.
Correlación bivariada



Posterior a esto, el SPSS emite una ventana (Figura 7), en la cual, del lado izquierdo, hay una caja con las variables a seleccionar por el procedimiento y debes pasarlas a la caja variables. El programa correlacionará todas las variables que se encuentren en esta caja. Los coeficientes de correlación que podremos correr son: el de Pearson, el de la Tau-b de Kendall y el de Spearman; los tres pueden ser seleccionados simultáneamente, aunque en este capítulo solo se revisan.

Figura 7
Cuadro de diálogo de la Correlación Bivariada





Así mismo, el programa da la opción de seleccionar la Prueba de significación, se puede seleccionar la unilateral o bilateral del contraste al que serán sometidos los coeficientes de correlación; SPSS ofrece, de entrada, el bilateral. Si se conoce de antemano la dirección de la asociación, a través de la revisión de la literatura, se selecciona “Unilateral”. Si no es así, se selecciona la “Bilateral”. El programa da también la posibilidad de marcar automáticamente los coeficientes de correlación que resulten significativos; de manera que, si se marcan correlaciones significativas, el programa marca con un asterisco (*) los coeficientes de correlación significativos al nivel 0,05 y, con dos asteriscos (**), los significativos al nivel 0, 01.

Una vez que se seleccionaron las variables y características requeridas en el análisis, se obtendrá un visor de resultados en forma de tabla cuadrada (Ver Tabla 2), en la cual aparecen las variables solicitadas, en el ejemplo Presión sistólica método 1 y Presión sistólica método 2. En la tabla de resultados donde se cruzan una fila y una columna, aparece el valor del coeficiente de correlación de las variables a estudiar (si se solicitó marcada con un asterisco), la significancia estadística del contraste de las variables y el número de casos sobre los que se calculó el coeficiente de correlación. Por último, es importante mencionar que la tabla es simétrica con respecto a su diagonal principal, también es llamada tabla espejo, ya que presenta los mismos datos en la parte superior e inferior de la tabla (Ver Tabla 3).

Tabla 2
Resultados de prueba de correlación

		Correlaciones	
		Presión sistólica método1	Presión sistólica método2
Presión sistólica método1	Pearson	Correlación de	1
		Sig. (bilateral)	,811**
	N	100	100
Presión sistólica método2	Pearson	Correlación de	,811**
		Sig. (bilateral)	,000
	N	100	100

** La correlación es significativa en el nivel 0,01 (bilateral).

Tabla 3
Resultados de prueba de correlación en tabla en triángulo inferior

		Correlaciones	
		Presión sistólica método1	Presión sistólica método2
Presión sistólica método1	Pearson	Correlación de	--
		N	100
Presión sistólica método2	Pearson	Correlación de	,811**
		Sig. (bilateral)	,000
	N	100	100

** La correlación es significativa en el nivel 0,01 (bilateral).

Ejemplos/ejercicios

Correlación de Spearman

Una profesora de la Universidad Autónoma de Coahuila realizó una investigación con Cuidadores familiares de adultos mayores con enfermedad crónica, cuyo objetivo fue el “Determinar la fuerza de correlación del tiempo desempeñando el rol de cuidador, la funcionalidad familiar, conocimientos sobre la enfermedad crónica, la sobrecarga, los niveles de incertidumbre y la resiliencia del cuidador familiar de adultos mayores con enfermedad crónica de la ciudad de Saltillo, Coahuila”. En este estudio participaron 196 cuidadores familiares, asumiendo la normalidad de las variables de interés, se procesa la prueba de Correlación de Spearman, obteniendo los siguientes resultados (Tabla 4). Se solicita que se interprete los resultados.

Hipótesis

Ho “No existe correlación del tiempo del tiempo desempeñando el rol de cuidador, la funcionalidad familiar, los conocimientos sobre la enfermedad y la sobrecarga del cuidador de la PAM”.

H1 “Existe correlación del tiempo del tiempo desempeñando el rol de cuidador, la funcionalidad familiar, los conocimientos sobre la enfermedad y la sobrecarga del cuidador de la PAM”.

Diagrama de decisión de

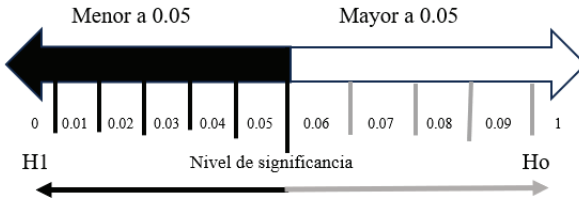


Tabla 4
Correlación entre las variables de estudio y datos sociodemográficos del CF

	1	2	3	4
1. Tiempo desempeñando el rol de cuidador	---	---	---	---
2. Funcionalidad familiar	---	---	---	---
3. Conocimientos sobre la enfermedad	0.17*	---	---	---
4. Sobrecarga del cuidar	0.18*	-0.26*	0.17*	---

* $p \leq 0.05$. ** $p \leq 0.001$.

De acuerdo con los resultados obtenidos en cuanto al valor de $p < 0.05$, se rechaza la hipótesis nula, por lo que se acepta la alterna donde se plantea que existe correlación del tiempo desempeñando el rol de cuidador, la funcionalidad familiar, los conocimientos sobre la enfermedad y la sobrecarga del cuidador de la PAM. La interpretación podría ser de la siguiente manera:

En la Tabla 4 se muestran los resultados de la prueba de correlación de Spearman, donde se encontró una correlación pequeña y positiva de los conocimientos sobre la enfermedad con el tiempo de desempeño del rol de cuidador ($r=0.17$, $p=0.01$), lo que significa que, a mayor nivel de conocimientos, mayor es el tiempo de desempeño como cuidador.

Así mismo, la sobrecarga del CF mostró una correlación pequeña y positiva con el tiempo de desempeño del rol de cuidador y los conocimientos de la enfermedad ($r=-0.18$, $p=0.00$; $r=-0.17$, $p=0.00$), es decir, que, a mayor sobrecarga del cuidado, mayor es el tiempo de desempeño del rol de cuidador y los conocimientos de la enfermedad. Por último, se encontró una correlación pequeña y negativa entre la sobrecarga y la funcionalidad familiar ($r= -0.26$, $p=0.00$), es decir, que, a mayor sobrecarga del cuidado, menor es la funcionalidad familiar.

Referencias

- CARREÑO Moreno, S. P., Mayorga Álvarez, J. H. (2020). *Estadística básica para la ciencia del cuidado humano (1ª Ed.)*. Colombia; Universidad Nacional de Colombia.
- COHEN, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Second Edition. Hillsdale, NJ: LEA.
- CHAN Y. H. (2003). Biostatistics 104: correlational analysis. *Singapore med journal*, 44(12), 614–619.
- DANIEL, W. W. (2010). *Bioestadística: Base para el análisis de las ciencias de la salud (4a. ed.)*. México: Limusa Wiley.
- GARCÍA, A. (2011). *Elementos de bioestadística* (España: Universidad de Extremadura, Ed.; Tercera, Vol. 1).
- GONZÁLEZ, F., Escoto, M., & Chávez, J. (2017). *Estadística aplicada en psicología y ciencias de la salud* (México: El manual moderno, Ed.; Primera edición, Vol. 1).
- GUERRA, T. (2014). *Bioestadístico* (México: UNAM-FEIS Zaragoza, Ed.; Primera, Vol. 1).
- MINITAB. (2019). Entendiendo las pruebas t: Valores t y distribuciones t. Minitab Blog.
- MONTES, A., Ochoa J., Juárez. B., Vázquez, M., Díaz, C. (Decimocuarta semana Internacional de la estadística y la probabilidad: 14 al 18 de junio, 2021). Aplicación del coeficiente de correlación de Spearman en un estudio de fisioterapia. Semana internacional de estadística y probabilidad.
- QUÍNTELA, A. (2019). *Estadística Básica Edulcorada* (1st ed., Vol. 1). México: Independently Published.
- SPSS Statics. (2021). Prueba T. IBM. Recuperado de: <https://www.ibm.com/docs/es/spss-statistics/25.0.0?topic=features-bivariate-correlations> (Consultado el 20 de octubre de 2023).

Capítulo 9.

Medidas de asociación para datos numéricos y categóricos. Correlación Tau B de Kendall. Tau C de Kendall

DCE CLAUDIA JENNIFER DOMÍNGUEZ CHÁVEZ
DCE ABRAHAM ISAAC ESQUIVEL RUBIO

Introducción

El coeficiente de correlación de rango Tau de Kendall es una medida no paramétrica que permite determinar la ausencia o presencia de la asociación lineal entre variables de tipo ordinal o de razón (Badii, M., Guillen, O., Lugo, O. & Aguilar, 2014; Gibbons & Chakraborti, 2010; International Business Machines, 2021a, 2021b). A través del coeficiente Tau () se realiza la transformación de los valores en rangos, para determinar la fuerza y dirección de la asociación entre las variables (Morales & Rodríguez, 2016). A partir de la consideración o no del empate entre los rangos, el coeficiente se puede clasificar como Tau-b y Tau-c de Kendall respectivamente.

Coefficiente de correlación Tau-b de Kendall

Esta medida no paramétrica considera los empates en los rangos y es ideal cuando las tablas cruzadas presentan el mismo número de categorías en las filas y en las columnas (Göktaş & Işçi, 2011; International Business Machines, 2021a; Leyva & Flores, 2014). Imagine que desea analizar la asociación entre el estado nutricional (estado de peso adecuado, riesgo de desnutrición y desnutrición) y el riesgo de úlceras por presión (bajo, moderado y alto), ambas variables tienen la misma cantidad de categorías con un orden jerárquico (Tabla 1). En este caso a través de la medida Tau-b se buscaría analizar si existe una asociación lineal entre el estado nutricional y las úlceras por presión.

Tabla 1
Tabla cruzada 3 x 3

Estado nutricional	Riesgo de úlcera por presión		
	Bajo	Moderado	Alto
Estado de peso adecuado			

Estado nutricional	Riesgo de úlcera por presión		
	Bajo	Moderado	Alto
Riesgo de desnutrición			
Desnutrición			

Fuente: Elaboración propia

Coeficiente de correlación Tau-c de Kendall

Por el contrario, Tau-c no tiene en cuenta los empates en los rangos (Göktaş & Işçi, 2011; International Business Machines, 2021b, 2021a). En este caso las tablas suelen presentarse con menos categorías en las filas en comparación con las columnas. La tabla 2 presenta el ejemplo de una tabla cruzada entre el estado cognitivo con dos categorías (adecuado y probable deterioro cognitivo) y los síntomas de depresión (sin síntomas, síntomas leves, síntomas moderados y síntomas graves).

Tabla 2

Tabla cruzada 2 x 3

Estado cognitivo	Síntomas de depresión			
	Sin síntomas	Síntomas leves	Síntomas moderados	Síntomas graves
Adecuado				
Probable deterioro cognitivo				

Fuente: Elaboración propia

Supuestos

Para realizar esta medida de asociación es necesario que las variables sean de tipo ordinal es decir que sus categorías impliquen un orden jerárquico (International Business Machines, 2021a). Es de utilidad cuando se trabaja con muestras pequeñas, se trata de una medición no paramétrica por lo que es necesario confirmar que las variables no presentan una distribución normal (International Business Machines, 2021a).

Hipótesis

H0: No existe asociación entre las variables

H1: Existe asociación entre las variables

Interpretación

Para la interpretación es necesario considerar la dirección y fuerza de la asociación. La dirección puede ser positiva o negativa, en donde el signo negativo en

caso de existir una asociación significativa ($p \leq .05$) indicará que mientras una variable aumenta la otra disminuye. Caso contrario en la asociación positiva, en la que se evidenciaría que ambas variables van en aumento, es decir van en la misma dirección.

La fuerza de asociación se establece a partir de los valores de los coeficientes Tau-b y Tau-c de Kendall que van de -1 a 1 (International Business Machines, 2021a), considerándose estos valores como asociaciones perfectas. Si las variables no presentan asociación el coeficiente tiene valor de cero (Bolboaca, S.D. & Jäntschi, 2006). En la tabla 3 se presentan los valores del coeficiente que corresponden al grado de fuerza baja, media o alta (Cohen, 1988).

Tabla 3
Grado de fuerza de los coeficientes

Grado de fuerza	Valor del coeficiente
Baja	.10 a .29
Mediana	.30 a .49
Alta	$\geq .50$

Fuente: Elaboración propia con base en Cohen, 1988

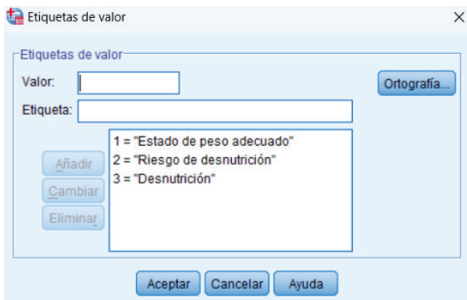
Es importante recordar que las medidas de asociación como lo es el caso de Tau-b y Tau-c de Kendall no permiten establecer relaciones causa efecto, por lo que el tener un resultado de asociación significativa no podría considerarse que una variable genera cambios en la otra ya que solo buscan determinar si las variables se relacionan y de ser así la dirección y grado de fuerza de esa relación (Morales, P. & Rodríguez, 2016).

Ruta SPSS

Para iniciar a trabajar con el proceso de análisis es necesario que considere un análisis preliminar de la base con la finalidad de asegurarse que no existen datos perdidos o información capturada incorrectamente. Debido a que las variables consideradas para esta prueba son de carácter ordinal, es ideal que las categorías se encuentren codificadas en el programa SPSS acorde a su jerarquía. Ejemplo, la variable “estado nutricional” cuenta con tres posibles categorías; estado de peso adecuado, riesgo de desnutrición y desnutrición. Al ingresar las etiquetas de valor, será necesario asignar el código numérico acorde a la jerarquía implícita en la variable (imagen 1).

Imagen 1

Ejemplo de codificación numérica para variables ordinales



Fuente: Elaboración propia

Para dar cumplimiento al segundo supuesto, recuerde que es necesario ejecutar la prueba de normalidad de Shapiro-Wilk o Kolmogorov-Smirnov (acorde al tamaño de la muestra) para identificar que se cumpla el supuesto en el que las variables no presentan una distribución normal ($p < .05$).

Rutas SPSS para Tau-b de Kendall

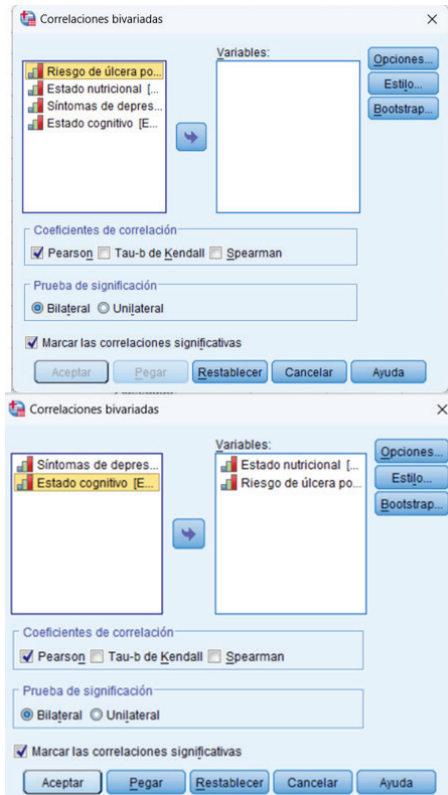
Esta medida tiene dos posibles rutas para efectuar el análisis en SPSS la primera de ellas es mediante el apartado “Correlaciones bivariadas” y la segunda que comparte con la ruta para Tau-c de Kendall que es a través de “Tablas cruzadas”.

Para realizar la prueba por medio de la ruta “correlaciones bivariadas” diríjase al menú en la parte superior y ejecute los siguientes pasos:

- Seleccione la pestaña analizar >> correlaciones >> correlaciones bivariadas.
- Identifique las variables que se analizaran, selecciónelas dando doble clic sobre ellas para pasarlas del lado derecho al apartado “variables” (imagen 2).

Imagen 2

Paso uno para realizar prueba Tau-b de Kendall por la ruta “Correlaciones bivariadas”



Fuente: Elaboración propia

- Diríjase al apartado “Coeficientes de correlación” elimine la selección automática que se hace en el cuadro Pearson y de clic en el cuadro para elegir Tau-b de Kendall y después en aceptar (imagen 3).
- Se desplegará el cuadro en el que deberá identificar la significancia estadística para determinar si existe asociación entre las variables, para posteriormente interpretar el valor de coeficiente de acorde a su dirección y fuerza.

Imagen 3

Paso dos para realizar la prueba Tau-b de Kendall por la ruta “Correlaciones bivariadas”



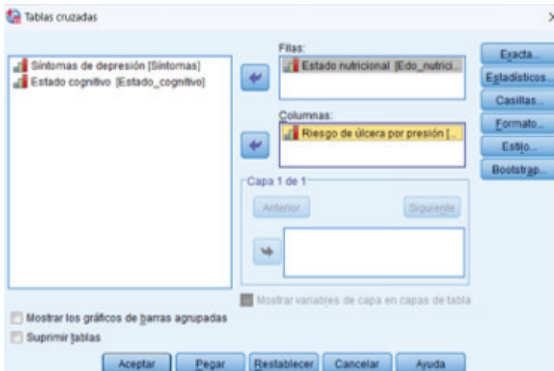
Fuente: Elaboración propia

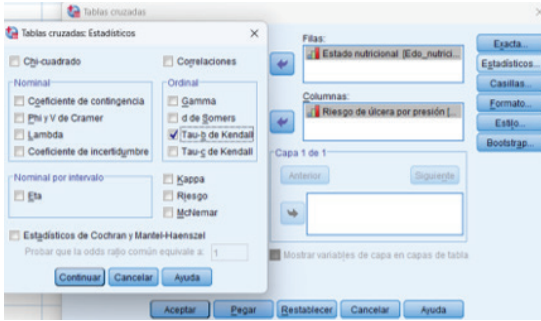
Para aplicar la ruta para Tau-b de Kendall por medio de tablas cruzadas siga los pasos que se presentan a continuación.

- Seleccione la pestaña analizar >> estadística descriptiva >> tablas cruzadas.
- Identifique las variables que se analizarán, seleccione una y envíela al apartado de filas y la otra al de columnas, dando doble clic sobre ellas.
- Del lado derecho de clic sobre la opción “Estadísticos” y seleccione “Tau-b de Kendall” y continuar (imagen 4).
- Finalice el proceso seleccionando aceptar
- Identifique el valor de la significancia, la fuerza y dirección de la asociación en caso de ser mayor a cero.

Imagen 4

Selección de Tau-b en tabla cruzada





Fuente: Elaboración propia

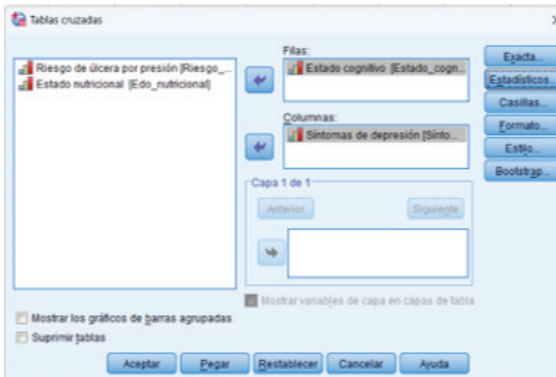
Ruta SPSS para Tau-c de Kendall

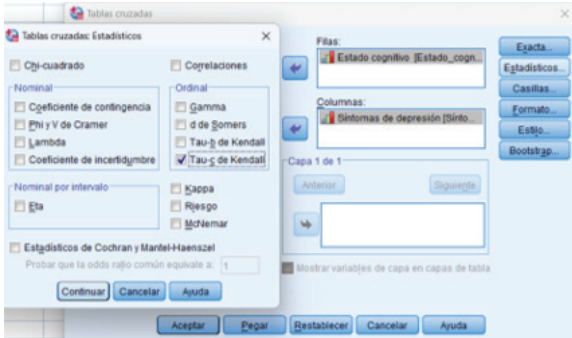
Para Tau-c de Kendall se aplica la misma ruta de tablas cruzadas que Tau-b de Kendall.

- Seleccione la pestaña analizar >> estadística descriptiva >> tablas cruzadas.
- Identifique las variables que se analizarán, selecciónelas dando doble clic sobre ellas. Para facilitar la visualización de la tabla de resultados, se sugiere que la variable con menos categorías se envíe al apartado de filas y la otra variable a columnas.
- Del lado derecho de clic sobre la opción “Estadísticos” y seleccione “Tau-c de Kendall” y continuar (imagen 4).
- Finalice seleccionando aceptar e interprete los datos considerando la significancia y el valor del coeficiente.

Imagen 5

Selección de Tau-c en tabla cruzada





Fuente: Elaboración propia

Ejemplos

Un grupo de estudiantes de enfermería realizaron una recolección de datos en personas adultas mayores residentes de una comunidad vulnerable. Entre los datos que recolectaron se encontraban las siguientes variables: estado nutricional (estado de peso adecuado, riesgo de desnutrición y desnutrición) que fue medido con valoración breve del estado nutricional (Mini Nutricional Assessment); riesgo de úlceras por presión (bajo, moderado y alto) empleando la escala de Braden; estado cognitivo (adecuado y probable deterioro cognitivo) con la evaluación cognitiva de Montreal (MoCA); y síntomas de depresión (sin síntomas de depresión, síntomas depresivos leves, síntomas depresivos moderados y síntomas depresivos graves) con la escala GDS-15.

Los estudiantes tenían las siguientes hipótesis:

H1. Existe asociación entre el estado nutricional y el riesgo de úlceras por presión en las personas adultas mayores.

H2. El estado cognitivo se asocia con los síntomas de depresión en las personas adultas mayores.

Para iniciar su proceso de análisis realizaron la prueba de normalidad y determinaron que no existía distribución normal en las variables por lo que consideraron aplicar los coeficientes Tau-b de Kendall para las variables estado nutricional y riesgo de úlceras por presión y Tau-c de Kendall para el estado cognitivo y los síntomas de depresión.

Los resultados indican que el estado nutricional y el riesgo de úlceras por presión se asocian significativamente ($p < .01$) con una fuerza alta y una dirección positiva ($\tau = .767$), lo que indicaría que cuanto mayor sea la desnutrición mayor será el riesgo de úlceras por presión en las personas adultas mayores (tabla 4).

Tabla 4
Resultado del análisis con la prueba Tau-b de Kendall

		Medidas simétricas			
		Valor	Error estándar asintótico	Aprox. Sb	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	.767	.074	11.523	.000
N de casos válidos		30			

- a. No se supone la hipótesis nula.
b. Utilización del error estándar asintótico que asume la hipótesis nula.
Fuente: Elaboración propia

En la tabla cinco, se presentan los resultados de la prueba de asociación entre el estado cognitivo (adecuado y probable deterioro cognitivo) y los síntomas de depresión (sin síntomas, síntomas leves, síntomas moderados, síntomas graves).

Tabla 5
Resultado del análisis con la prueba Tau-c de Kendall

		Medidas simétricas			
		Valor	Error estándar asintótico	Aprox. Sb	Aprox. Sig.
Ordinal por ordinal	Tau-c de Kendall	.898	.067	13.453	.000
N de casos válidos		30			

- a. No se supone la hipótesis nula.
b. Utilización del error estándar asintótico que asume la hipótesis nula.
Fuente: Elaboración propia

Los resultados indican que el estado cognitivo y los síntomas de depresión se asocian significativamente ($p < .01$) con una fuerza alta ($\tau = .898$) y una dirección positiva, lo que indicaría que a mayor presencia deterioro cognitivo mayores serán los síntomas de depresión.

Referencias

- BADII, M., Guillen, O., Lugo, O. & Aguilar, J. (2014). Correlación No-Paramétrica y su Aplicación en la Investigaciones Científica. *International Journal of Good Conscience* Agosto, 9(2), 31–40.
- BOLBOACA, S.D. & Jäntschi, L. (2006). Pearson versus Spearman, Kendall's Tau Correlation Analysis on Structure-Activity Relationships of Biologic Active Compounds. *Ljş*, 9, 179–200.
- COHEN, J. W. (1988). *Statistical power analysis for the behavioral sciences* (Second). Lawrence Erlbaum Associates.
- GIBBONS, J. D., & Chakraborti, S. (2010). Nonparametric statistical inference. In *Nonparametric Statistical Inference*. https://doi.org/10.5005/jp/books/10313_14

- GÖKTAŞ, A., & IŞÇI, Ö. (2011). A comparison of the most commonly used measures of association for doubly ordered square contingency tables via simulation. *Metodoloski Zvezki*, 8(1), 17–37. <https://doi.org/10.51936/milh5641>
- INTERNATIONAL Business Machines. (2021a). *Estadísticos de tablas cruzadas*. Estadísticos de Tablas Cruzadas. <https://www.ibm.com/docs/es/spss-statistics/25.0.0?topic=crosstabs-statistics>
- INTERNATIONAL Business Machines. (2021b). *Tablas cruzadas*. Tablas Cruzadas. <https://www.ibm.com/docs/es/spss-statistics/25.0.0?topic=features-cross-tabs>
- LEYVA, C. O. & FLORES, H. M. (2014). Capítulo 19. Análisis de correlaciones bivariadas y parciales con SPSS. In G. G. Sáenz, L. K & Tamez (Ed.), *Métodos Y Técnicas Cualitativas Y Cuantitativas Aplicables a La Investigación En Ciencias Sociales* (Tirant Hum, pp. 421–436). <http://eprints.uanl.mx/8625/1/7>. Metodología cap 19 solo capítulo.pdf
- MORALES, P. & RODRÍGUEZ, L. (2016). *Aplicación de los coeficientes correlación de Kendall y Spearman*. 1–23.

Capítulo 10.

Prueba de hipótesis paramétrica. Prueba T para muestras independientes. Prueba T para muestras relacionadas

DSP JOSUÉ ARTURO MEDINA FERNÁNDEZ
DCS NISSA YAING TORRES SOTO

Definición de la prueba

La prueba T o también conocida como prueba T de Student, es una prueba paramétrica que se emplea para determinar la significatividad de la diferencia entre las medias de dos muestras, cuando estas no son amplias se dice que contrasta con la prueba Z. La prueba T demuestra que las muestras poseen una distribución normal y varianzas similares, y trabaja con los valores de las medias y con el error típico de la diferencia entre estas medias. Si el valor calculado de la prueba T es mayor o igual al valor crítico, la diferencia entre las medias será significativa, es decir, será suficiente para aceptar la hipótesis del investigador, es decir, que las medias de sus poblaciones respectivas son diferentes (Palacios et al., 2019). Por lo que, para usar esta prueba, previamente se asume que:

- Los datos son continuos
- La muestra de datos se seleccionó aleatoriamente de la población estudiada
- Hay homogeneidad en la varianza (es decir, la variabilidad de datos de cada grupo es similar)
- La distribución es aproximadamente normal
- Para pruebas t de dos muestras, se debe tener muestras independientes, si estas no son independientes, se deberá considerar las pruebas t pareada

Pero, para poder entender un poco más esta prueba estadística, es necesario explicar que es una muestra independiente y relacionada (o pareada). La primera es aquella que compara las medias de dos grupos de casos. Lo ideal es que para esta prueba los participantes se asignen aleatoriamente a dos grupos, de forma que cualquier diferencia en la respuesta sea debida al tratamiento o falta de tratamiento, y no a otros factores externos. Mientras que la segunda se aborda en aquellas que poseen dos variables o poblaciones de medición, pero dependientes (en una misma muestra), pudiéndose aplicar como ejemplo el hecho si deseas verificar el efecto de un nuevo método de curación de enfermería, por lo que es necesario que se prueba en un grupo de personas y se miden los efectos antes y después (Quintela, 2019).

Supuestos

La prueba T entre sus principales usos, busca comparar la media con un valor hipotético u objetivo utilizando una prueba t de una muestra, sin embargo, también se puede comparar las medias de dos grupos con una prueba t de dos muestras y si tiene dos grupos con observaciones pareadas como lo son las mediciones antes y después, se utiliza la prueba t pareada (Minitab, 2019). La siguiente tabla resume las características de cada una de las pruebas:

Tabla 1
Comparación de tipos de pruebas T

	Prueba T de una muestra	Prueba T de dos muestras independientes	Prueba T pareada o relacionadas
Sinónimos	Prueba t de Student	Prueba t de grupos independientes Prueba t de muestras independientes Prueba t de varianzas iguales Prueba t de varianzas desiguales	Prueba t de grupos pareados Prueba t de muestras dependientes
Número de variables	Uno	Dos	Dos
Tipo de variable	Medida continua	Medida continua, escalor o numérica. Ordinal dicotómica o nominal dicotómica para definir grupos	Medida continua, escalor o numérica. Ordinal dicotómica o nominal dicotómica para definir pares en un grupo
Objetivo de la prueba	Determinar si la media de población es igual o no que un valor específico	Determinar si las medias de población de dos grupos distintos son o no iguales	Determinar si la diferencia entre medidas pareadas de una población es o no igual a cero
Ejemplo: comprobar si...	El porcentaje de masa muscular de un grupo de personas adultas mayores es igual al 30% o no	La velocidad de la marcha de dos grupos de personas adultas mayores es la misma o no	La diferencia media de frecuencias cardíacas de un grupo de personas antes y después de hacer ejercicio es cero o no

Fuente: (Minitab, 2019)

Hipótesis

Para una prueba t para varianzas iguales, las observaciones deben ser muestras aleatorias independientes de una distribución normal con la misma varianza de población. Para una prueba t de varianzas desiguales, las observaciones deben ser muestras aleatorias independientes de una distribución normal. Una prueba t de 2 muestras es muy resistente a la desviación de la normalidad. Al contrastar las distribuciones gráficamente, compruebe que son simétricas y que no contienen valores atípicos (Guerra, 2014).

En otras palabras, para muestras grandes, las pruebas t suelen ser válidas incluso cuando se viola el supuesto de normalidad. Esta propiedad lo convierte

en uno de los métodos más útiles para estimar las medias poblacionales. Sin embargo, con un tamaño de muestra pequeño y distribuciones no normales y muy asimétricas, podría ser más conveniente usar pruebas no paramétricas (García, 2011).

Dicho de otra manera, para poder realizar una hipótesis estadística para esta prueba es necesario que este se observe la variable estudiada, el grupo o grupos. Aquí algunos ejemplos:

Hipótesis para prueba T de una muestra

H_i = La puntuación para riesgo de sarcopenia es diferentemente significativo entre las mujeres adultas mayores de Chetumal en el periodo 2013-2023.

Hipótesis para prueba T de dos muestras independientes

Previamente es necesario hacer una prueba de igualdad de varianzas: cuando realice una prueba de hipótesis de dos muestras independientes debe realizar primero la prueba de igualdad de varianzas.

H_o : Las varianzas son iguales

H_a : Las varianzas son diferentes

Recuerde elegir su alfa antes de realizar la prueba de hipótesis.

Decisión: rechazar H_o si el valor de calculado es menor que el valor de p seleccionado.

H_i = El autocuidado es diferentemente significativo entre el grupo de adultos mayores de hombres y mujeres.

Hipótesis para prueba T pareada

H_i = El efecto de la adopción del rol es diferentemente significativo entre la preprueba y post prueba en los cuidadores de adultos mayores.

Interpretación

El valor t de Student (para muestras independientes) está asociado con la probabilidad (p) del valor t obtenido (es decir, significación estadística). Ese valor de p es la probabilidad de obtener una diferencia distinta de cero si ambas muestras pertenecen a la misma población y no son más que diferencias aleatorias (no es la probabilidad de obtener la misma diferencia en comparaciones similares) (Wayne, 2002).

Dado que nuestro nivel de confianza es $p = 0,05$, la diferencia es estadísticamente significativa siempre que este valor de p sea menor o igual a 0,05. En este caso se fija el valor exacto de la probabilidad, basta con fijar $p < 0,05$ o $p > 0,05$, si este es nuestro nivel de confianza, como es práctica habitual. (González et al., 2017).

También es usual utilizar tres valores de referencia para afirmar las diferencias: 0.05 (5%), 0.01 (1%) y 0.001, pero basta que el nivel de significancia sea inferior a 0.05. El valor t asociado a su probabilidad (p) nos da confianza de que podemos confirmar la diferencia. Lo que el valor t no nos dice es si la diferencia

es grande o pequeña, solo nos da la probabilidad. Con muestras grandes, es fácil lograr diferencias estadísticamente significativas que pueden ser muy pequeñas. Un resultado estadístico significativo no quiere decir que sea relevante o importante (Wayne, 2002).

Si el valor de t nos remite a una probabilidad mayor de 0.5 ($p > 0.05$) pero tampoco es una probabilidad muy grande (ejemplo $p = 0.06$) y las muestras comparadas son más bien pequeñas, no podemos afirmar la diferencia (con nuestro nivel de confianza), pero si se puede proponer como hipótesis, que, comparando muestras mayores, es probable obtener una diferencia estadísticamente significativa, sobre todo si la magnitud de la diferencia (tamaño del efecto) es apreciable (González et al., 2017).

Finalmente, el tamaño del efecto se denota con la letra d , el signo del tamaño del efecto es indiferente, pero hay que tenerlo en cuenta en la interpretación. Para valorar la magnitud del tamaño del efecto (de la diferencia entre las medias) suelen tenerse en cuenta las orientaciones de Cohen 1988, donde menciona que $d = 0.20$ tiene una diferencia pequeña, $d = 0.50$ tiene una diferencia moderada y $d = 0.80$ tiene una diferencia grande (Palacios et al., 2019).

Ruta SPSS

Para obtener una prueba T para una muestra (SPSS Statics, 2021)

1. Elija en los menús: Analizar > Comparar medias > Prueba T para una muestra.
2. Seleccione una o más variables para contrastarlas con el mismo valor hipotetizado.
3. Introduzca un valor de contraste numérico para compararlo con cada media muestral.
4. También, puede pulsar en Opciones para controlar el tratamiento de los datos perdidos y el nivel del intervalo de confianza.

Para obtener una prueba T para dos muestras independientes (SPSS Statics, 2021)

1. Elija en los menús: Analizar > Comparar medias > Prueba T para muestras independientes.
2. Seleccione una o más variables de contraste cuantitativas. Se calcula una prueba t para cada variable.
3. Seleccione una sola variable de agrupación y pulse en Definir grupos para especificar dos códigos para los grupos que desee comparar.
4. También puede pulsar en Opciones para controlar el tratamiento de los datos perdidos y el nivel del intervalo de confianza.

Para las variables de agrupación de cadena, escriba una cadena para el Grupo 1 y otra para el Grupo 2; por ejemplo, sí y no. Los casos con otras cadenas se excluyen del análisis.

Definir grupos en una prueba T para muestras independientes (SPSS Statics, 2021)

1. Seleccione en los menús: Analizar > Comparar medias > Prueba T para muestras independientes.
2. Seleccione una variable de agrupación y, a continuación, pulse en Definir grupos.
3. Escriba los valores para definir los grupos.

Para obtener una prueba T para muestras relacionadas (pareada) (SPSS Statics, 2021)

1. Elija en los menús: Analizar > Comparar medias > Prueba T para muestras relacionadas.
2. Seleccione uno o más pares de variables
3. También puede pulsar en Opciones para controlar el tratamiento de los datos perdidos y el nivel del intervalo de confianza.

Ejercicio

Prueba T para muestras independientes

1. Un profesor de la Universidad Autónoma del Estado de Quintana Roo desea realizar una investigación cuyo objetivo se centra en determinar la diferencia de medias de la autotrascendencia, autocuidado y funcionalidad familiar en grupos de adultos mayores con y sin enfermedades crónicas. En este participaron 240 adultos mayores con una igualdad en grupos, por lo cual se solicita que se interprete la siguiente tabla posterior al seguimiento de la ruta en el SPSS.

Tabla 2
Diferencia de medias de la autotrascendencia, autocuidado y funcionalidad familiar en adultos mayores con y sin enfermedades crónicas

Variable	AM con enfermedad		AM sin enfermedad		t	p	d
	crónica		crónica				
	M	DE	M	DE			
Autotrascendencia	47.06	6.204	49.40	6.335	4177.000	0.008	0.64
Funcionalidad familiar	7.37	8.000	7.55	2.853	5029.000	0.377	0.001
Autocuidado	87.11	13.886	91.62	12.762	4365.500	0.024	0.55

Nota: AM= Adulto mayor, M= Mediana, DE= Desviación estándar, t= t de student para muestras independientes, p=nivel de significancia, d= tamaño de efecto.

RESPUESTA

En la tabla 2 se realizó la prueba de T de student para muestras independientes, se encontró diferencia de medias significativa de la autotranscendencia ($p=0.008$) y autocuidado ($p=0.024$), observándose las medias mayores en el grupo de adultos mayores sin enfermedades crónicas, teniendo un tamaño de efecto moderada para ambas variables.

Prueba T para muestras relacionadas

2. Una profesora de la Universidad Autónoma de Quintana Roo desea comprobar si la intervención educativa en salud en un grupo de 40 mujeres que viven con diabetes en la comunidad mejora los niveles de glucosa, malestar emocional y empoderamiento de la diabetes, por lo cual realizó una medición pretest y postest mediante la aplicación de escalas. Se analizó en el programa de SPSS, resultando la siguiente tabla. Interpreta los resultados.

Tabla 3
Diferencia de medias de una intervención educativa en un grupo de mujeres que viven con diabetes de la comunidad

Variable	Pretest	Postest	t	p	d
	M ± DE	M ± DE			
Glucosa	178.70± 5.656	156.31± 2.076	15891.500	0.014	0.52
Malestar emocional	31.98±8.789	31.72±7.677	18098.000	0.657	0.04
Empoderamiento en diabetes	110.50±4.432	113.84±2.654	16085.500	0.023	0.63

Nota: M= Mediana, DE= Desviación estándar, t= t de student para muestras relacionadas, p=nivel de significancia, d= tamaño de efecto.

RESPUESTA

En la tabla 3 se reportó una diferencia de medias significativa entre el pretest y postest de la glucosa y el empoderamiento en diabetes en un grupo de mujeres de la comunidad, teniendo un tamaño de efecto moderado, analizándose mediante las medias que los niveles de glucosa disminuyeron y el empoderamiento de la diabetes mejoró.

Referencias

- GARCÍA, A. (2011). *Elementos de bioestadística* (Universidad de Extremadura, Ed.; Tercera, Vol. 1).
- GONZÁLEZ, F., Escoto, M., & Chávez, J. (2017). *Estadística aplicada en psicología y ciencias de la salud* (El manual moderno, Ed.; Primera edición, Vol. 1).
- GUERRA, T. (2014). *Bioestadístico* (UNAM-FEIS Zaragoza, Ed.; Primera, Vol. 1).
- MINITAB. (2019). *Entendiendo las pruebas t: Valores t y distribuciones t*. Minitab Blog .
- PALACIOS, I., Alonso, R., Cal, M., & Fernández, F. (2019). *Diccionario electrónico de enseñanza y aprendizaje de lenguas* (L. Gómez, Ed.; 1st ed., Vol. 1).
- QUINTELA, A. (2019). *Estadística Básica Edulcorada* (1st ed., Vol. 1).
- SPSS Statics. (2021). *Prueba T*. IBM . <https://www.ibm.com/docs/es/spss-statistics/25.0.0?topic=tests-independent-samples-t-test>
- WAYNE, D. (2002). *Bioestadística base para el análisis en ciencias de la salud* (John Wiley & Sons, Ed.; Cuarta, Vol. 1).

Capítulo 11

Prueba de hipótesis paramétrica. Prueba T para una muestra

Análisis de varianza de un factor (ANOVA)

DCE JUANA EDITH CRUZ QUEVEDO
DCE MARCELA TENORIO-GÓMEZ

Definición de la prueba ANOVA

Análisis de Varianza (ANOVA) se refiere al procedimiento estadístico paramétrico que valora la diferencia entre las medias de una variable continua (dependiente) cuando hay tres o más grupos o condiciones de un mismo Factor (variable independiente); es decir, compara la variabilidad de los puntajes *entre* los diferentes grupos (supuestamente atribuible a la variable independiente) con la variabilidad *dentro* de cada grupo (en estadística se usa el término “*intragrupos*”).

Supuestos

En general los supuestos que se presentan para ANOVA son similares a todas las pruebas paramétricas; sin embargo, deben puntualizarse a fin de evitar lo más posible sesgos en los resultados. Los supuestos más importantes se refieren al nivel de medición de las variables, la distribución de los datos, el tipo de muestreo que se usa para la recolección de datos, independencia de las observaciones y homogeneidad de la varianza.

Nivel de medición de las variables. Respecto al nivel de medición de las variables, en ANOVA se asume que la variable dependiente (en la que se buscarán las diferencias de medias) es una medición continua (intervalo o razón). El uso de variables continuas en investigación garantiza mayores posibilidades de manejo, por lo que es recomendable incluir este tipo desde el diseño de los estudios; así por ejemplo es mejor tener la variable edad en forma continua (años) y de ser necesario poder después categorizar si es necesario o la evidencia empírica lo sugiere. Mientras que la variable independiente se proponga como categórica (tres o más categorías).

Distribución de los datos. Este es uno de los supuestos básicos de toda prueba paramétrica en la que se asume que los datos se distribuyen normalmen-

te en la muestra por lo que es recomendable tener siempre un análisis de normalidad como por ejemplo la prueba de Kolmogorov Smirnov. No obstante, en el caso de la disciplina de enfermería (y en la mayoría de las disciplinas sociales) puede encontrarse frecuentemente una distribución no normal de la variable dependiente y en estos casos se debe recurrir a la valoración de robustez de los datos que permita hacer uso del procedimiento para que sea posible sobrepasar este supuesto sin mayor problema. La robustez de la variable se da con muestras grandes (Cohen, 1988), generalmente consideradas mayor de 30.

Muestreo en la recolección de los datos. Idealmente los datos para pruebas paramétricas, incluyendo la prueba ANOVA, deben surgir a partir de muestreos aleatorios. Aunque existe evidencia que es frecuente la subrepresentación de este tipo de muestreo en los estudios (Williamson, 2003), es un tema que debe considerarse cautelosamente para evitar sesgos metodológicos.

Independencia de las observaciones. Las observaciones de los datos deben forzosamente ser independientes entre los grupos por lo que es necesario evitar la contaminación cruzada en el caso de, por ejemplo, grupos de intervención cuando ésta es la variable independiente que se ve involucrada en el análisis. Asimismo, cuando el análisis ANOVA se realiza asumiendo una variable independiente teóricamente construida (por ejemplo, estado civil: casado, viudo, divorciado; estado de salud percibido: malo, regular, bueno) es importante garantizar que las categorías sean mutuamente excluyentes.

Homogeneidad de la varianza. En las pruebas paramétricas como lo es ANOVA se acepta que las poblaciones de las cuales emergen las muestras tienen varianzas iguales (homogéneas), lo que significaría que la variabilidad de los datos se conduce en forma similar en cada grupo. En el caso del análisis un factor ANOVA con el programa SPSS, el proceso permite identificar si se cumple con este supuesto a través de la prueba de Levene por lo que previo al reporte de los resultados debe interpretarse.

Hipótesis a probar en ANOVA

La hipótesis a probar en la prueba ANOVA es que existe diferencia de medias de una variable (dependiente) entre tres (o más) grupos, denominado para la prueba Factor, que se encuentra representado por categorías. Por ejemplo, tipos de tratamiento en una intervención (modalidad de tratamiento A, modalidad de tratamiento B, Grupo control) o bien diferencias por categorías de una variable (estado civil: soltero, viudo, divorciado o casado). La hipótesis de investigación se representaría de la siguiente manera:

$$H_1 = \mu_a \neq \mu_b \neq \mu_c$$

Es importante aclarar que en estadística se trabaja con la hipótesis nula que en este caso establecería que no existe diferencia entre los grupos o que las medias de los grupos son iguales. Se representa como:

$$H_0 = \mu_a = \mu_b = \mu_c$$

El estadístico resultante para ANOVA es el **valor F**, también reportado como estadístico F o razón F (F-statistic, F ratio). El valor F representa justamente la varianza entre los grupos dividida entre la varianza intra grupos. Cuando el resultado del valor F es suficientemente grande, permite la significancia por lo que se interpreta como la existencia de mayor variabilidad entre los grupos que la que existe entre cada grupo con lo que es posible rechazar la hipótesis nula y aceptar la hipótesis de investigación.

Un dato adicional para considerar al probar la hipótesis es el **tamaño de efecto** (o fuerza de la asociación) que permite no solo confirmar la significancia de la diferencia sino además valorar la magnitud con la que las dos variables son asociadas una de la otra. Aunque existen diferentes formas para calcular e interpretar el tamaño de efecto, la mayoría de la literatura en estadística e investigación consideran la propuesta por Cohen (1988), denominada *Eta Cuadrada*, como la más clásica y confiable. Cohen considera el tamaño de efecto como pequeño (.01 o 1%), mediano (.06 o 6%) y grande .138 o (13.8%). En algunos procesadores estadísticos el cálculo de Eta Cuadrada no es automático, así es por ejemplo para el caso del cálculo de ANOVA en SPSS, sin embargo, el valor de Eta Cuadrada puede calcularse fácilmente pues es el resultado de la división entre la suma de cuadrados entre grupos entre el total de la suma de cuadrados; puede representarse de la siguiente manera:

$$\text{Eta Cuadrada} = \frac{\text{Suma de cuadrados entre grupos}}{\text{Total de la suma de cuadrados}}$$

Ruta en SPSS

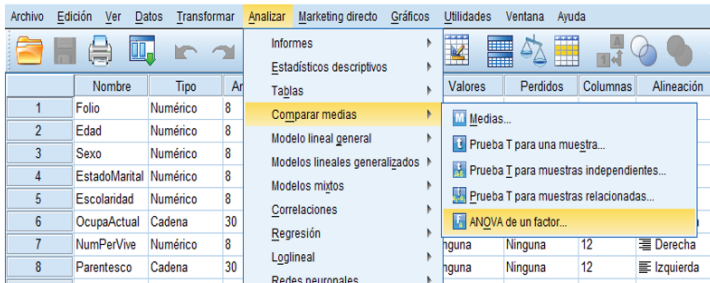
Una vez considerados los elementos esenciales y con los cuidados previos de la base como por ejemplo limpieza de la misma, identificación, y en su caso, tratamiento de valores perdidos, prueba de normalidad en los datos, entre otros, es tiempo de proceder al análisis inferencial propuesto: Análisis de Varianza entre grupos de un factor (One-way between-groups ANOVA). Para guiar el proceso, procederemos a plantearnos el siguiente caso:

Se está desarrollando una investigación donde se plantea la pregunta: ¿Existen diferencias en los puntajes del estado nutricional de acuerdo al estado de salud percibido en adultos mayores? En este caso la propuesta es que el estado de salud percibido juegue como variable independiente (categórica con tres niveles: 1= malo, 2= regular, 3= bueno) mientras que el estado nutricional se

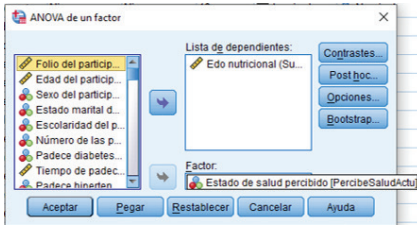
propone como variable dependiente, se valoró con la Valoración Mínima de Estado Nutricional (MNA, por sus siglas en inglés) considerando para el análisis, la sumatoria de esta valoración (continua).

Para acceder al análisis se deben seguir los siguientes pasos:

1. Identificar en la pestaña “Analizar” en el menú superior del programa, la opción “comparar medias” y en el menú desplegado seleccionar la prueba “ANOVA de un factor”



2. La pantalla resultante será la siguiente:

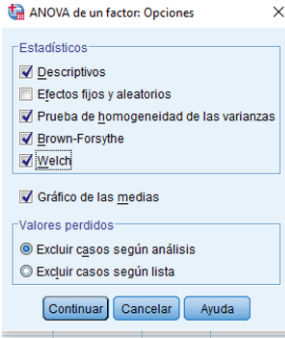


Del listado de variables, seleccionar la variable que fungirá como variable dependiente (Edo nutricional, en este ejemplo) y la trasladamos al recuadro titulado “Lista de dependientes” mientras que en el recuadro denominado “Factor”, trasladamos la variable que se comportará como variable independiente (en este caso el Estado de salud percibido). Nótese el nivel de medición de las variables, continua para la variable dependiente y categórica para la variable independiente, identificadas por los íconos de la regla y la figura de conjuntos respectivamente.

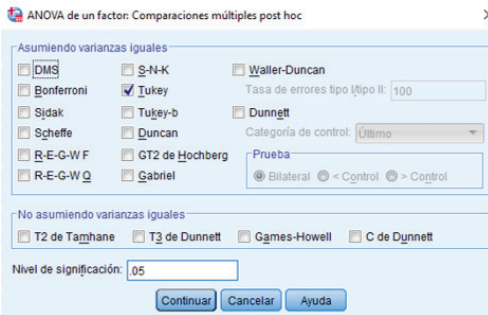
Hasta este momento, tenemos solo las variables; es importante detallar algunos aspectos que permitirán asegurarnos de cumplir con los supuestos de las pruebas, así como obtener datos precisos para una buena interpretación por lo que procederemos a identificar las opciones y especificar en el apartado Post Hoc

3. Seleccione el botón “opciones”. En la pantalla resultante asegúrese seleccionar en el recuadro de Estadísticos las opciones: descriptivos,

prueba de homogeneidad de las varianzas, Brown-Forsythe, Welch y gráfico de medias. En el recuadro de Valores perdidos, asegúrese que este seleccionada la opción de “Excluir casos según la lista”. Dé clic en continuar.



4. Posteriormente lo regresa a la pantalla inicial, ahora seleccione el botón señalado con la leyenda “Post hoc”. En este paso seleccione en el recuadro “Asumiendo varianzas iguales” la opción “Tukey” y establezca el nivel de significación (en este caso se estableció un nivel de .05).



5. Dé clic en continuar y lo regresará a la pantalla inicial.
6. Finalmente, de clic en “Aceptar”

Presentación de resultados e interpretación de ANOVA

En el archivo resultados se desplegarán los mismos de la siguiente manera:

Primero se identifica la sintaxis generada, en ella se debe verificar que los comandos hayan sido los que se tienen planeados previamente, como por ejemplo el valor de nivel de significancia.

```

ONEWAY SumMNA BY PercibeSaludActu
/STATISTICS DESCRIPTIVES HOMOGENEITY BROWNFORSYTHE WELCH
/PLOT MEANS
/MISSING ANALYSIS
/POSTHOC=DUKEY ALPHA (.05) .
    
```

Posteriormente se presentan los datos descriptivos y prueba de homogeneidad.

ANOVA de un factor

Descriptivos

Edo nutricional (Sumatoria de MNA)

	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo
					Límite inferior	Límite superior		
Malo	5	22.7000	2.16795	.96954	20.0081	25.3919	20.00	25.00
Regular	60	22.8583	3.71790	.47998	21.8979	23.8188	10.00	28.50
Bueno	46	25.0870	2.76748	.40804	24.2651	25.9088	15.00	28.50
Total	111	23.7748	3.45671	.32810	23.1246	24.4250	10.00	28.50

Es importante observar en la tabla de datos descriptivos cada uno de los datos que presenta: el número de casos en cada uno de los grupos (N) la media de la variable dependiente (Estado nutricional) en cada grupo, su desviación estándar, el error típico, el intervalo de confianza y los valores mínimo y máximo. Aquí se verifica que los datos de tamaño de los grupos sean correctos y los datos descriptivos restantes permiten tener una idea del comportamiento de la variable entre los grupos.

Enseguida es necesario verificar la homogeneidad de las varianzas. Recordemos que un supuesto importante para esta prueba es que las muestras tienen varianzas iguales (homogéneas), lo que significaría que la variabilidad de los datos se conduce en forma similar en los grupos. Para ello la prueba de Levine prueba que los puntajes de varianza son iguales en los grupos; por lo que en esta prueba **una significancia mayor a .05 indica que no se ha violado este supuesto** y que efectivamente existe homogeneidad entre los grupos.

Prueba de homogeneidad de varianzas

Edo nutricional (Sumatoria de MNA)

Estadístico de Levene	g1	g2	Sig.
1.496	2	108	.229

Posteriormente se muestran los cuadros de la prueba titulado ANOVA, en el que se muestran los resultados de la suma de cuadrados, grados de libertad (gl), media cuadrática, el valor de F y la significancia (sig).

ANOVA

Edo nutricional (Sumatoria de MNA)

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	135.371	2	67.686	6.200	.003
Intra-grupos	1178.998	108	10.917		
Total	1314.369	110			

Este es el cuadro principal de resultados de la prueba por lo que aquí se establece primero la significancia de la prueba. **Un valor de significancia igual o menor a .05 indica que existen diferencias significativas entre las medias de los grupos propuestos** y como ya se planteó anteriormente un valor de F “grande” representa justamente esa varianza. Los grados de libertad con los que se maneja la prueba son dos, inter grupos ($gl= 2$) e intra-grupos ($gl=108$), por lo que para el valor de F se reportarían los grados de libertad como 2, 108.

Otro dato importante que debe verse en este cuadro es el tamaño de efecto, el valor Eta no viene descrito específicamente, pero puede calcularse con la fórmula ya descrita previamente. En este ejemplo, el cálculo del tamaño de efecto resultaría:

$$\text{Eta Cuadrada} = \frac{\text{Suma de cuadrados entre grupos}}{\text{Total, de la suma de cuadrados}}$$

$$\text{Eta Cuadrada} = \frac{135.371}{1314.369} = 0.10$$

El tamaño de efecto resultante fue de 0.10 que de acuerdo a la clasificación de Cohen (1988) indica un efecto mediano (mayor de .06 pero menor a .13).

También se observa el recuadro donde muestra los resultados de las pruebas de robustez de igualdad.

Pruebas robustas de igualdad de las medias

Edo nutricional (Sumatoria de MNA)

	Estadístico ^a	gl1	gl2	Sig.
Welch	6.900	2	12.286	.010
Brown-Forsythe	8.834	2	38.080	.001

a. Distribuidos en F asintóticamente.

Estos valores se utilizan cuando la prueba de Levine resultó contraria a lo esperado y por lo tanto el valor de ANOVA puede resultar no confiable. Es decir, los valores Welch y Brown-Forsythe son preferibles cuando el supuesto de homogeneidad no se cumple (Glanz et al., 2016). Para este ejemplo, dado que, si se cumplió con el supuesto demostrado con la prueba de Levine, no se considerarán los resultados de Welch y Brown-Forsythe.

Finalmente se presentan las pruebas Post Hoc que presenta los resultados para las comparaciones múltiples.

Pruebas post hoc

Comparaciones múltiples

Edo nutricional (Sumatoria de MNA)
HSD de Tukey

(I) Estado de salud percibido	(J) Estado de salud percibido	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95%	
					Límite inferior	Límite superior
Malo	Regular	-.15833	1.53794	.994	-3.8132	3.4965
	Bueno	-2.38696	1.55584	.279	-6.0844	1.3104
Regular	Malo	.15833	1.53794	.994	-3.4965	3.8132
	Bueno	-2.22862*	.64750	.002	-3.7674	-.6899
Bueno	Malo	2.38696	1.55584	.279	-1.3104	6.0844
	Regular	2.22862*	.64750	.002	.6899	3.7674

*. La diferencia de medias es significativa al nivel .05.

La prueba Post hoc muestra las comparaciones múltiples entre los grupos. Es en estas comparaciones que se visualiza exactamente en donde están las diferencias. En este ejemplo podemos ver que la única diferencia significativa está entre los grupos regular y bueno con una diferencia de medias significativas de .002.

Reporte de los resultados

Con frecuencia se encuentra difícil el redactar los resultados, una orientación precisa es considerar siempre que la redacción debe ser lo suficientemente clara para dar respuesta a la hipótesis y/o pregunta planteada y que deben mostrarse los datos suficientes que faciliten la comprensión del resultado y orienten la congruencia lógico-teórica de la investigación. A manera de ejemplo, se redacta a continuación el reporte que derivaría del ejemplo planteado.

A fin de explorar la diferencia entre el estado nutricional de acuerdo al Edo de salud percibido se realizó un ANOVA de un factor. El estado de salud percibido se propuso como el factor de análisis dividido en tres grupos (1: malo, 2: regular y 3: bueno). Los resultados muestran una diferencia estadísticamente significativa al nivel propuesto de <.05 en el estado nutricional entre los tres grupos ($F_{(2, 108)} = 6.20; p = .003$). El tamaño de efecto entre las medias de los grupos fue mediano ($\eta^2 = 0.10$). El análisis de comparaciones múltiples Post-hoc a través de la prueba de Tukey HSD indican que las diferencias significativas ($p = .002$) estuvieron entre las medias del grupo bueno ($M = 25.08, DE = 3.45$) y el grupo regular ($M = 22.85, DE = 3.71$). El grupo de estado de salud percibido como malo ($M = 22.70, DE = 2.16$) no difirió significativamente de los que percibieron su estado de salud como regular o malo.

Referencias

- COHEN, J.W. (1988). *Statistical Power analysis for the behavioral sciences* (2nd Ed). Lawrence Erlbaum Associates.
- GLANZ, S.A., Slinker, B.K., & Neilands, T.B. (2016). *Primer of Applied Regression & Analysis of Variance*. 3rd Edition. McGraw Hill.
- WILLIAMSON, G. R. (2003). Methodological issues in nursing research: Misrepresenting random sampling? A systematic review of research papers in the *Journal of Advanced Nursing*. *Journal of Advanced Nursing* (Wiley-Blackwell), 44(3), 278–288. <https://doi.org.ezproxy.uv.mx/10.1046/j.1365-2648.2003>.

Capítulo 12.

Prueba de hipótesis no paramétrica. Prueba H de Kruskal Wallis. Prueba U de Mann-Whitney. Rangos de Wilcoxon

DSP ISAÍ ARTURO MEDINA FERNÁNDEZ
DCE DAFNE ASTRID GÓMEZ MELASIO

La comparación de medias abarca la comparación una variable continua (variable de razón) que se puede resumir en dos o más categorías (variables dicotómicas, nominales o politómicas) y que se engloba dentro de las pruebas para datos independientes, así como la comparación de los valores de una variable que es medida continua evaluada en dos o más momentos en el tiempo (por ejemplo dos mediciones pre y post intervención)(Moral, 2016)entre otras.

Existen diversas pruebas estadísticas que comparan las medias de una variable entre dos o más grupos, esta dependerá si cumple con ciertos supuestos para su aplicación. Así mismo, en todas la hipótesis de comparación que se desea evaluar se puede analizar mediante pruebas paramétricas o no paramétricas. En el caso de las pruebas no paramétricas puede ser mediante U de Mann-Whitney para comparar dos grupos y Prueba H de Kruskal Wallis para más de dos grupos para comparar mediciones dependientes.

U de Mann-Whitney

Definición de la prueba

La prueba de t U Mann Whitney es un criterio estadístico utilizado para evaluar diferencias entre dos muestras independientes asumiendo que los datos no siguen una distribución normal (Grishaev & Гришаев, 2015). Esta prueba tiene su base en la diferencia de rango y es la contraparte de la t de Student que se emplea en las variables cuantitativas con distribución normal (Rivas-Ruiz et al., 2013).

Supuestos

Los siguientes supuestos deben aplicarse para poder realizar esta prueba estadística (IBM Corporation, 2021; Landeros & González, 2009).

- Las dos muestras son independientes

- La variable observada es una variable aleatoria continua
- Comprueba diferencias entre dos distribuciones
- Las dos muestras de tamaños se utilizan para el análisis independiente y al azar de sus poblaciones respectivas
- Si las poblaciones son diferentes, difieren solo en lo que respecta a sus medianas
- Las funciones de distribución de las dos poblaciones diferente, si lo hacen, solo respecto de la posición de la mediana

Hipótesis

La prueba U de Mann-Whitney supone que las dos variables son independientes. Si los datos son independientes, entonces las observaciones recopiladas de un grupo no influirá ni tendrá ningún efecto en el segundo grupo de muestreo (Weaver et al., 2018). Referido lo anterior cuando se satisfacen las suposiciones antes referidas (Milton, 2007), puede probarse que:

1. Las poblaciones no tienen medianas iguales
2. La mediana de la población 1 es mayor que la median de la población 2
3. La mediana de la población 1 es menor que la mediana de la población 2

Comenzamos formulando nuestras hipótesis: las hipótesis comparan la mediana de población A con la de la población B y evalúan las diferencias entre las poblaciones, la hipótesis nula (H_0) establece que las dos muestras provienen de la misma población, lo que no indica diferencia (H_0): Los rangos de A = los rangos de B.

Ejemplo: “No existe diferencia significativa de la calidad de vida relacionada con la salud en adultos mayores con y sin sarcopenia”

Por otra parte, la hipótesis alternativa (H_1) establece que las dos muestras independientes son de diferentes poblaciones. Los rangos de A > los rangos de B o el rango de A < los rangos de B (Weaver et al., 2018).

Ejemplo: “La calidad de vida relacionada con la salud en adultos mayores es diferentemente significativamente con el grupo de sarcopenia comparado a los que no tiene sarcopenia”

Interpretación

Al realizar el análisis nos indicará dos cuadros, el primero se presenta una descripción de los grupos comparados, la suma de rangos y un valor que representa cual grupo tiene una mediana mayor, este dato es el rango promedio, que se obtiene de dividir la suma de rangos de cada grupo entre la cantidad de casos en el grupo (Sundjaja et al., 2021).

Posteriormente saldrá otro cuadro donde indique el resultado de la U , Z y p . En cuanto al nivel de significancia de la prueba se debe verificar el valor p , si esta es menor a 0.05 se rechaza la hipótesis nula.

La U se obtiene al ir comparando cada individuo de un grupo con cada individuo del otro para contabilizar el número de veces que alguien de un grupo presenta un valor superior a alguien del otro (Martínez et al., 2020). El denominador z es el error de estándar estimado.

Ruta SPSS

1. Seleccione Analizar > cuadro de diálogos antiguos > Pruebas No paramétrica > 2 pruebas independientes
2. Pulse seleccionar variables numéricas que se desea comparar entre grupo y arrastre en la lista de variables.
3. Pulse Seleccionar variable nominal dicotómica hacia la sección de variables de grupo.
4. Posteriormente de click sobre las variables de grupos y seleccione definir grupos, posteriormente indicar grupo 1 o 2.
5. Pulse Ok para ejecutar análisis.

Ejemplo

En la siguiente tabla se compara la edad, fuerza de empuñadora, actividades básicas e instrumentales de la vida diaria, estas son variables numéricas que se busca comparar entre el grupo con sarcopenia y sin sarcopenia. La interpretación observamos de primera instancia las p , donde todas las variables son menores a 0.05, por lo tanto, se rechaza la hipótesis nula, es decir existe una diferencia significativa.

Tabla 1

Prueba U de Mann-Whitney de edad, fuerza de empuñadora, ABVD y AIVD para adultos con y sin sarcopenia

	Con sarcopenia		Sin sarcopenia		U	p
	n=64		n=80			
	Md	DE	Md	DE		
Edad	74.20	7.10	70.20	7.90	-3.158	.002
Fuerza de empuñadora	16.95	6.07	20.97	5.49	1563.50	<.001
ABVD	90.54	14.	98.62	3.72	1528.50	<.001
AIVD	6.51	1.91	7.55	.89	1771.50	<.001

Nota: =mediana; DE = desviación estándar; U = U Mann Whitney, p =significancia bilateral $n=144$, ABVD= actividades básicas de la vida diaria, AIVD= actividades instrumentales de la vida diaria (Medina, 2019).

Ejercicio

Instrucciones: interpreta los resultados de la siguiente tabla y responde las preguntas de forma correcta, en ese estudio se tuvo como objetivo comparar la calidad de vida y bienestar en adultos mayores cuidadores de nietos del norte y suroeste de México.

Tabla 2
Diferencia de medias de calidad de vida y bienestar en adultos mayores cuidadores de nietos

	Coahuila	Yucatán	<i>U</i>	<i>p</i>
	<i>M</i>	<i>M</i>		
Edad	68.60	68.50	12426.500	.0869
Bienestar subjetivo	16.44	8.62	98884.500	.001
Calidad de vida	156.56	140.97	8206.500	.000

Nota: *M*= Media, *U*= *U* de Mann-Whitne, *p*= significancia bilateral (Carrillo et al., 2022).

- 1.-¿Qué variable obtuvo una diferencia significativa?
- 2.-¿Qué grupo presento mejor calidad de vida y bienestar subjetivo?

H de Kruskal Wallis

Definición de la prueba

En el área de salud es frecuente el uso de H Kruskal Wallis para la comparación de dos o más grupos (Ventura-León, 2017). La prueba de Kruskal Wallis es la alternativa no paramétrica al ANOVA unidireccional que determina si las medianas más de dos grupos son diferentes (Glen, 2022).

Así mismo, esta prueba comprueba de una sola vez tres o más muestras independientes, podría decirse que compara las medianas de los grupos, pues usa solo la posición de cada dato y no su valor (Martinez et al., 2020), es por ello que para realizar podemos utilizar una variable tipo ordinal o nominal politómica.

Supuestos

- La variable independiente tiene una escala categórica de más de 2 categorías.
- La prueba se usa más comúnmente cuando tienes tres o más niveles.
- Para dos niveles, considere utilizar la prueba *U* de Mann Whitney

- Variables dependientes de escala ordinal o escala de razón.
- las variables deben ser independientes, en otras palabras, no debe haber relación entre los miembros de cada grupo ni entre grupos (Glen, 2022).

Hipótesis

La formulación de la hipótesis se debe comparar la mediana de la población donde se encontrará si las medianas de las poblaciones son iguales o no son iguales

Para plantear la hipótesis nula se podría establecer de la siguiente manera: Ejemplo: (H0) “No existe diferencia significativa de la ansiedad por semestre académico en la licenciatura en enfermería”

En la hipótesis alternativa (H1) se observará una diferencia significativa. Ejemplo: “La ansiedad es diferente significativamente por semestre académico en la licenciatura en enfermería”

Interpretación

Al realizar el análisis de los resultados se observará las tablas arrojadas con el SPSS, se contará con dos tablas, la primera es la de rangos, aquí se tomará en cuenta el rango promedio para observar si existencia una diferencia, posteriormente se analiza los estadísticos de prueba, donde si la p (sig asintótica) es menor .05 se concluye que existe una diferencia significativa entre los grupos independientes.

Ruta SPSS

1. Seleccione Analizar > cuadro de diálogos antiguos > Pruebas No paramétricas > K de muestras independientes
2. Pulse seleccionar variables numéricas que se desea comparar entre grupo y arrastre en la lista de variables
3. Pulse Seleccionar variable nominal politómica u ordinal hacia la sección de variables de grupo.
4. Posteriormente de click sobre las variables de grupos y seleccione definir grupos,
5. Posteriormente indicar el número mínimo y máximo de los grupos.
6. Pulse Ok para ejecutar análisis.

Ejemplos

En la tabla 3 se compara la mediana de los componentes del desempeño físico con los grupos de no tener enfermedad crónica, diabetes tipo 2 e hipertensión arterial, donde la variable de SPPB (desempeño físico) fue la única con diferen-

cia significativa, siendo mayor el desempeño físico en el grupo de los adultos mayores sin enfermedad crónica.

Tabla 3

Diferencia de medias de los componentes de desempeño físico con presencia o no de enfermedades crónicas

	Sin enfermedad crónica	Diabetes tipo 2	Hipertensión arterial	<i>P</i>
	<i>Md</i>	<i>Md</i>	<i>Md</i>	
Prueba de balance	8.36	6.84	6.28	.061
Velocidad de la marcha	1.29	1.23	1.28	.762
Prueba de levantarse de la silla	13.54	12.16	13.69	.160
SPPB	9.17	9.11	8.90	0.36

Nota: x=media, SPPB prueba de desempeño físico, p=significancia bilateral

Ejercicio

Instrucciones: interpreta los resultados de la siguiente tabla y responde de forma correcta las preguntas.

Tabla 4

Agilidad, capacidad aeróbica y fuerza de segmento superior por grupo de edad

	60 a 64 años	65 a 69 años	70 a 79 años	<i>P</i>
	<i>x</i>	<i>x</i>	<i>x</i>	
Agilidad	6.30	6.35	6.45	.610
Capacidad aeróbica	52.42	50.20	45.98	.049
Fuerza segmento superior	16.00	15.00	14.00	00.160

Nota: p= significancia bilateral

- 1.-¿Qué variable obtuvo una diferencia significativa?
- 2.-¿Qué grupo presento mejor calidad de vida y bienestar subjetivo?

Referencias

CARRILLO, A. L., Sifuentes, D., Yam, A., Cándila, J., González, J., & Esparza, S. C. (2022). Calidad de vida en adultos mayores cuidadores de nietos del

- norte y suroeste de México. *Enfermería Global*, 21(2), 46–63. <https://doi.org/10.6018/eglobal.489371>
- GLEN, S. (2022). *Kruskal Wallis H Test: Definition, Examples, Assumptions, SPSS*. StatisticsHowTo.Com: Elementary Statistics for the Rest of Us! <https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/kruskal-wallis/>
- GRISHAEV, V. Y., & Гришаев, В. Ю. (2015). *Mann-whitney criterion*. <https://www.revistaseden.org/files/12-CAP 12.pdf>
- IBM Corporation. (2021). *Prueba U de Mann-Whitney*.
- LANDEROS, R., & González, M. (2009). *Estadística con SPSS y metodología de la investigación* (Trillas (ed.); 3era edición).
- MARTINEZ, M., Sánchez, A., Toledo, E., & Faulin, J. (2020). *Bioestadística amigable* (Trillas (ed.); 4ta edición).
- MEDINA, I. (2019). *Influencia de la sarcopenia sobre los componentes físicos de calidad de vida y capacidad funcional de adultos mayores*. Universidad Autónoma de Coahuila.
- MILTON, S. (2007). *Estadística para biología y ciencias de la salud* (Mc Graw Hill (ed.); Tercera ed).
- MORAL, I. (2016). Comparación de medias. In *Métodos Estadísticos Para Enfermería Nefrológica* (pp. 165–184). <https://www.revistaseden.org/files/12-CAP 12.pdf>
- RIVAS-RUIZ, R., Moreno-Palacios, J., & Talavera, J. (2013). Investigación clínica XVI Diferencias de medianas con la U de Mann-Whitney Rodolfo. *Rev Med Inst Mex Seguro Soc*, 51(4), 414–419. <https://www.medigraphic.com/pdfs/imss/im-2013/im134k.pdf>
- SUNDJAJA, J. H., Shrestha, R., & Krishan, K. (2021). *McNemar And Mann-Whitney U Tests*. StatPearls. <http://www.ncbi.nlm.nih.gov/pubmed/32809534>
- VENTURA-LEÓN, J. L. (2017). Tamaño del efecto para Kruskal-Wallis: aportes al artículo de Domínguez-González et al. *Investigación En Educación Médica*, 8, 135–136. <https://doi.org/10.1016/j.riem.2017.07.002>
- WEAVER, K., Morales, V., Dunn, S., Godde, K., & Weaver, P. (2018). Mann-Whitney U and Wilcoxon Signed-Rank Learning. In John Wiley & Sons (Ed.), *An Introduction to Statistical Analysis in Research: With Applications in the Biological and Life Sciences* (First Edit).

Capítulo 13

Análisis de regresión lineal.

Regresión lineal simple y múltiple

DCE REYNA TORRES OBREGÓN

Definición de prueba

El análisis de regresión lineal se utiliza para predecir el valor de una variable sobre el valor de otra variable. La variable que desea predecir se llama variable dependiente. Las variables que se utilizan para predecir el valor de otras variables se denominan variables independientes.

Este análisis estadístico calcula los coeficientes de una ecuación lineal que contiene una o más variables independientes que mejor predice el valor de una variable dependiente. La regresión lineal se ajusta a una línea recta o a una superficie curva que reduce la diferencia entre los valores de salida previstos y reales. Existe una calculadora de regresión lineal simple que utiliza mínimos cuadrados para determinar la línea recta que mejor se adapta a conjuntos de datos pareados (Freund, J. & Simon, G., 2000).

Los modelos de regresión lineal proporcionan fórmulas simples y fáciles de interpretar que pueden generar predicciones. Se utilizan en algunos campos como la biología, las ciencias ambientales, las ciencias sociales y las ciencias del comportamiento. Cabe destacar que se ha convertido en un método fiable de predicción. (Montero Granados, 2016).

Supuestos (para que se usa)

Es un método estadístico útil que podemos utilizar para comprender la relación entre dos variables, X e Y . Sin embargo, antes de realizar una regresión lineal, primero debemos asegurarnos de que se cumplan cuatro supuestos (Morton, R. Hebel, R. & McCarter, R., 1993):

Supuesto 1: Relación lineal

Existe una relación lineal entre la variable independiente “ x ” y la variable dependiente “ y ”.

Cómo determinar si se cumple este supuesto:

La forma más sencilla de demostrar si esta suposición es correcta es crear un diagrama de dispersión X/Y . Esto le permite ver visualmente si existe una relación lineal entre dos variables. Si se observa que los puntos del gráfico se encuentran en una línea recta, entonces existe algún tipo de relación lineal entre las dos variables y se cumple este supuesto.

Qué hacer si se viola esta suposición: si crea un diagrama de dispersión de los valores de x y x y descubre que no existe una relación lineal entre las dos variables, todavía tiene una opción:

Se aplica la función lineal. Transformación de variables independientes y/o dependientes. Por ejemplo, puedes calcular el logaritmo, la raíz cuadrada o el recíproco de la variable independiente y/o dependiente.

Supuesto 2: Independencia

Los residuales son independientes. Es decir, no existe correlación entre residuos consecutivos en los datos. Esto es especialmente relevante cuando se trabaja con datos de series temporales. Idealmente, no desea que exista un patrón entre residuos consecutivos. Por ejemplo, los residuos no deberían seguir aumentando con el tiempo.

Para determinar si esta suposición es correcta:

La forma más natural de probar si esta suposición es correcta es ver un gráfico de series de tiempo residuales, que es un gráfico de residuos versus tiempo. Idealmente, la mayor parte de la autocorrelación residual debería estar dentro de una banda de confianza del 95% alrededor de cero. Al mismo tiempo, también podemos utilizar la prueba de Durbin-Watson para comprobar formalmente si se cumple este supuesto. Qué hacer si se viola este supuesto: Dependiendo de la naturaleza de la violación de este supuesto, tiene varias opciones:

- Para obtener una correlación serial positiva, modele los rezagos de las variables dependientes y/o independientes. Considere agregarlo/a.
- Si hay correlación serial negativa, asegúrese de que ninguna de las variables esté sobrediferenciada.
- Para correlaciones estacionales, debería considerar agregar variables ficticias estacionales a su modelo.

Supuesto 3: Homocedasticidad

Este supuesto considera que los residuos tienen varianza constante en cada nivel de x . Esto se conoce como homocedasticidad. Cuando este no es el caso, se dice que los residuos sufren de heterocedasticidad.

Cuando existe heterocedasticidad en un análisis de regresión, la varianza de las estimaciones del coeficiente de regresión aumenta, lo que dificulta confiar en los resultados del análisis. Sin embargo, el modelo de regresión no detecta esto. Como resultado, es más probable que los modelos de regresión muestren que los términos del modelo son estadísticamente significativos cuando en realidad no lo son.

Para determinar si se cumple este supuesto:

La forma más sencilla de detectar la heteroscedasticidad es utilizar una gráfica residual para crear un ajuste. Cuando ajusta una línea de regresión a un conjunto de datos, puede crear un diagrama de dispersión que muestra los valores ajustados del modelo en comparación con los residuos de esos valores ajustados.

Qué hacer si se viola este supuesto:

Hay tres formas generales de corregir la heteroscedasticidad:

1. Transformar la variable dependiente. Una modificación común es utilizar el logaritmo de la variable dependiente. Es decir, si usa el tamaño de la población (variable independiente) para predecir el número de floristas en una ciudad (variable dependiente), puede usar el tamaño de la población para predecir el logaritmo del número de floristas en una ciudad. El uso del logaritmo de la variable dependiente en lugar de la variable dependiente original a menudo elimina la heteroscedasticidad.
2. Redefinir la variable dependiente. Una forma común de redefinir la variable dependiente es utilizar proporciones en lugar de valores brutos.
3. Utilice regresión ponderada, la cual es otra forma de corregir la heterocedasticidad, este tipo de regresión asigna un peso a cada punto de datos en función de la varianza de su valor ajustado. Básicamente, esto da pequeños pesos a los puntos de datos que tienen variaciones altas, lo que disminuye sus residuos al cuadrado. Por ende, cuando se usan los pesos adecuados, se puede eliminar el problema de la heterocedasticidad.

Supuesto 4: Normalidad

Los residuos del modelo se distribuyen normalmente.

Es importante determinar si se cumple este supuesto.

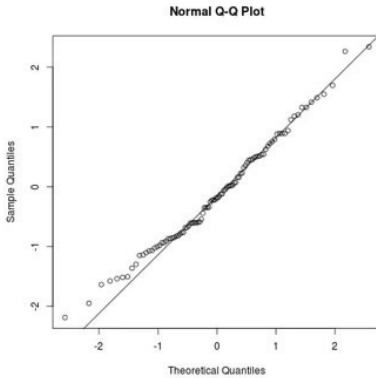
Hay que considerar que si se violan uno o más de estos supuestos, los resultados del análisis lineal pueden ser poco fiables.

Hay dos formas habituales de comprobar si se cumple este supuesto:

1. Verifique la suposición visualmente usando gráficos QQ .

Un gráfico QQ (abreviatura de gráfico cuantil-cuantil) es un gráfico que se utiliza para determinar si los residuos de un modelo siguen una distribución

normal. Los puntos del diagrama deben formar una línea diagonal aproximadamente recta y se cumple el supuesto de normalidad (Daniel, W. 2008).



Fuente https://statologos.com/lineal-regression-assumptions/#google_vignette

2. También puede verificar el supuesto de normalidad mediante pruebas estadísticas formales como Kolmogorov-Smirnov o Shapiro-Wilk, a su vez hay que tomar en cuenta que estas pruebas son sensibles a los grandes tamaños de muestra, es decir, frecuentemente determinan que los residuos no son normales debido a que el tamaño de la muestra es grande. Esta es la razón por la que a menudo es más fácil usar métodos gráficos como una gráfica QQ para verificar esta suposición.

Qué hacer si se quebranta esta suposición

Si se vulnera el supuesto de normalidad, existen varias opciones:

- Primero, verifique que los valores atípicos no tengan un gran impacto en la distribución y si los hay, verificar que sean valores reales y que no sean errores de entrada de datos.
- Aplicar una transformación no lineal a la variable independiente y/o dependiente; es decir, tomar la raíz cuadrada o el logaritmo de la variable independiente y/o dependiente (Martínez, Sánchez, Toledo & Faulin, 2020).

Interpretación

Después de tener la base de datos lista y haber pulsado **Aceptar** en el SPSS, se desglosarán los resultados de la regresión lineal.

La primera tabla que nos debe interesar es la que se identifica como: Resumen del modelo: Ejemplo de tabla de resumen del modelo en SPSS.

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación
1	,765 ^a	,585	,481	43,64452

a. Predictores: (Constante), VAR00001

ANOVA^a

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	10747,456	1	10747,456	5,642	,076 ^b
	Residuo	7619,378	4	1904,844		
	Total	18366,833	5			

a. Variable dependiente: VAR00002

b. Predictores: (Constante), VAR00001

Coeficientes^a

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	95,0% intervalo de confianza para B	
		B	Error estándar	Beta			Límite inferior	Límite superior
1	(Constante)	-7,061	38,560		-,183	,864	-114,121	99,999
	VAR00001	7,615	3,206	,765	2,375	,076	-1,286	16,516

a. Variable dependiente: VAR00002

Fuente: <https://estamatica.net/tipos-de-regresion-en-estadistica/>

A continuación, se menciona se explica cómo interpretar los resultados más importantes de las tablas:

R Cuadrado: Es la proporción de la varianza en la variable de respuesta que se puede explicar por la variable explicativa. En el ejemplo anterior, se explica el 58,5% de la variabilidad de la variable dependiente o la variable de respuesta. Cabe resaltar que el resultado se debe reportar en porcentaje.

Error estándar de la estimación: El error estándar es la distancia promedio que los valores observados caen desde la línea de regresión.

B no estandarizado (constante): Esto nos dice el valor promedio de la variable de respuesta cuando la variable predictora es cero.

B no estandarizado: Esto nos dice el cambio promedio en la variable de respuesta asociado con un aumento de una unidad en la variable predictora.

Sig: Es el valor de p asociado con la estadística de prueba. Cuando este valor es menor que 0.05, podemos concluir que la variable predictora es estadísticamente significativa, y por el contrario si el valor de p es mayor de 0.05 se puede concluir que la variable predictora NO es estadísticamente significativa (Landeros, R., & González, M., 2009).

Ruta SPSS

A continuación, se desglosan los pasos para realizar una regresión lineal.

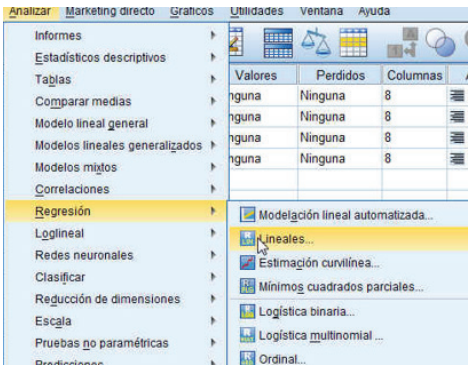
Paso 1: Visualice los datos

Realizar previo un análisis de correlación entre las variables, esto con el fin de confirmar que se cumplen los supuestos.

Suponiendo que se cumple los supuestos y se obtiene una relación lineal entre todas las variables, el siguiente paso es ajustar un modelo de regresión lineal.

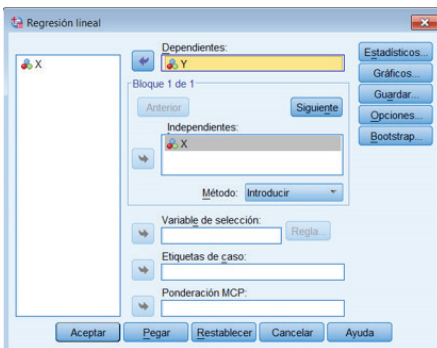
Paso 2: Ajuste un modelo de regresión lineal simple

Haga clic en la pestaña Analizar >> Regresión >> Lineal:



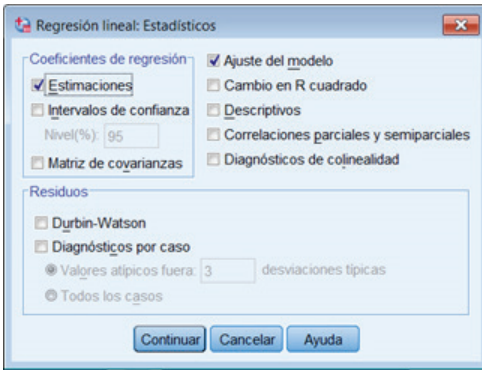
Fuente: https://wpd.ugr.es/~bioestad/wp-content/uploads/cuerpo3_5.jpg

Aparecerá una nueva ventana, arrastre la variable al cuadro llamado: Dependiente y después la variable correspondiente al cuadro dominado Independiente, finalmente hacer clic en **Aceptar**.



Fuente: https://wpd.ugr.es/~bioestad/wp-content/uploads/cuerpo3_5-2.jpg

En la ventana aparecerá el botón **Estadísticos**, se selecciona lo que se va a considerar en los análisis, después Continuar y Aceptar (Pallant, J., 2007).



Fuente: https://wpd.ugr.es/~bioestad/wp-content/uploads/cuerpo3_6.jpg

Paso 3: Interprete los resultados

Una vez que haga clic en Aceptar, aparecerán los resultados de la regresión lineal, los cuales serán en tablas como se mostraron anteriormente. A continuación, se muestra un ejemplo de interpretación de resultados:

Se ajustó un modelo de regresión, el cual consideró como variables independientes la edad, depresión, autoestima sexual, escolaridad, estado civil y religión y como variable dependiente la conducta sexual asertiva. Los resultados muestran que la edad, depresión, autoestima sexual y religión influyen en la conducta sexual asertiva de las mujeres (tabla 1). Las variables explicaron la varianza en un 46% de la conducta sexual asertiva

($F_{[6,598]} = 87.396, p < .001$). Con estos resultados se pudiera decir que a mayor edad, mayor conducta sexual asertiva; a menor depresión, mayor conducta sexual asertiva; a mayor autoestima sexual mayor conducta sexual asertiva y a menor prácticas religiosas, mayor conducta sexual asertiva (Torres, R., 2017).

Tabla 1

Modelo de regresión de edad, depresión, autoestima sexual, escolaridad, estado civil, religión con conducta sexual asertiva

Modelo	<i>B</i>	<i>SE</i>	<i>Beta</i>	<i>p</i>
Constante	7.101	5.505	-	.198
Edad	.266	.105	.077	.012
Depresión	-.198	.052	-.141	.001
Autoestima sexual	.380	.036	.368	.001
Escolaridad	.658	.371	.053	.076
Estado civil	-1.410	3.150	-.031	.654
Religión	-2.790	3.190	.456	.001

Nota: Variable dependiente = conducta sexual asertiva; *B*= b; *SE*= Error estándar; *Beta*= valor de beta; *n*=605

Ejercicio

Interpreta los resultados de la tabla 2:

Resultados del modelo: $R^2 = 0.23$; $F = 29.733$; $p < .001$

Tabla 2

Modelo de regresión de edad, depresión, autoestima sexual, escolaridad, estado civil y religión con apoyo de pareja

Apoyo de pareja	<i>B</i>	<i>SE</i>	<i>Beta</i>	<i>p</i>
Constante	52.684	3.601	-	.001
Edad	.076	.069	.040	.268
Depresión	-.155	.034	.204	.001
Autoestima sexual	.136	.023	.242	.001
Escolaridad	-.151	.243	-.022	.535
Estado civil	-.355	2.065	-.014	.864
Religión	.188	1.089	.209	.013

Nota: Variable dependiente= apoyo de pareja; *B*= b; *SE*= Error estándar; *Beta*= valor de beta; *n*=605

Referencias

- DANIEL, W. Bioestadística: bases para el análisis de ciencias de la salud. (2008). México. LIMUSA WILEY.
- FREUND, J. & Simon, G. (2000). Estadística elemental. 8a. Ed. Prentice Hall.
- LANDEROS, R., & González, M. (2009). *Estadística con SPSS y metodología de la investigación* (Trillas (ed.); 3era edición).
- MARTINEZ, M., Sánchez, A., Toledo, E., & Faulin, J. (2020). *Bioestadística amigable* (Trillas (ed.); 4ta edición).
- MONTERO Granados, R (2016): Modelos de regresión lineal múltiple. Documentos de Trabajo en Economía Aplicada. Universidad de Granada. España.
- MORTON, R. Hebel, R. & McCarter, R. (1993). Bioestadística y Epidemiología. 3a. Ed. Interamericana.
- PAGANO, M. & Gauvreau, K. (2001). Fundamentos de bioestadística. 2a. Ed. Math Learning.
- PALLANT, J. (2007). SPSS. Survival Manual. Open University Express.
- TORRES Obregón, Reyna (2017). Modelo de conducta sexual asertiva en mujeres. Doctorado thesis, Universidad Autónoma de Nuevo León

Capítulo 14

Regresión logística

DCE EDNA IDALIA PAULINA NAVARRO OLIVA

Definición de la prueba

La Regresión Logística (RL) es un grupo de pruebas estadísticas multivariantes que permite relizar predicciones, en donde se tiene como objetivo comprobar la probabilidad de que suceda o no un acontecimiento (variable dependiente - binaria) con relación a un conjunto de variables (independientes – no necesariamente binaria) (Martínez y Pérez, 2024). Ahora bien, se considera Regresión Logística Binaria o Binominal (RLB) cuando la variable dependiente tiene dos categorías y como Regresión Logística Multinomial (RLM) cuando tiene tres categorías o más (Berlanga-Silvente y Vilà-Baños, 2014; Maurandi-López, Del Río, González-Vidal y Ferre, 2019).

Cada regresión incluye opciones que no están disponibles en otras regresiones. Cada regresión es teóricamente distinta y RLB genera todas las predicciones, residuos, estadísticas de efectos y pruebas de bondad de ajuste al nivel del caso individual, independientemente de cómo se ingresan los datos o del número de patrones de covariables. Menor o mayor que el número total de casos. Luego, el RLM agrega internamente los casos para formar subpoblaciones con patrones idénticos en las covariables de las variables predictoras. Esto genera predicciones, residuos y pruebas de bondad de ajuste basadas en estas subpoblaciones (International Business Machines Corporation [IBM], 2019).

Si todas las variables predictoras son categóricas o si las variables predictoras continuas toman solo un número limitado de valores (por ejemplo, cuando hay más de un caso para cada patrón de covariables), el enfoque de subpoblación puede producir una bondad de -La prueba de ajuste y los residuales son informativos, pero el método descrito anteriormente no lo permite (IBM, 2019). Por ejemplo, en una investigación se trabaja con la variable dependiente binaria: consumo de alcohol en jóvenes universitarios, en donde se puede presentar consumo o no consumo de alcohol, partiendo de variables independientes (edad, sexo, nivel educativo, estrés percibido, etc); en este caso se propondría utilizar un modelo de RLB. Así mismo, se podría trabajar con una variable dependiente multinomial: los tipos de consumo de alcohol en jóvenes universitario (bajo riesgo, intenso, perjudicial y dependiente), partiendo de variables independientes; en este caso se propondría utilizar un modelo de RLM.

Es importante resaltar, que la diferencia al utilizar modelos de RL a otros tipos de modelos de regresión, las variables dependientes no necesariamente deben ser cuantitativas, ni tampoco cumplir con supuestos de normalidad, así como permite introducir variables categóricas como independientes en el modelo. Con relación a las variables independientes, a diferencia de en otro tipo de modelos, en la RL no es necesario cumplir con supuestos de normalidad o de igualdad de varianzas (Berlanga-Silvente y Vilà-Baños, 2014).

Por sus características, los modelos de RL permiten cumplir con tres propósitos, como cuantificar la relación que existe entre las variables independientes y la variable dependiente, así como conocer la existencia de interacción o confusión entre las primeras y la segunda, al conocer el Odds Ratio para cada variable independiente y clasificar a los sujetos en estudio dentro de categorías como: presente o ausente (Berlanga-Silvente y Vilà-Baños, 2014; Maurandi-López, et al., 2019).

Supuestos

Se menciona que el modelo de RL es una prueba estadística robusta, al no requerir de diversos supuestos a diferencia de algunos otros modelos de regresión; ya que no presenta condiciones de aplicación restrictiva; sin embargo, es de vital importancia que el modelo este bien fundamentado teóricamente (Berlanga-Silvente y Vilà-Baños, 2014). Además, para llevar a cabo la RL se requiere contar con una variable dependiente cualitativa y una o más variables independientes que pueden ser cualitativas o cuantitativas (Maurandi-López, et al., 2019)

Sin embargo, se pueden identificar cuatro supuestos, el primero es la *linealidad en el registro de las probabilidades*, ya que la RL asume que existe relación lineal entre las variables predictoras y el logaritmo de la razón de probabilidades; el segundo supuesto es la *independencia de observaciones*, donde se supone que las observaciones son independientes entre sí, es decir, que los resultados de un individuo no están influenciados por los resultados de otros individuos; el tercer supuesto es la *ausencia de multicolinealidad*, se espera que las variables predictoras no estén altamente correlacionadas entre sí, ya que esto puede dificultar la interpretación de los coeficientes y finalmente el cuarto supuesto es la *ausencia de interacciones no lineales*, se asume que no hay interacciones complejas entre las variables predictoras, es decir, que el efecto de una variable no depende del valor de otra variable (Maurandi-López, et al., 2019).

Métodos para realizar el modelo de RL

Al contar con un modelo inicial se debe proceder a la reducción con la finalidad de obtener el modelo que mejor explique los datos, para esto se puede utilizar diversos métodos que ofrece el paquete estadístico, es importante mencionar que alguna de estas opciones se necesario conocer como se contruye el modelo con base en la literatura, ya que se puede ir introduciendo las variables que otorguen mayor peso al modelo, sin embargo el mismo paquete estadístico ofrece opciones automáticas que de acuerdo a los resultados observados se van agregando variables que aportan al modelo o retirando aquellas que no lo hacen (Berlanga-Silvente y Vilà-Baños, 2014; Maurandi-López, et al., 2019).

El primer método se puede encontrar como *Introducir*, este modelo le permite al investigador tener la capacidad de decidir cuales variables se introducen o extraen del modelo, basado en lo que ha encontrado en la literatura, como se mencionó anteriormente. Así mismo se puede utilizar el método *Pasos hacia Adelante*, este es considerado como uno de los métodos automáticos que permite que el programa introduzca las variables en el modelo, empezará introducir aquellas variables que muestran coeficientes de regresión más grandes y estadísticamente significativos; es importante mencionar que en cada paso se realizará una evaluación de los coeficientes y su significancia estadística, esto permitirá que se puedan eliminar del modelo las variables que no se consideran estadísticamente significativas o que aporten al modelo (Berlanga-Silvente y Vilà-Baños, 2014; Maurandi-López, et al., 2019).

Otro de los métodos se le conoce como *Pasos hacia Atrás*, al igual que el modelo anteriormente descrito, es otro de los métodos automáticos, sin embargo en este caso se parte de un modelo en donde se tienen disponibles todas las variables independientes que se hayan seleccionado en el cuadro de diálogo y elimina del modelo aquellas variables sin significancia estadística. Es importante resaltar que en los métodos por *Pasos hacia Adelante o hacia Atrás* el programa estadístico permite elegir tres diferentes estadísticos: *razón de verosimilitud, condicional y Wald*, los cuales servirán al investigador en la toma de decisiones estadísticas con el objetivo de comprobar la significación estadística que cada coeficiente de regresión aporta al modelo (Berlanga-Silvente y Vilà-Baños, 2014; Maurandi-López, et al., 2019).

La primera opción, llamada “Selección directa (condicional)”, es un paso que ayuda a comparar la entrada de la variable con la significancia estadística y eliminarla en función de la probabilidad de la estadística del índice de verosimilitud basada en la estimación condicional basada en la comparación. Este es un método de selección. parámetros. De manera similar, la selección directa (índice de probabilidad) es un método de selección de paso adelante que contrasta la entrada de variables según la importancia de la estadística de evaluación con la eliminación basada en la probabilidad de la estadística del índice de proba-

bilidad. Esto está respaldado por estimaciones parciales de máxima verosimilitud (Berlanga-Silvente y Vilà-Baños, 2014; Maurandi-López, et al., 2019). Otro método es la selección directa (bosque). Esto compara la entrada de variables según la importancia de la estadística de puntuación y la eliminación según la probabilidad de la estadística de Wald. Finalmente, los métodos de cancelación hacia atrás (condicionales) realizan cancelación de contraste basada en la probabilidad de una estadística de razón de verosimilitud basada en estimaciones condicionales de parámetros (Berlanga-Silvente y Vilà-Baños, 2014; Maurandi-López, et al., 2019).

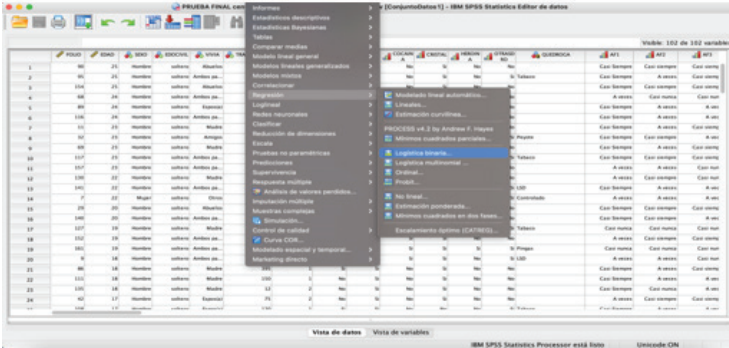
Interpretación

El propósito de realizar un análisis a través de una RL es predecir la probabilidad de que una situación ocurra o no. Por lo regular, se otorgan valores de cero y uno, en donde cero será la probabilidad de que dicha situación no ocurra y uno de que la situación ocurrirá. Así mismo, se debe analizar cuáles de las variables independientes son las que aportan mayor peso al modelo, es decir, cuales variables aumentan o disminuyen la probabilidad de que a algún sujeto le suceda o no la situación, por lo tanto para llevar a cabo el análisis nos enfocamos en las características (variables) en las que se puede presentar o no la situación o fenómeno de estudio (Maurandi-López, et al., 2019).

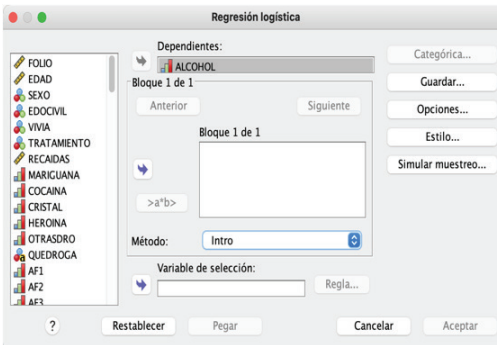
En este sentido la interpretación de la RL se basa en los coeficientes de regresión asociados a cada variable independiente, estos coeficientes representan el cambio en el logaritmo de Odds Ratio (razón de probabilidades) del evento en función de un cambio unitario en la variable independiente, manteniendo constantes las demás variables. En términos más sencillos, los coeficientes nos indican cómo influye cada variable en la probabilidad de que ocurra o no el evento de interés. Un coeficiente positivo significa que a medida que aumenta el valor de la variable independiente, la probabilidad del evento también aumentará. Por otro lado, un coeficiente negativo indica que a medida que aumenta el valor de la variable independiente, la probabilidad del evento disminuye. Es importante tener en cuenta que la interpretación de los coeficientes puede variar dependiendo de la escala y la codificación de las variables. Además, es recomendable considerar otros aspectos como el valor p y el intervalo de confianza para evaluar la significancia estadística de los coeficientes (Maurandi-López, et al., 2019).

Ruta SPSS

Para obtener un análisis de RLB se deberán seguir los siguientes pasos (IBM, 2019):

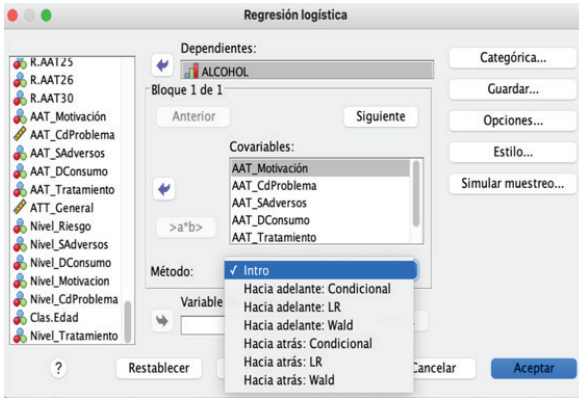
1. Elija en el menú: *Analizar >> Regresión >> Logística binaria ...*

2. Seleccione una variable dependiente dicotómica (puede ser numérica o de cadena).

3. Seleccione una o varias covariables. Para incluir términos de interacción, seleccione todas las variables implicadas en la interacción y, a continuación, seleccione $>a*b>$. Para escribir variables en grupos (Bloques), seleccione las covariables para un bloque y pulse: *Siguiente* para especificar un nuevo bloque. Repita estos pasos hasta que haya especificado todos los bloques. Si lo desea, puede seleccionar casos para el análisis, elija una *Variable de selección* e introduzca los criterios de *Regla*.

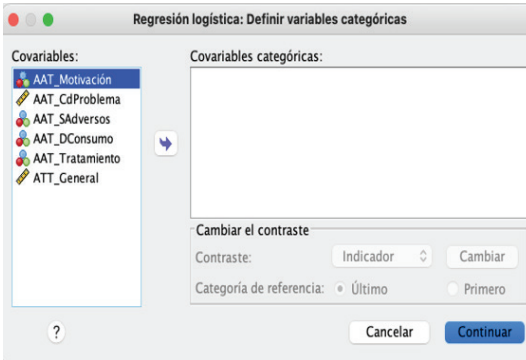
Esta opción de *Regla de selección* permite incluir especificaciones en la estimación del modelo. Por ejemplo, si se selecciona una variable y la opción *igual que* se especifica a 5 como el valor de regla, el programa sólo incluirá en el análisis aquellos casos el valor sea 5 en la variable seleccionada.

4. Seleccione el método que desea utilizar (recuerde tomar en cuenta lo abordado en el apartado Métodos para realizar el modelo de RLB).



Categorica ... definir variable categórica

En esta comando *Categorica ...* se puede especificar detalles sobre cómo el procedimiento de RLB manipulará las variables categóricas. En la opción *Covariables* contiene una lista de todas las variables independientes o covariables especificadas en el cuadro de diálogo *Covariables*, ya sea por ellas mismas o como parte de una interacción, en cualquier capa. Es importante mencionar que si alguna de las variables son de tipo cadena o categóricas, sólo se podrán utilizar como covariables categóricas (IBM, 2019).



En *Covariables categóricas*, se posee una lista de las variables identificadas como categóricas, cada variable incluye una notación entre paréntesis indicando el esquema de codificación de contraste que va a utilizarse. Las variables de cadena (señaladas con el símbolo < a continuación del nombre) estarán presentes ya en la lista *Covariables categóricas*. Se puede seleccionar cualquier otra covariable categórica de la lista *Covariables* y puede moverse a la lista *Covariables categóri-*

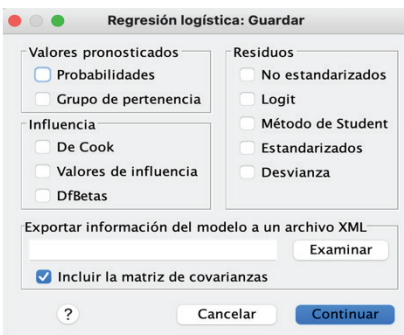
cas. La opción *Cambiar el contraste*, le permite cambiar el método de contraste. Los métodos de contraste disponibles son: *Indicador* indica la presencia o ausencia de la pertenencia a una categoría, la categoría de referencia se representa en la matriz de contrastes como una fila de ceros. La opción *Simple* permite que cada categoría del predictor (excepto la propia categoría de referencia) se compare con la categoría de referencia (IBM, 2019).

La opción de *Diferencia*, posibilita que cada categoría del predictor, excepto la primera categoría, se compare con el efecto promedio de las categorías anteriores, también se le conoce como contrastes de Helmert inversos. La opción de *Helmert*, permite que cada categoría del predictor, excepto la última categoría, se compare con el efecto promedio de las categorías subsiguientes. La opción de *Repetido*, posibilita que cada categoría de la variable de predictor (excepto la última categoría) se compare con la categoría siguiente. Así mismo en la opción *Polinómico*, permite que los contrastes polinómicos ortogonales que las categorías estén espaciadas *equidistantemente* (sólo están disponibles para variables numéricas) (IBM, 2019).

Finalmente, la opción *Desviación* le permite comparar cada categoría del predictor, excluyendo la categoría de referencia, con el efecto general. Es importante tener en cuenta que, si selecciona *Desviación*, *Simple* o *Indicador*, debe seleccionar *Primero* o *Último* como categoría de referencia. Así como las covariables de cadena deben ser covariables categóricas, para eliminar una variable de cadena de la lista de covariables categóricas, elimine todos los términos que contengan esa variable de la lista de covariables en el cuadro de diálogo *Variables* (IBM, 2019).

Guardar ... nuevas variables

En el comando *Guardar ...* se puede indicar al paquete estadístico que guarde los resultados de la RLB como variables nuevas, las cuales se incluirán en el conjunto de datos (IBM, 2019).



Los comandos disponibles para guardar incluyen Valores previstos, que tienen dos opciones: Probabilidad y Grupo de membresía. La opción Probabilidad almacena la probabilidad prevista de que ocurra el evento en cada caso. Los resultados muestran una tabla que muestra el nombre y el contenido de la nueva variable. Y los eventos son la categoría de la variable dependiente con mayor valor. Es decir, H. Si a la variable dependiente se le asignan valores de 0 y 1, la probabilidad predicha se almacena como 1. Otra opción disponible es 'Membresía de grupo'. Esta opción guarda en el grupo más probable después del análisis y se basa en el valor discriminante (IBM, 2019).

En el comando de *Influencia* es la opción que permite guardar los valores de estadísticos que miden la influencia de los casos sobre los valores pronosticados. Y se tienen tres opciones disponibles *De Cook*, *Valores de influencia* y *DfBetas*. En la opción *De Cook* permite conocer cuánto cambiarían los residuos de todos los casos, si se llegara a excluir del cálculo de los coeficientes de regresión un caso en particular. La opción de *Valor de influencia* permite conocer la influencia relativa de una observación en el ajuste del modelo y la opción de *DfBetas* es la diferencia en el valor de beta o el cambio en el valor de un coeficiente de regresión que resulta de la exclusión de un caso particular, en donde se calculará un valor para cada término del modelo, incluyendo la constante (IBM, 2019).

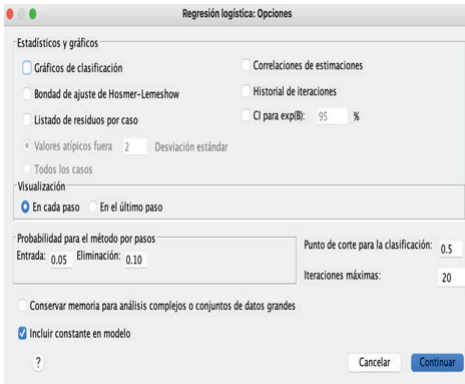
Otro de los comandos disponibles es *Residuos* el cual permite guardar los residuos y se tienen disponibles las opciones de *No estandarizados*, *Logit*, *Método de Student*, *Estandarizados* y *Desviianza*. En la opción de *No estandarizados* se podrá conocer la diferencia entre un valor observado y el valor predicho por el modelo. En *Logit* se conocerá el residuo del caso si se pronostica en la escala logit y se considera al residuo dividido por la probabilidad pronosticada multiplicada por 1, menos la probabilidad pronosticada. En el *Método de Student* es el cambio en la desviianza del modelo, si se excluye el caso. La opción *Estandarizados* se puede obtener el residuo dividido por una estimación del error estándar; a este tipo de residuo también se le conocer como residuos de Pearson o residuos tipificados, tienen una media de 0 y una desviación estándar de 1. Y en la opción de *Desviianza* otorga los residuos basados en la desviianza del modelo (IBM, 2019).

Finalmente el comando *Guardar ...* otorga la opción de *Exportar información del modelo a un archivo XML* el cual permitirá guardar las estimaciones de los parámetros y las covarianzas en un archivo en formato XML, el cual se podrá utilizar para aplicar la información del modelo a otros archivos de datos (IBM, 2019).

Opciones ...

Este comando tiene la opción de *Estadísticos y gráficos*, el cual permite solicitarle al paquete estadístico que desglose estadísticos y gráficos con las siguientes

opciones: *Gráficos de clasificación*, *Bondad de ajuste de Hosmer-Lemeshow*, *Listado de residuos por caso*, *Correlaciones de estimaciones*, *Historial de iteraciones* e *CI para exp(B)*. La opción de *Bondad de ajuste de Hosmer-Lemeshow* es un estadístico de bondad de ajuste más robusto que el estadístico de bondad de ajuste que normalmente se utiliza en la RL, se sugiere utilizar en los modelos con covariables continuas y en los estudios con muestra pequeñas; se basa en agrupar los casos en deciles de riesgo y comparar la probabilidad observada con la probabilidad esperada dentro de cada decil. En la opción de *Visualización* otorga la posibilidad de mostrar estadísticos y gráficos *En cada paso* o sólo el modelo final con la opción *En el último paso* (IBM, 2019).

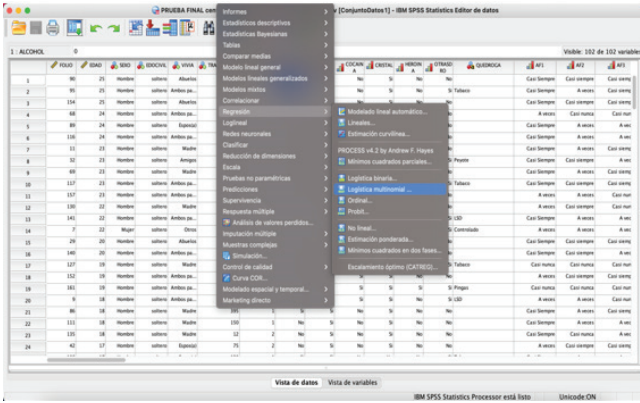


Los métodos de probabilidad escalonada le permiten controlar los criterios mediante los cuales se introducen o eliminan variables del modelo. Esta opción le permite establecer criterios para ingresar o eliminar variables. Se introduce una variable en el modelo si la probabilidad de la estadística de evaluación es menor que el valor de entrada y se elimina si la probabilidad es mayor que el valor de eliminación. Es importante mencionar que, para anular el valor predeterminado, los campos de entrada y exclusión deben ingresarse con valores positivos, y el valor de entrada debe ser menor que el valor de exclusión (IBM, 2019).

La opción de *Punto de corte para la clasificación* permite determinar el punto de corte para clasificar los casos, cuando los valores pronosticados sobrepasan el punto de corte para la clasificación se clasifican como casos positivos y los que están por debajo del punto de corte se clasifican como casos negativos, en esta opción también se puede cambiar el valor predeterminado, y se debe introducir un valor entre 0,01 y 0,99 (IBM, 2019).

Continuando con la RLM, para obtener un análisis se deberán seguir los siguientes pasos (IBM, 2019):

1. Elija en el menú: *Analizar >> Regresión >> Regresión logística multinomial...*

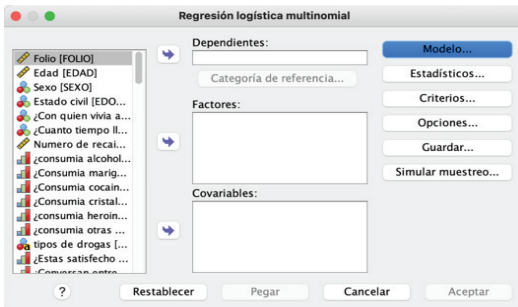


2. Seleccione una variable dependiente.
3. Los factores son opcionales y pueden ser numéricos o categóricos.
4. Las covariables son opcionales, pero si se especifican deben ser numéricas.

El paquete estadístico de manera predeterminada genera un modelo con los efectos principales del factor y la covariable, sin embargo, se puede especificar un modelo personalizado o solicitar una selección de modelos por pasos.

Modelo ...

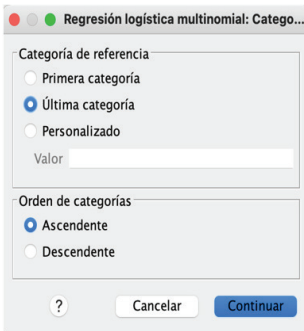
Se tiene el comando *Modelo ...* la opción *Especificar modelo*, el cual tiene las opciones de *Efectos principales*, *Factorial completo* y *Personalizado / Pasos sucesivos*. La opción de *Efectos principales* contiene los efectos principales de las covariables y los factores, pero no contiene efectos de interacción. En el caso del *Factorial completo* esta opción permite conocer todos los efectos principales y todas las interacciones factor por factor, sin embargo no contiene interacciones de covariable (IBM, 2019).



En la opción *Personalizado / Pasos sucesivos* es posible crear un modelo personalizado, en el cual se puede especificar subconjuntos de interacciones entre los factores o las interacciones entre las covariables, así como solicitar una selección por pasos de los términos del modelo. En la opción de *Factores y covariables* permiten mostrar una lista de los factores y las covariables. Así como *Términos de entrada forzada* estos términos siempre se incluirán en el modelo. Y en la opción de *Términos por pasos*, los términos añadidos a la lista por pasos se incluirán en el modelo de acuerdo con uno de los métodos por pasos seleccionados (revisar el apartado Métodos para realizar el modelo de RLB) (IBM, 2019).



Con la opción Categoría de referencia, el paquete estadístico utiliza de forma predeterminada la última categoría como categoría de referencia. El cuadro de diálogo Variables le permite controlar las categorías de referencia y el orden de las categorías. Especifique si la opción Categoría de referencia es la primera, la última o la categoría personalizada. Similar al orden de categorías, esta opción especifica el orden ascendente. Los valores más bajos definen la primera categoría, los valores más altos definen la última categoría. O orden descendente. El valor máximo define la primera categoría y el valor mínimo define la última categoría. (IBM, 2019).



Estadísticos ...

En este comando, estará de manera predeterminada ciertos estadísticos, como el *Resumen del procesamiento de casos*, en donde se podrá encontrar una tabla con información sobre las variables categóricas especificadas. En la opción de *Modelo* se encontrarán los estadísticos para el modelo global. De manera predeterminada estará seleccionado el estadístico *Pseudo R cuadrado*, el cual imprime los estadísticos *R² de Cox y Snell*, *Nagelkerke*, y *McFadden*. Así mismo podemos observar el *Resumen de pasos*, en esta tabla se resume los efectos introducidos o eliminados en cada paso en un método por pasos (es importante mencionar que no se generará, a menos que se haya especificado un modelo por pasos en el diálogo Modelo). En *Información de ajuste de los modelos*, en esta tabla se comparan los modelos ajustados y nulos o de solo intersección (IBM, 2019).

Las siguientes opciones no estarán preseleccionadas por el paquete estadístico, pero se podrán seleccionar de manera manual si el investigador lo requiere. En la opción de *Criterios de información*, se mostrará una tabla que imprime el criterio de información de *Akaike (AIC)* y el criterio de información *Bayesiano (BIC) de Schwarz*. Así mismo en la opción de *Probabilidades de casillas*, se imprime una tabla de las frecuencias observadas y esperadas (con los residuos) y las proporciones por patrón en las covariables y por categoría de respuesta. En la *Tabla de clasificación* se imprime una tabla de las respuestas observadas respecto a las respuestas pronosticadas. En la *Bondad de ajuste* se imprime los estadísticos de *chi-cuadrado de Pearson* y de *chi-cuadrado de la razón de verosimilitud*. Los estadísticos se calculan para los patrones en las covariables determinados por todos los factores y las covariables o también el investigador puede definir

un subconjunto. En las *Medidas de monotonicidad*, se muestra una tabla con información sobre el número de pares concordantes, pares discordantes y empates; así mismo se muestran los estadísticos: *D de Somers*, la *gamma de Goodman y Kruskal*, la *tau-a de Kendall* y el *índice de concordancia C* (IBM, 2019).

En el comando de *Parámetros* se podrá encontrar los estadísticos relacionados con los parámetros de modelo. Dentro de las opciones disponibles se encuentra preseleccionada la de *Estimaciones*, el cual imprime las estimaciones de los parámetros del modelo con un nivel de confianza especificado por el usuario. En la opción de *Prueba de razón de verosimilitud*, se imprimen los contrastes de la razón de verosimilitud para los efectos parciales del modelo y el contraste para el modelo global se imprime de manera automática (IBM, 2019).

Así mismo, se tienen disponibles otras opciones que no están preseleccionadas, pero que si el investigador lo requiere las puede seleccionar, como las *Correlaciones asintóticas*, donde se imprime la matriz de las correlaciones entre las estimaciones de los parámetros. Las *Covarianzas asintóticas*, imprime la matriz de las covarianzas de las estimaciones de los parámetros. Otra de las opciones es la de *Definir subpoblaciones* que le permite seleccionar un subconjunto de factores y covariables de manera que pueda definir los patrones en las covariables utilizados por las probabilidades de casilla y las pruebas de bondad de ajuste (IBM, 2019).

Crterios

En este comando se permite especificar los criterios como *Iteraciones*, esta opción permite especificar el número máximo de veces que se desea recorrer el algoritmo, así como el número de *máxima subdivisión por pasos*, las tolerancias de *convergencia para los cambios* en el *logaritmo de la verosimilitud* y los *parámetros*, la *frecuencia con la que se imprime el progreso del algoritmo iterativo* y en qué iteración el procedimiento debería empezar a *comprobar la separación completo o cuasi completa de los datos* (IBM, 2019).

Regresión logística multinomial: Criterios de convergencia

Iteraciones

Iteraciones máximas: 100

Máxima subdivisión por pasos: 5

Convergencia de log-verosimilitud: 0

Convergencia de los parámetros: 0.000001

Imprimir el historial de iteraciones para cada 1 pasos

Comprobar separación de los puntos de los datos desde la iteración 20 hacia adelante

Delta: 0 Tolerancia para la singularidad: 0.00000001

? Cancelar Continuar

En la *Convergencia del logaritmo de la verosimilitud*, se asume la convergencia si el cambio absoluto en la función *log-verosimilitud* es menor que el valor especificado, este criterio no se aplica si el valor es igual a 0 y es importante que se especifique un valor no negativo. En la *Convergencia de los parámetros*, se asume la convergencia si el cambio absoluto en las estimaciones de parámetro es menor que este valor y este criterio no se aplica si el valor es igual a 0. La opción *Delta*, le permite especificar un valor no negativo inferior a 1, este valor se añade a cada casilla vacía de la tabulación cruzada de las categorías de respuesta por patrones de covariables, con lo cual se ayuda a estabilizar el algoritmo y evitar sesgos en las estimaciones. La *Tolerancia para la singularidad* permite especificar la tolerancia empleada en la comprobación de la singularidad (IBM, 2019).

Opciones ...

En este comando se puede especificar las opciones de *Escala de dispersión* el cual permitirá especificar el valor de escalamiento de dispersión que se utilizará para corregir la estimación de la matriz de covarianzas de parámetros se tienen cuatro opciones: ninguna, la *Desviación*, que estima el valor de escalamiento mediante el estadístico de la función de desviación (chi-cuadrado de la razón de verosimilitud. La opción de *Pearson* que estima el valor de escalamiento mediante el estadístico chi-cuadrado de Pearson. Y también se puede especificar el propio valor de escalamiento, el cual debe ser un valor numérico positivo (IBM, 2019).

En *Opciones por pasos* se ofrece el control de los criterios estadísticos cuando se utilizan métodos por pasos para generar un modelo (se utilizan si se especificó un modelo por pasos). En la opción de *Probabilidad de entrada*, se trata de la probabilidad del estadístico de la razón de verosimilitud para la entrada de variables, cuanto mayor sea la probabilidad especificada la variable entra en el modelo con mayor facilidad (este criterio se utiliza si se seleccionó un modelo

de pasos hacia delante, pasos hacia adelante o pasos hacia atrás). La *Prueba de entrada*, es el método para entrar términos en métodos por pasos y se escoge entre la prueba de la razón de verosimilitud y la prueba de puntuación (este criterio se utiliza si se seleccionó los métodos de pasos hacia delante, pasos hacia adelante o pasos hacia atrás) (IBM, 2019).

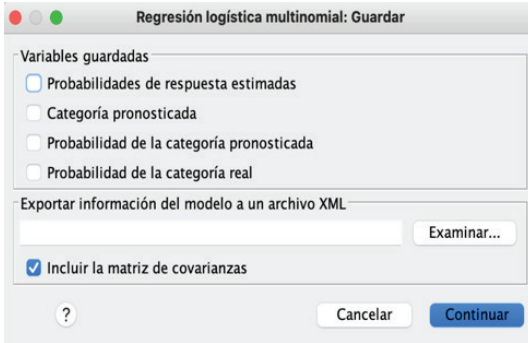
La *Probabilidad de eliminación*, es la probabilidad del estadístico de la razón de verosimilitud para la eliminación de variables, entre mayor sea la probabilidad especificada, más fácil resultará que una variable permanezca en el modelo (este criterio se utiliza si se seleccionó los métodos de eliminación hacia atrás, pasos hacia adelante o pasos hacia atrás). La *Prueba de eliminación*, es el método para eliminar términos en métodos por pasos, se puede elegir entre la prueba de la razón de verosimilitud o la prueba de Wald (este criterio se utiliza si se seleccionó los métodos de eliminación hacia atrás, pasos hacia adelante o pasos hacia atrás). *Efectos por pasos mínimos en el modelo (para métodos hacia atrás)*, al utilizar los métodos de eliminación hacia atrás o pasos sucesivos hacia atrás, se especifica el mínimo número de términos que puede incluirse en el modelo (IBM, 2019).

Los *Efectos por pasos máximos en el modelo (para métodos hacia adelante)*, al utilizar los métodos de entrada hacia delante o pasos sucesivos hacia adelante, se especifica el máximo número de términos que puede incluirse en el modelo. La opción de *Restringir jerárquicamente la entrada y la eliminación de términos*, esta opción permite elegir si se desea aplicar restricciones a la inclusión de términos de modelo, la jerarquía precisa que para que se incluya un término, todos los inferiores que formen parte del que se desea incluir se encuentren antes en el modelo (IBM, 2019).

Guardar ...

El comando *Guardar ...* permite guardar variables en el archivo de trabajo y exportar información de modelo a un archivo externo. Se tiene la opción de *Variables guardadas*, que se le solicita al programa estadístico guardar las variables de *Probabilidades de respuesta estimadas*, estas son las probabilidades estimadas de la clasificación de un patrón de factor/covariable en las categorías de respuesta; hay tantas probabilidades estimadas como categorías de la variable de respuesta y sólo se guardarán hasta 25. En la opción de *Categoría pronosticada*, es la categoría de respuesta con la mayor probabilidad esperada para un patrón de factores/covariables. La *Probabilidades de categoría predichas*, que otorga el máximo de las probabilidades de respuesta estimadas. Así como la *Probabilidad de la categoría real*, que es la probabilidad estimada de la clasificación de un patrón de factores/covariables en la categoría observada. Y finalmente la opción de *Exportar información del modelo a un archivo XML*, permitirá guardar las estimaciones de los parámetros y las covarianzas, y se exportan en un archivo en

formato XML y se podrá utilizar este archivo de modelo para aplicar la información del modelo a otros archivos de datos (IBM, 2019).



Referencias

- BERLANGA – Silvente, V. Y Vilà – Baños, R. (2014). Cómo obtener un Modelo de Regresión Logística Binaria con SPSS. *Revista d'innovació i Recerca en Eduació*, 7(2), 105-118. DOI:10.1344/reire2014.7.2727
- INTERNATIONAL Business Machines Corporation. (2019). IBM SPSS – Regresión 26. Disponible: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.ibm.com/docs/en/SSLVMB_26.0.0/pdf/es/IBM_SPSS_Regression.pdf
- MARTÍNEZ, P. J. A. y Pérez, M. P. S. (2024). Logistic regression. *Medicina de Familia. SEMERGEN*, 50(1). <https://doi.org/10.1016/j.semerg.2023.102086>
- MAURANDI – López, A., Deñ Río, A. L., González – Vidal, A. Ferre, J. M. E. Hernández, V. A. (2019). *Fundamentos Estadísticos para Investigación. Introducción a R y Modelos*. DOI: 10.5281/zenodo.2628915

Capítulo 15.

Pruebas epidemiológicas

DCE PATRICIA ENEDINA MIRANDA FÉLIX

DCE ROSARIO EDITH ORTIZ FÉLIX

Pruebas Epidemiológicas

La epidemiología es una disciplina científica que estudia la frecuencia y distribución de los fenómenos relacionados a la salud-enfermedad y sus determinantes de riesgo en poblaciones vulnerables, parte de un enfoque cuantitativo que puede ser aplicado a una amplia variedad de problemas, desde la prevención, diagnóstico, tratamiento y rehabilitación del paciente (Last J. M, 2001). En las últimas décadas la epidemiología ha logrado un gran desarrollo, gracias a la implementación de herramientas informáticas para el tratamiento de los datos, asimismo ha hecho uso de nuevas técnicas y procedimientos innovadoras que han permitido garantizar el nivel científico necesario para el trabajo del profesional sanitario en sus decisiones y prácticas diagnósticas (De Irala J, Martínez González M.A. Seguí-Gómez M., 2008).

Un aspecto importante de la epidemiología es la medición de frecuencias de enfermedades en el campo de la salud, estas, analizan la ocurrencia de un fenómeno de salud, describen el evento de una enfermedad o salud en relación con el tamaño de una población en riesgo la cual puede describirse desde una perspectiva transversal (en un solo momento del tiempo) o longitudinal (seguimiento). El profesional de enfermería debe contar con conocimientos sobre las pruebas epidemiológicas existentes para el cálculo de la frecuencia de fenómenos de salud y de magnitud de asociación entre variables que se utilizan en los diferentes estudios epidemiológicos.

Medidas de frecuencia

Existen dos tipos de medidas de frecuencia, que son las más utilizadas, la prevalencia y la incidencia.

Prevalencia

La prevalencia (P) establece la proporción de un grupo de individuos que tienen una enfermedad en un momento determinado del tiempo (t). La prevalen-

cia se determina mediante el muestreo representativo de una población definida en un momento determinado que contiene individuos con y sin el problema bajo estudio (Miron & Sardon, 2008). Representa la medida de frecuencia de la enfermedad que se calcula en los estudios transversales o de prevalencia. Existen 2 tipos de prevalencia:

Prevalencia de punto o puntual; Se considera la más común, determina en el momento del muestreo para cada individuo, aunque no necesariamente en el mismo momento para todos los individuos de la población definida. Un ejemplo podría ser determinar el número de individuos hospitalizados por COVID-19; donde el número de pacientes hospitalizados por COVID-19 es el numerador y el total de pacientes hospitalizados en dicho hospital sería el denominador.

$$PP= \frac{\text{Número de pacientes hospitalizados por COVID-19}}{\text{Total de pacientes hospitalizados en dicho hospital}}$$

Prevalencia de periodo (Pp). Es la frecuencia de una enfermedad que se presenta en un tiempo determinado, muestra la probabilidad en que un individuo la desarrolle. Un ejemplo, serían los individuos que presentaron COVID-19 antes y durante el periodo de estudio tomando el rol de numerador y el denominador aquellos individuos que enfermaron durante todo el periodo de estudio.

$$Pp= \frac{\text{Individuos que presentaron COVID-19 antes y durante el periodo de estudio}}{\text{Individuos que presentaron COVID-19 durante todo el periodo de estudio}}$$

En el presente cálculo se puede presentar que el total de la población cambie durante el período. Por lo general, se toma como punto medio del periodo a la población situada como denominador. Es importante considerar que cuando se presenta la prevalencia sin referencia en el tiempo de estudio pudiera confundirse con el indicador de prevalencia puntual. Otros criterios a considerar es conocer la relación entre prevalencia, incidencia y el tiempo que dura la enfermedad, dónde la prevalencia pudiera aumentar, disminuir o será estable según el comportamiento de la incidencia y la duración promedio de la enfermedad.

Dónde:

P = Prevalencia de la enfermedad

I = Incidencia

T = Duración promedio de la enfermedad

Hipótesis

H_0 . La prevalencia de consumo de cigarro electrónico en adolescentes es igual a la reportada a la literatura.

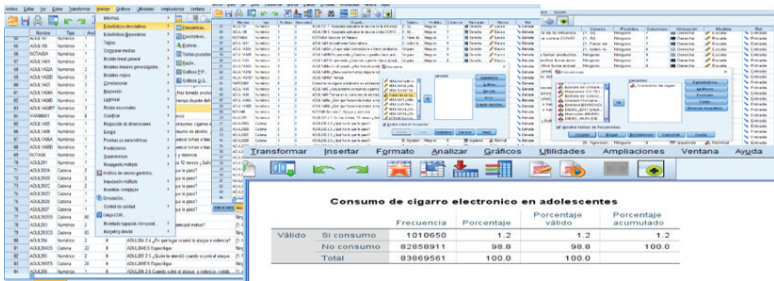
H_1 . La prevalencia de consumo de cigarro electrónico en adolescentes es mayor o menor a la reportada en la literatura.

Ruta SPSS para calcular prevalencia

Analizar >> Estadísticos descriptivos >> Frecuencias...

Esta opción brinda el menú que permite seleccionar la variable de interés, una vez seleccionada dar en aceptar, por ejemplo: Consumo de cigarro electrónico en adolescentes, de febrero a julio del 2023 (Figura 1).

Figura 1
Estadística descriptiva en SPSS



Incidencia

Es importante resaltar que para estudios epidemiológicos el propósito es conocer cómo se distribuye la enfermedad en los individuos de una población. La incidencia se define como el número de casos nuevos que se presentan en una población en riesgo durante un periodo de tiempo (Tapia Granados, 1994).

Como *supuesto*, lo primero a considerar para la distribución de la enfermedad es:

- Incidencia según la persona,
- Incidencia según lugar de residencia
- Incidencia según el tiempo.

Ejemplo:

Incidencia de adultos jóvenes de 18 a 59 años de edad con COVID-19 derechohabientes de un Hospital Público de segundo nivel de atención de los estados de la Zona Pacífico (Sinaloa, Nayarit, Jalisco y Colima) de la República Mexicana durante el periodo 2020-2022. Como podemos observar el riesgo de la población (incidencia) de adultos jóvenes con COVID-19, habitantes de determinado lugar (estados de la Zona Pacífico) y durante el periodo de 3 años (2020-2022).

Tomando como referencia el ejemplo anterior, si en el periodo del tiempo de estudio (2020-2022) se presentó un aumento de casos de acuerdo a la tendencia, no se puede afirmar que el riesgo para desarrollar la enfermedad va en aumento en la población de estudio, para este caso debiera calcularse la tasa de incidencia. En relación al aumento de número de casos solo indica que el denominador pudo aumentar y por eso incrementaron los casos, pero no que el riesgo de desarrollar la enfermedad se ha incrementado. Asimismo, al mostrar la tendencia de las tasas en determinado periodo tiempo (2020 a 2022) y se observa que con el tiempo la incidencia aumenta habrá de interpretarse que el riesgo para el desarrollo de la enfermedad (COVID-19) va en aumento o disminuye, por ello, se debe considerar las causas de esa tendencia.

Existen dos tipos de medidas de incidencia: la *incidencia acumulada* (IA) y la *tasa de incidencia o densidad de incidencia* (DI).

Incidencia acumulada

La IA se calcula dividiendo el número de casos nuevos de una enfermedad que surge en un periodo de tiempo por el total de la población en riesgo al principio del período.

Dónde:

$$IA = \frac{\text{Número de casos nuevos en un periodo determinado}}{\text{Total de la población en riesgo}}$$

Para obtener la IA se requiere de un grupo de individuos sin la enfermedad que se requiere estudiar, que después de un tiempo determinado pasan del estado de salud a la enfermedad. En esta medida, el numerador lo constituyen los individuos que enfermaron (A) y el denominador, los que no lo hicieron (A + B), en este tipo de incidencia el tiempo (t) es fijo para todos los individuos, es decir, no se toma en cuenta el momento en que enfermaron, todos tienen el mismo tiempo de seguimiento; por ello, el tiempo se elimina de la ecuación y solo queda A/AB. Para el análisis, es como una proporción, la cual puede multiplicarse por 100.

Por ejemplo: si se ha realizado el seguimiento a 1000 individuos sanos durante un año y se enfermaron 200, entonces se tendría que la IA en un año fue de 200 % (IA = 200/1000; IA = 0.2 × 1000; IA = 200%).

El tiempo de no exposición en los individuos que enferman es cambiante en esta tasa de incidencia y depende del momento en que enferman durante el periodo de estudio, situación que debería descontarse del tiempo total de seguimiento, sin embargo, no se toma en cuenta para el cálculo final.

Tasa de incidencia o densidad de incidencia. En esta medida se reflejan los casos nuevos relativos al tamaño de la población. Se calcula dividiendo el número de casos nuevos ocurridos durante un periodo de seguimiento entre la suma de todos los periodos de observación de los sujetos bajo estudio.

Dónde:

$$TI = \frac{\text{Número de casos nuevos en un periodo determinado}}{\text{Personas- Tiempo de observación}}$$

Su interpretación es colectiva y representa el potencial de cambio a la que una población sana va enfermando.

Hipótesis

H_0 . El número de casos nuevos de COVID-19 en adolescentes en los últimos 3 meses es igual a lo reportado en la literatura.

H_1 . El número de casos nuevos de COVID-19 en adolescentes en los últimos 3 meses es mayor o menor a lo reportado en la literatura.

Ruta SPSS para calcular incidencia

Analizar >> Estadísticos descriptivos >> Frecuencias...

Esta opción brinda el menú que permite seleccionar la variable de interés, una vez seleccionada se da click en aceptar, por ejemplo: Diagnostico de COVID-19 en adolescentes en los últimos 3 meses (Figura 2).

Figura 2 Incidencia en SPSS

Número	Nombre	Tipo	Alcance	Formato	Etiquetas	Valores	Perdidos	Columnas	Asociación	Mostrar	Alto
47	er	Numérico	5	3	SELECCIONANDO PARA SANGRE VENOSA	Ninguno	Ninguno	3	Derecha	Escalada	✓ Escalada
48	lva	Numérico	5	3	SELECCIONANDO PARA SANGRE COVID 19	Ninguno	Ninguno	3	Derecha	Escalada	✓ Escalada
49	notaf	Numérico	5	3	Comentario para realizar la entrevista	{ } (A) Nota	Ninguno	3	Derecha	Escalada	✓ Escalada
51	notaf1	Numérico	5	3	TOMA DE MUESTRA	{ } (F) Fecha de	Ninguno	3	Derecha	Escalada	✓ Escalada
52	cap01	Numérico	5	3	SANGRE REGISTRAR EL RESULTADO DE LA TOMA DE MUESTRA DE SANGRE VENOSA	{ } (S) Sexo	Ninguno	12	Derecha	Escalada	✓ Escalada
53	cap02	Cadena	8	3	SANGRE REGISTAR LA PERSONA EN LA QUE SE TOMA LA MUESTRA UTILIZANDO EL FORMATO	{ } (S) Sexo	Ninguno	12	Derecha	Escalada	✓ Escalada
54	cap03	Cadena	8	3	SANGRE REGISTAR LA PERSONA EN LA QUE SE TOMA LA MUESTRA UTILIZANDO EL FORMATO	{ } (S) Sexo	Ninguno	5	Izquierda	Horizontal	✓ Escalada
55	com01	Cadena	76	3	(¿Algun comentario de la entrevista?)	{ } (S) Sexo	Ninguno	25	Izquierda	Horizontal	✓ Escalada
56	tempo	Numérico	10	3	Fecha de entrevista	{ } (F) Fecha de	Ninguno	3	Derecha	Escalada	✓ Escalada
57	lva_1a	Cadena	8	3	Fecha de toma de muestra	{ } (F) Fecha de	Ninguno	10	Izquierda	Horizontal	✓ Escalada
58	lva_1b	Cadena	10	3	Fecha de toma de muestra	{ } (F) Fecha de	Ninguno	10	Izquierda	Horizontal	✓ Escalada
59	com02	Numérico	5	3	Fecha de entrevista	{ } (F) Fecha de	Ninguno	3	Derecha	Escalada	✓ Escalada
60	com03	Cadena	1	3	Fecha de entrevista	{ } (F) Fecha de	Ninguno	3	Derecha	Escalada	✓ Escalada
61	Sexo	Numérico	11	3	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
62	Etnia	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
63	edad_1	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
64	edad_2	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
65	edad_3	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
66	edad_4	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
67	edad_5	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
68	edad_6	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
69	edad_7	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
70	edad_8	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
71	edad_9	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
72											
73											
74											
75											
76											
77											

Número	Nombre	Tipo	Alcance	Formato	Etiquetas	Valores	Perdidos	Columnas	Asociación	Mostrar	Alto
47	er	Numérico	5	3	SELECCIONANDO PARA SANGRE VENOSA	Ninguno	Ninguno	3	Derecha	Escalada	✓ Escalada
48	lva	Numérico	5	3	SELECCIONANDO PARA SANGRE COVID 19	Ninguno	Ninguno	3	Derecha	Escalada	✓ Escalada
49	notaf	Numérico	5	3	Comentario para realizar la entrevista	{ } (A) Nota	Ninguno	3	Derecha	Escalada	✓ Escalada
51	notaf1	Numérico	5	3	TOMA DE MUESTRA	{ } (F) Fecha de	Ninguno	3	Derecha	Escalada	✓ Escalada
52	cap01	Numérico	5	3	SANGRE REGISTRAR EL RESULTADO DE LA TOMA DE MUESTRA DE SANGRE VENOSA	{ } (S) Sexo	Ninguno	12	Derecha	Escalada	✓ Escalada
53	cap02	Cadena	8	3	SANGRE REGISTAR LA PERSONA EN LA QUE SE TOMA LA MUESTRA UTILIZANDO EL FORMATO	{ } (S) Sexo	Ninguno	3	Izquierda	Horizontal	✓ Escalada
54	cap03	Cadena	76	3	(¿Algun comentario de la entrevista?)	{ } (S) Sexo	Ninguno	25	Izquierda	Horizontal	✓ Escalada
56	tempo	Numérico	10	3	Fecha de entrevista	{ } (F) Fecha de	Ninguno	3	Derecha	Escalada	✓ Escalada
57	lva_1a	Cadena	8	3	Fecha de toma de muestra	{ } (F) Fecha de	Ninguno	10	Izquierda	Horizontal	✓ Escalada
58	lva_1b	Cadena	10	3	Fecha de toma de muestra	{ } (F) Fecha de	Ninguno	10	Izquierda	Horizontal	✓ Escalada
59	com02	Numérico	5	3	Fecha de entrevista	{ } (F) Fecha de	Ninguno	3	Derecha	Escalada	✓ Escalada
60	com03	Cadena	1	3	Fecha de entrevista	{ } (F) Fecha de	Ninguno	3	Derecha	Escalada	✓ Escalada
61	Sexo	Numérico	11	3	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
62	Etnia	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
63	edad_1	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
64	edad_2	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
65	edad_3	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
66	edad_4	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
67	edad_5	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
68	edad_6	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
69	edad_7	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
70	edad_8	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
71	edad_9	Numérico	8	2	Registro de entrevista	{ } (S) Sexo	Ninguno	3	Derecha	Escalada	✓ Escalada
72											
73											
74											
75											
76											
77											

Estadísticos

DIAGNOSTICO DE COVID-19 EN LOS ULTIMOS 3 MESES

N	Válido	9464
	Perdidos	0

DIAGNOSTICO DE COVID-19 EN LOS ULTIMOS 3 MESES

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido POSITIVO	2152	22.7	22.7	22.7
NEGATIVO	7312	77.3	77.3	100.0
Total	9464	100.0	100.0	

Medidas de asociación o efecto en los estudios epidemiológicos

Las medidas de asociación son ampliamente utilizadas en los estudios epidemiológicos, actúan como estimador para medir la causa-efecto en un evento o enfermedad. En la gran mayoría de las investigaciones se hace uso de estos para trasladar el conocimiento científico a la práctica cotidiana. Estos estimadores pretenden medir el grado de asociación entre uno o más factores independientes (exposición) y un evento adverso o enfermedad (variable dependiente). Las medidas de asociación se relacionan con la identificación del riesgo, para determinar si ciertos factores aumentan o disminuyen la probabilidad de desarrollar un evento adverso o enfermedad (riesgo relativo y razón de momios; Fuentes, F.M.E. & Del Prado, G.N, 2013; Rendón-Macias, M.E., García, H., & Villalís-Keever, M.Á. (2021).

Para el cálculo de las medidas de asociación los datos suelen presentarse en tablas de contingencia (2 x 2), ver tabla 1. Las medidas de asociación varían en función del tipo de estudio epidemiológico y se pueden evaluar mediante contraste de hipótesis.

Tabla 1
Tabla de contingencia 2x2

	Evento o Enfermedad		TOTAL
	NO	SI	
Expuesto	A	b	a+b
No expuesto	C	d	c+d
TOTAL	a+c	b+d	a+b+c+d

Nota: a= personas expuestas al riesgo que no presentan el evento o enfermedad, b= personas expuestas al riesgo que presentan el evento o enfermedad, c= personas no expuestas al riesgo que no presentan evento o enfermedad, d= personas no expuestas al riesgo que no presentan evento o enfermedad, a+c= sumatoria de personas expuestas y no expuestas que no presentan el evento o enfermedad, b+d= sumatoria de personas expuestas y no expuestas que presentan el evento o enfermedad, a+b= sumatoria de personas expuestas que presentan y no el evento o enfermedad, c+d= sumatoria de personas no expuestas que presentan y no el evento o enfermedad, a+b+c+d= total de personas expuestas, no expuestas, que presentan el evento o no.

Riesgo relativo

Es una medida de efecto ya que representa la relación entre dos variables. Se usa generalmente para expresar la probabilidad de que una persona expuesta a un factor de riesgo presente un evento o enfermedad. Este estimador, es la medida de elección en los estudios observacionales de cohorte y estudios experimentales (Kappes & Riquelme, 2021).

Para realizar el cálculo de riesgo relativo (RR), se divide la incidencia de la enfermedad en expuesto (IE) entre la incidencia de enfermedad en no expuestos (IE°) con la siguiente fórmula:

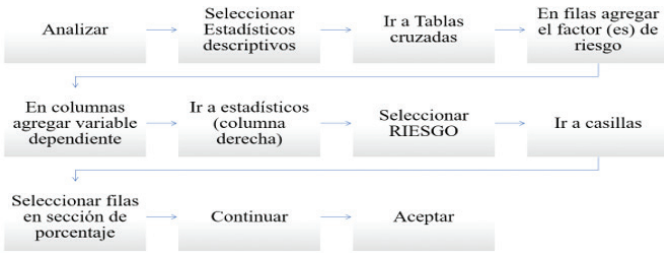
$$RR = IE / IE^{\circ} = a/a+b / c/c+d$$

El RR expresa el número de veces que es más probable que un evento o enfermedad se desarrolle en el grupo de expuestos en relación con el grupo de no expuestos. Se caracteriza por tomar solo valores positivos. Se argumenta que, para su interpretación, un valor mayor a 1 ($RR > 1$) la exposición es un factor de riesgo para el evento o enfermedad, por el contrario, si el valor es menor a 1 ($RR < 1$) la exposición es un factor protector del evento o enfermedad. Si el RR toma un valor de 1, significa que no existe asociación entre la exposición y el evento o enfermedad, ya que la incidencia de expuestos es la misma que los no expuestos.

La hipótesis nula plantea que no hay diferencias entre los grupos, lo que se cumple si el RR es igual a 1. Además, cuando el intervalo de confianza incluye a 1, la diferencia de riesgo no es estadísticamente significativa (Dagnino, 2014). Se fija un nivel de significancia por lo general 0.05, se cumple así la hipótesis nula (no hay relación entre la exposición y evento o enfermedad) bajo el supuesto de que si la p resulta menor que la significancia establecida entonces se puede afirmar que hay asociación entre la exposición y el evento o enfermedad, por tanto, se puede buscar la fuerza de esa relación (Tamargo, B., Gutiérrez, R., Quesada, P., López, N & Hidalgo, C. 2019).

El cálculo de este estimador se puede realizar a través de diversos programas estadísticos, sin embargo, para fines prácticos del presente capítulo a continuación se desarrolla la ruta a seguir para calcular el RR en el SPSS (figura 3).

Figura 3
Ruta que seguir para calcular el RR en SPSS



Para finalizar este apartado, se agrega un ejemplo del cálculo del RR aplicado a un estudio experimental (Ortiz-Félix, Cárdenas-Villarreal, Miranda-Félix & Guevara-Valtier, 2021) donde las participantes fueron mujeres gestantes sometidas a una intervención educativa prenatal para prevenir el sobrepeso en lactantes.

H_0 = Las mujeres embarazadas que recibieron la intervención educativa, no presentaron diferencias en las prácticas de alimentación en comparación con las mujeres embarazadas del grupo control.

Resultados de la aplicación de la ruta a seguir para calcular el RR en el SPSS

Tabla cruzada

		Clasificación				Total	
		Adecuada		Inadecuada			
		N	%	N	%	N	%
Grupos de la intervención	grupo control	3	12,0%	27	77,1%	30	50,0%
	grupo intervencion	22	88,0%	8	22,9%	30	50,0%
Total		25	100,0%	35	100,0%	60	100,0%

Estimación de riesgo

	Valor	Intervalo de confianza de 95 %	
		Inferior	Superior
Razón de ventajas para Grupos de la intervención (grupo control / grupo intervencion)	,040	,010	,171
Para cohorte Clasificación = Adecuada	,136	,046	,408
Para cohorte Clasificación = Inadecuada	3,375	1,842	6,182
N de casos válidos	60		

El RR es igual a 3.37 y el IC 95% va de 1.84 a 6.18 por tanto, se rechaza la hipótesis nula.

Interpretación: Las mujeres del grupo control tuvieron 3.37 mayor RIESGO de continuar con prácticas de alimentación inadecuadas para con sus hijos en comparación con las mujeres que tomaron la intervención educativa.

Odds Ratio

La odds ratio (OR) también conocida como razón de momios, razón de probabilidades o productos cruzados resulta de la división de una odds de casos de exposición entre otra odds en los expuestos del grupo control. La odds oscila entre 0 e infinito y se interpreta en una escala multiplicativa como el número de veces que es mayor la odds de exposición en los casos que en los controles. Estos estimadores se representan a través de una tabla de contingencia y son de uso frecuente para estimar la relación entre dos variables binarias, para investigar el efecto de otras variables sobre esa relación a través de la regresión logística y para analizar el resultado de estudios de casos y controles donde no es factible usar el riesgo relativo (Dagnino, 2014; Kappes & Riquelme, 2021).

Para realizar el cálculo de OR, es importante considerar la relación que existe en la columna de frecuencia de la exposición dentro de los casos y controles, por tanto, se puede calcular el *odds* de exposición dentro de los casos (a/c) y el odds de exposición en los controles (b/d), así se obtiene dos *odds* que se pueden relacionar mediante una razón y es a lo que se le denomina razón de *odds* u *odds ratio*. La OR, se representa a través de la siguiente formula:

$$OR = a \times d / b \times c$$

Para su interpretación se considera, que si el OR es mayor que 1 y los valores del IC 95 % rebasan el 1 significa que la asociación es positiva, entonces estaríamos ante un factor de riesgo. Si el OR es menor que 1, y los valores del IC de 95 % están por debajo de 1 la asociación es negativa y estaríamos bajo un factor protector para presentar un evento o enfermedad. Si el límite inferior del IC de 95 % para el OR, es menor que 1 y el superior mayor que dicho valor, se interpreta como que no hay ninguna relación estadísticamente significativa, entre la exposición y la enfermedad o el evento de interés.

Al igual que RR, el cálculo del estimador OR se puede realizar a través de diversos programas estadísticos, sin embargo, para fines prácticos del presente capítulo a continuación se desarrolla la ruta a seguir para calcular el OR en el SPSS (figura 4).

Figura 4
Ruta que seguir para calcular el OR en SPSS



Para finalizar este apartado, se agrega un ejemplo del cálculo del OR
H0= A mayor PMPH inadecuada menor exceso de peso en los niños.

Los resultados de la aplicación de la ruta a seguir para calcular el OR mediante regresión logística binaria en el SPSS.

Las pruebas ómnibus, indican que los valores del modelo deben ser menores de 0.05 para que la PMPH puedan predecir mejor el estado nutricional del niño.

Pruebas ómnibus de coeficientes de modelo

		Chi-cuadrado	gl	Sig.
Paso 1	Paso	26,075	1	,000
	Bloque	26,075	1	,000
	Modelo	26,075	1	,000

Esta prueba se interpreta de manera similar a R cuadrada donde los valores oscilan entre 0 y 1. Por tanto, nuestro modelo explica el 47%.

Resumen del modelo

Paso	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	55,429 ^a	352	.474

Finalmente vamos a la tabla principal, la cual nos indica los valores de los coeficientes predictores en este caso PMPH.

		Variables en la ecuación					95% C.I. para EXP(B)		
		B	Error estándar	Wald	gl	Sig.	Exp(B)	Inferior	Superior
Paso 1 ^a	Percepción materna del peso del hijo por palabras	2,738	,661	17,166	1	< ,001	15,461	4,233	56,468
	Constante	-4,681	1,128	17,225	1	< ,001	,009		

a. Variables especificadas en el paso 1: Percepción materna del peso del hijo por palabras.

OR=15.46

Es así como se concluye que, la PMPH predice el estado nutricional del niño. A mayor PMPH inadecuada mayor exceso de peso en los niños. O bien, cuando las madres no perciben adecuadamente el peso del hijo, los niños tienen 15.46 veces mayor posibilidad de presentar exceso de peso.

Referencias

- AGUILAR Ticona Juan Pablo, Arriaga Gutiérrez María Belen, Chaves Torres Ninfa Marlen, Zeballos Rivas Diana Reyna. (2023). Entendiendo la Odds Ratio. *SCIENTIFICA* ; 15(1): 27-30. Disponible en: http://revistasbolivianas.umsa.bo/scielo.php?script=sci_arttext&pid=S1813-00542017000100008&lng=es.
- CERDA, Jaime, Vera, Claudio, & Rada, Gabriel. (2013). Odds ratio: aspectos teóricos y prácticos. *Revista médica de Chile*, 141(10), 1329-1335. <https://dx.doi.org/10.4067/S0034-98872013001000014>

- DAGNINO, J. (2014). Riesgo relativo y odds ratio (razón de ventajas). *Rev Chil Anest*, 43, 317-321.
- DE Irala J, Martínez González M. A. Seguí-Gómez M., (2008). Epidemiología aplicada. 2da Edición. Barcelona.
- FAJARDO-GUTIÉRREZ, Arturo (2017). Medición en epidemiología: prevalencia, incidencia, riesgo, medidas de impacto. *Revista alergia México*, 64(1), 109-120. <https://doi.org/10.29262/ram.v64i1.252>
- FUENTES Ferrer ME., del Prado González N. (2013). Desde el laboratorio a la clínica, medidas de frecuencia y de asociación en epidemiología clínica. *An Pediatr Contin*;11(6):346-349.
- JEREMY A Labrecque, Myriam M G Hunink, M Arfan Ikram, M Kamran Ikram, Do Case-Control Studies Always Estimate Odds Ratios?, (2021). *American Journal of Epidemiology*, Volume 190, Issue 2, February, Pages 318–321, <https://doi.org/10.1093/aje/kwaa167>
- KAPPES, María Soledad, & Riquelme, Verónica. (2021). El valor P, y medidas de efecto: su interpretación en investigación cuantitativa en enfermería. *Ene*, 15(2), 1247. Epub 31 de enero de 2022. Recuperado en 03 de octubre de 2023, de http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1988-348X2021000200004&lng=es&tlng=es.
- LAST J. (2001). Dictionary of Epidemiology. 4th ed. International Epidemiological Association. Oxford University Press.
- MIRON, C. & Sardon. M. (2008). Medidas de frecuencia asociación e impacto en investigación aplicada, *Journal Medicina y Seguridad del Trabajo*, 54 (211), 93-102.
- ORTIZ-FÉLIX, Rosario E., Cárdenas-Villarreal, Velia M., Miranda-Félix, Patricia E., & Guevara-Valtier, Milton C.. (2021). Impacto de una intervención de educación prenatal en mujeres embarazadas para prevenir el sobrepeso en lactantes. *Gaceta médica de México*, 157(1), 3-9. Epub 18 de junio de 2021. <https://doi.org/10.24875/gmm.19005610>
- PARDO, S.A. (2023). Probabilidades, odds ratios y comparación de proporciones. En: *Statistical Methods and Analyses for Medical Devices*. Springer, Cham. https://doi.org/10.1007/978-3-031-26139-8_17
- RENDÓN-MACÍAS, Mario Enrique, García, Heladia, & Villasis-Keever, Miguel Ángel. (2021). Medidas de riesgo, asociación e impacto en los estudios de investigación clínica. Cómo interpretarlas para su aplicación en la atención médica. *Revista alergia México*, 68(1), 65-75. 2021.<https://doi.org/10.29262/ram.v68i1.886>
- TAMARGO Barbeito, Teddy Osmin, Gutiérrez Rojas, Ángela Rosa, Quesada Peña, Susel, López León, Nirka, & Hidalgo Costa, Tania. (2019). Algunas consideraciones sobre aplicación, cálculo e interpretación de odds ratio y riesgo relativo. *Revista Cubana de Medicina*, 58(3), e497. Recuperado en

03 de octubre de 2023, de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0034-75232019000300008&lng=es&tlng=es.

TAPIA Granados J.A (1994). Incidencia; concepto, terminología y análisis dimensional. *Medicina Clínica*, 103:140-2.

Capítulo 16

Normas APA para la presentación de datos estadísticos (tablas, figuras y gráficos)

DCE ISAÍAS IVÁN BRICEÑO RODRÍGUEZ

Los trabajos de investigación generan una gran cantidad de información y datos estadísticos, los cuales tendrán que ser presentada ante los lectores de una manera sintética, organizada y eficiente que permita la comprensión e interpretación de los mismo, además deberá ser de una forma que sea atractiva visualmente para mantener la atención de aquellos que se encuentran en la búsqueda de información.

Son los datos una parte importante, el núcleo de una investigación y el segmento más valorado por los lectores, es en la sección de resultados donde los datos comunican a la comunidad científica, estos podrán tener aportaciones que impliquen modificaciones en el campo de conocimiento y en algunas ocasiones orientara la toma de decisiones. Es por ello, la importancia de conocer las principales consideraciones al momento de redactar y de presentar con mayor precisión la gran cantidad de datos encontrados, a través de una exposición sencilla, clara y precisa, son las tablas y figuras que complementan los resultados principales y secundarios y apoyen la argumentación textual en la sección de resultados (Castro et al., 2018).

Las tablas como elementos visuales son un conjunto de datos, cuya estructura debe de ser lo más simple posible y presentar la información en una secuencia lógica, fácilmente cuantificable y comprensible. Entre los datos que se pueden presentar por medio de tablas se encuentran algunos datos de diseño metodológico, características de los grupos, pruebas estadísticas. Con la finalidad de mejorar la comprensión de los resultados de una manera sencilla, permitiendo que el lector realizar interpretaciones y generar conclusiones que pueden ser utilizados para la toma de decisiones (Jiménez et al., 1999).

Tablas y Figuras

Formato General del Texto

Como primer paso, retomaremos una serie de recomendaciones que son aplicables a todo el texto establecidas por la American Psychological Association (APA, 2020), y las cuales se enlistan a continuación:

- 2.5 de margen

- Letra Times New Roman 12 pt, también pueden considerarse estilo:
 - Fuentes Sans Serif: Calibri y Arial 11 puntos, Lucida Sans Unicode 10 puntos
 - Fuentes Serif: Times New Roman de 12 puntos
- Texto a doble espacio y alineado a la izquierda, excepto en tablas y figuras.
- Sangría (Indet) a 5 espacios o ½ pulgada en todos los párrafos.
- Se enumeran utilizando números arábigos
- La secuencia se establece a partir del orden de aparición de las tablas en el texto.

Componentes de una Tabla

Estructura básica

La estructura básica está compuesta por columnas y filas, el número de estas dependerá de la cantidad de información que se desee plasmar. Primeramente, es necesario realizar un esquema de los datos que se presentaran, para calcular el número de filas y columnas que se necesitaran.

Tipo y tamaño de fuente e interlineado. La fuente será la misma que se utiliza en el resto del texto. El interlineado para el cuerpo de la tabla puede utilizarse, sencillo, uno y medio o interlineado doble, para seleccionar el interlineado se tomará en cuenta aquel que facilite la lectura de los resultados (ver Figura 1).

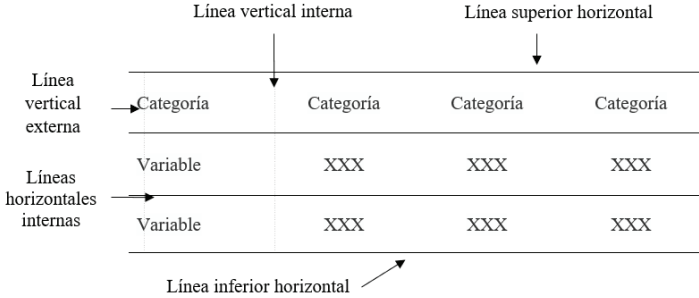
Figura 1

Identificación de filas y columnas de una tabla

Categoría	Categoría	Categoría	Categoría
Variable	XXX	XXX	XXX
Variable	XXX	XXX	XXX

Uso de líneas dentro de la tabla. Las líneas verticales externas e internas no serán visibles en el formato, las líneas horizontales serán visibles en la estructura de la tabla. Se mantendrá una línea superior horizontal que delimite el título de la tabla con los títulos de cada columna. Las líneas horizontales internas dividen el contenido de cada una de las filas, finalmente se colocará una línea inferior horizontal que marca el final de la tabla (ver Figura 2).

Figura 2
Uso de líneas en tablas estilo APA



Número y título de la tabla. El número de la tabla estará colocado en la parte superior, en negritas y alineado a la izquierda, usando números arábigos. El título será colocado por debajo del número de la tabla, estará alineados a la izquierda y estará escrito en cursiva. Recordemos que el título de la tabla deberá ser breve, describiendo el contenido central de la tabla. El interlineado entre el número de la tabla, el título y la línea superior de la tabla, será a doble espacio (ver Figura 3). El orden de las tablas es consecutivo, siguiendo el orden en el cual se mencionan en el texto.

Figura 3
Numero y titulo en tablas estilo APA



Categoría	Categoría	Categoría	Categoría
Variable 1	XXX	XXX	XXX
Variable 1	XXX	XXX	XXX

Encabezados de la tabla. El contenido de la tabla incluye una variedad de encabezados (títulos), estableciendo una organización de la información y permite identificar lo que se encuentra en cada una de las columnas. El encabezado proporcionará información del contenido de manera muy breve, el encabezado no será más ancho que la información que ese encontraba por debajo de él y cada una de las columnas deberá de tener un encabezado, algunos tendrán un encabezado secundario dentro de una misma columna (ver Figura 4). Todo en-

cabezado iniciara con una letra mayúscula y deberá estar centrado a menos que la alineación a la izquierda mejore la legibilidad.

Figura 4
Encabezado en tablas estilo APA

Tabla 1

Características sociodemográficas

Características	Encabezado Grupo A		Grupo B		Encabezado Secundario Grupo C	
	n	%	n	%	n	%
Genero						
Masculino	26	44.82	20	57.14	22	48.88
Femenino	32	55.17	15	42.85	23	51.11
Estado Civil						
Casado	37	63.79	15	42.85	23	51.11
Soltero	21	36.20	20	57.14	22	48.88

Columna Auxiliar Punto de Intersección Valor decimal

Columna auxiliar. La columna auxiliar permite, aquella que se encuentra a la izquierda de la tabla, identificar las principales variables independientes que serán explicadas con la información que se encuentre en las columnas siguientes. El contenido de una línea, que se encuentre en la columna auxiliar, se unirá en un punto (punto de intersección) de algunas de las columnas siguientes, permitiendo la interpretación de los resultados obtenidos (ver Figura 4).

Valor decimal. Los resultados numéricos serán reportados con el numero decimal que permita una mejor precisión de la medición, para ello, el resto de los resultados deberá de contener la misma cantidad de decimales. Los valores estarán alineado a la izquierda o en la parte central, si ello permite mejor la lectura y comprensión de los resultados. Cundo no exista dato por colocar en alguna de las celdas se dejará el espacio en blanco o se colocara un guion, colocando una nota explicando al pie de la tabla el ¿Por qué? la falta de información (ver Figura 4).

Notas de la tabla. Una nota se coloca por debajo del cuerpo de la tabla, la cual es aplicable solo a la tabla que se describe, en caso de que la nota aplique para otra tabla, se colocara la misma descripción para relacionar la información de ambas tablas con el objetivo de leerlas e interpretarlas en su conjunto. Existen tres tipos de notas, generales, específicas y notas de probabilidad, las notas generales explican o proporcionan información con relación al contenido, puede explicar abreviatura o símbolos utilizados dentro de la misma; para su identificación se colocará la palabra “Nota” en formato cursiva, seguida de

un punto. Las notas específicas, hacen referencia a una columna, fila o celda, para identificarlas se colocará una letra minúscula en superíndice. Las notas de probabilidad describen los símbolos para indicar el valor de p, utilizando dos o tres decimales para proporcionar una probabilidad más exacta (por ejemplo, $p = .023$), cuando el valor es menor a “.001” se colocará el valor como “<.001”. Cada tipo de nota iniciara por debajo del cuerpo de la tabla, inicia con una nota general seguida de una nota específica en una nueva línea por debajo de la nota general, una nota de probabilidad comenzara en una línea debajo de una nota general o específica. Cuando existen más de una nota específica de probabilidad, continuaran en la misma línea separadas entre sí por un punto y un espacio (ver Figura 5).

Es uso de la nota permite eliminar la repetición dentro del cuerpo de una tabla. Cuando exista valores de probabilidad o los tamaños los datos son numerosos, se agrega una columna para evitar muchas notas; caso contrario, si dentro de una fila o columna contiene pocas entradas se eliminará o se evitara el uso de una columna para agregar una nota a la tabla. Dentro del formato de la tabla no use bordes verticales para separar los datos, evite el uso del sombreado o cursiva dentro de la tabla; se puede usar cursiva o negrita con explicación dentro de la nota general de la tabla.

Figura 5
Tipos de notas en tablas estilo APA

Tipo de bienestar	N	X	DE	t	gl	p
Bienestar físico	18 ^a	21.95	1.61	-0.310	38	0.758 [*]
Bienestar emocional	20	22.10	1.45			

Nota. t = prueba “t” de Student. DE = Desviación Estándar. gl = grados de libertad.
 a Dos participantes no completaron las mediciones.
 * $p < .05$

Tablas largas o anchas. Si una tabla sobrepasa el largo de la página esta será dividida colocando nuevamente la línea de encabezados para poder ser identificada.

Figura 5
Formato para la división de tablas largas en formato APA

Características	Grupo A		Grupo B		Grupo C	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Genero	26					
Masculino	32	44.82	20	57.14	22	48.88
Femenino		55.17	15	42.85	23	51.11

Características	Grupo A		Grupo B		Grupo C	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Estado Civil						
Casado	37	63.79	15	42.85	23	51.11
Soltero	21	36.20	20	57.14	22	48.88

Las tablas cuya anchura sobrepasan los márgenes de la hoja pueden solventarse realizando modificaciones en la orientación de la hoja, de estar en vertical colocarla en horizontal. Si a pesar de ello, la tabla continúa sobrepasando los márgenes establecidos, se dividirá la tabla manteniendo el contenido de la columna auxiliar. En caso de colocar alguna nota, estas estarán colocadas al final de la segunda tabla.

Figura 6
Formato para la división de tablas en formato APA

Características	Grupo A		Grupo B		Grupo C	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Genero	26					
Masculino	32	44.82	20	57.14	22	48.88
Femenino		55.17	15	42.85	23	51.11

Características	Grupo D		Grupo E		Grupo F	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Genero						
Masculino	29	63.79	35	42.85	28	51.11
Femenino	27	36.20	25	57.14	27	48.88

Componentes de una Tabla con Datos Cualitativos

Los datos cualitativos se pueden presentar en tablas, adecuándolos a la mejor forma en la cual se puedan leer e interpretar los resultados. Pueden presentarse datos cualitativos o datos que incluyan resultados cualitativos y cuantitativos, limitado a presentar solo frecuencias y porcentajes de las categorías presentadas.

Figura 7

Formato para la presentación de datos cualitativos en formato APA

Tabla 1 “Motivos” consumo de drogas en adolescentes

Tema	Subtemas		f	%
Todo acto social tiene una intención y un significado. el significado de una acción se objetiva. Cada acción nace del acontecimiento y se autonomiza respecto de su agente. Lo cual constituye la dimensión social de la acción	Motivos para	<i>“...recuerdo que el primer día fue con marihuana y pues me agradó el sentirme así, el cómo me puso, me sentía bien, me sentía contento, me reía por todo, fue en sí una experiencia chida que todavía recuerdo...” (Q1)</i>	6	30
	Son los objetivos que se pretenden alcanzar con la acción, estos motivos son conscientes y están dominados por el tiempo futuro.	<i>“Ese primer día pues muy feliz era porque empecé a fumar marihuana... me daba un frigo de risa, me gustaba el efecto y seguía fumando...quería ser igual que ellos...” (R1)</i>	7	35
		<i>...se dio entrando a la secundaria y no entré y me fui ahí con los cholos...y empezamos a consumir...nosotros podemos decir que taba chida la marihuana...se me venían a la mente recuerdos tristes...cosas de mi infancia...pues que me sacaban un sentimiento no y empecé a llorar.. (C1)</i>	7	35
	“Motivos porque”	<i>“Pues fue primero por curiosidad, porque así que decía para saber ¿qué se siente?, ¿por qué lo hacen?, a ver si es cierto, a ver si esta chido...” (Q1)</i>	6	30
	Son los que se explica sobre la base de los antecedentes, ambiente o predisposición psíquica del actor, están dominados por el tiempo pasado, es una cuestión causal.	<i>“ porque quería era, que ¿por qué me habían dejado cuando estaba morrillo?, pos sí desde morrillo empecé a drogarme. Quería ser igual que ellos...y yo como estaba chiquillo dije pos es barrio era, yo quiero ser de ese barrio...” (R1)</i>	6	30
		<i>“...con un compa que estaba ahí, ya consumía marihuana y le dije dame y dijo que no y le dije que me diera y al último me dio... me llamaba la atención, a ver que se sentía por eso le pedí...” (C1)</i>	8	40

Figuras

Una figura es un elemento visual, una ilustración o una imagen que no es una tabla. Dado que su estructura no se presenta por medio de filas o columnas.

Los componentes básicos de una figura

Numero. El numero de la figura se colocará en la parte superior de la figura en negrita. Las figuras que formen parte del texto serán numeradas usando números arábigos

Título. Cada figura deberá contener un título que sea breve y que sea explicativo para el contenido de la figura. El título será colocado por debajo del número a doble espacio en formato cursiva, por debajo del número de la figura, con espaciado a doble espacio.

Imagen. Una figura será considerada un gráfico, una fotografía, un dibujo, un cuadro, un diagrama u otra ilustración.

Leyenda. Una leyenda de la figura debe de colocarse dentro de los bordes de la figura, en la cual se puede utilizar para explicar los símbolos utilizados en la figura.

Nota. Las notas permiten describir el contenido de la figura que no puede ser entendido a partir del título de la figura, la imagen o la leyenda. No todas las figuras deberán de incluir notas, solo aquellas que permitan complementar o explicar la figura. Al igual que en las tablas, pueden colocarse tres tipos de notas, general, específica y de probabilidad, siguiendo los mismos lineamientos ya establecidos.

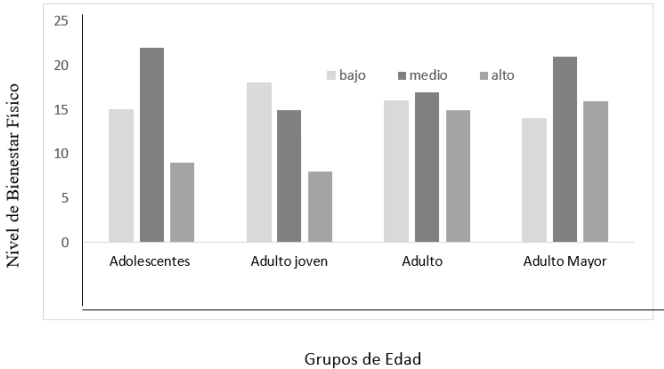
Sombreado en imágenes. En caso de utilizar sombreado para para distinguir dos o más conjuntos de barras o líneas, lo más idóneo es utilizar dos conjuntos de sombrados, en caso de necesitar más de dos sombrados utilizar patrones para la identificación.

Color en imágenes. El color, por lo general, es utilizado con fines decorativos. No obstante, en la búsqueda de publicación de resultados se deberá evitar el color, a menos que el medio de publicación lo permita o sea indispensable para la comprensión de la información, se utilizará una escala de grises para hacer la diferencia en la información presentada.

Figura 8
Elementos básicos en figura formato APA

Figura 1

Nivel de bienestar emocional en pacientes hospitalizados



Nota. Esta figura representa los elementos básicos de una figura.

Nota específica. Una nota específica describe algún elemento específico que permitirá comprender la gráfica.

Nota de probabilidad. Describe algunos resultados específicos, como el valor de p , o algunos elementos específicos de probabilidad.

Fotografías

Las fotografías son un tipo de figuras, pueden imprimirse en color o en casas de grises, en ambos casos el contraste deberá garantizar el contenido, dependerá del contenido de la fotografía y el lugar de la publicación. Para publicación deberá de consultar las pautas de la editorial. Los principios éticos de toda publicación prohíben tergiversar las imágenes, así como la publicación por medio de fotografía en la cual se pueda identificar una persona, a menos que se cuente con una autorización firmada.

Datos biológicos

Los resultados electrofisiológicos, radiológicos, genéticos son considerados biológicos. Cada una de las imágenes incluirá una nota que describa los mecanismos de procesamiento o mejora de las imágenes, cuando se presentan más de una figura con el mismo tema, mantenga los mismos elementos de estilo y formato. Las notas permitirán describir cada uno de los detalles de la imagen, ya esta puede mejorar el valor comunicativo de la figura.

Referencias

- AMERICAN Psychological Association. (2020). Publication manual of the American Psychological Association (7th ed.). <http://doi.org/10.1037/0000165-000>
- CASTRO, Y., Mattos, M., Aliaga A. (2018). Consideraciones en redacción científica: los resultado, tablas y figuras. *Odontología Sanmarquina*, 21(3): 241-248.
- JIMÉNEZ, J., Argimón J. (1999). Artículo original (IV): Tablas y figuras. *Atención Primaria*, 23(5): 310-311.

Coordinadores

Carrillo Cervantes Ana Laura

Doctora en Ciencias de Enfermería (Universidad de Guanajuato). Maestra en Ciencias de Enfermería (UANL). Licenciada en Enfermería (UAdeC). Profesor de Tiempo Completo, Facultad de Enfermería “Dr. Santiago Valdés Galindo” Universidad Autónoma de Coahuila. Perfil deseable PRODEP. Certificación Docente COMCE. Investigador del Sistema Nacional de Investigadores Nivel Candidato, Investigador Honorífico del Sistema Estatal de Investigadores Coahuila. Miembro de la Red de Enfermería en Salud del Adulto Mayor México (Red-ESAM). Miembro Red CRONIFAM. Miembro del Capítulo Tau Alpha, Sigma Theta Tau International. Integrante del Cuerpo académico “Cuidado para el envejecimiento activo y saludable”. Autor de libros, capítulo de libros, artículos científicos y contenido digital (Capítulo 8).

Cortes Montelongo Diana Berenice

Doctora en Ciencias de Enfermería (UANL). Máster en Administración de la Atención de Enfermería (UASLP). Licenciada en Enfermería (UAdeC). Profesor de Tiempo Completo, Facultad de Enfermería “Dr. Santiago Valdés Galindo Universidad Autónoma de Coahuila. Unidad Saltillo. Saltillo, Coahuila. Perfil deseable PRODEP. Certificación Docente COMCE. Investigador del Sistema Nacional de Investigadores Nivel Candidato, Investigador Honorífico del Sistema Estatal de Investigadores Coahuila. Miembro de la Red de Enfermería en Salud del Adulto Mayor México (Red-ESAM). Miembro Red CRONIFAM. Miembro del Capítulo Tau Alpha, Sigma Theta Tau International. Responsable del Cuerpo académico “Cuidado para el envejecimiento activo y saludable”. Autor de libros, capítulo de libros, artículos científicos y contenido digital (Capítulo 1).

Cortez González Luis Carlos

Doctor en Ciencias de Enfermería. Máster en Enfermería con Acentuación en la Atención del Adulto Mayor. Licenciado en Enfermería. Profesor de Tiempo Completo, Facultad de Enfermería “Dr. Santiago Valdés Galindo” Universidad Autónoma de Coahuila. Unidad Saltillo. Saltillo, Coahuila, 25000, México.

Perfil deseable PRODEP. Certificación Docente COMCE. Candidato a Investigador del Sistema Nacional de Investigadores. Investigador Honorífico del Sistema Estatal de Investigadores Coahuila. Miembro de la Red de Enfermería en Salud del Adulto Mayor-México (Red-ESAM). Miembro Red CRO-NIFAM. Miembro del Colegio de Actualización de Enfermería de Coahuila. Líder Académico GREEFES Saltillo. Integrante del Cuerpo académico “Cuidado para el envejecimiento activo y saludable” (Capítulo 7).

Colaboradores

Amaro Hinojosa Marily Daniela

Doctora en Ciencias de Enfermería; Profesor de Tiempo Completo de la Universidad Autónoma de Chihuahua; Miembro del Sistema Nacional de Investigadores (candidato), Sigma Theta Tau Internacional, Capítulo Tau Alpha, Cuerpo en Formación Respuestas humanas a la salud sexual y reproductiva. Coordinador del Programa Doctorado en Enfermería de la FEN-UACH. Autor y coautor de Investigaciones en enfermería, relacionadas con conductas sexuales en jóvenes y autoeficacia para prevenir VIH (Capítulo 4)

Briceño Rodríguez Isaías Iván

Doctor en Ciencias de Enfermería y Maestro en Enfermería, por la Universidad de Guanajuato. Especialidad en Medicina Interna y Terapia Intensiva y Licenciado en Enfermería por la Facultad de Enfermería y Obstetricia de la Universidad Juárez del Estado de Durango. Certificaciones como Docente y Licenciado en Enfermería por el COMCE, Elaboración de Notas de Enfermería. REMEMPRO – IMEXCE, Perito en Enfermería. Instituto Nacional de Enfermería Pericial, Centro de Capacitación Tabasco S.C. – Universidad Salazar, Instituto de estudios Superiores de Chiapas. Miembro de la Red Mexicana de Enfermería Pericial y Procesos S.C. REMEMPRO y al Colegiado en el Colegio de Enfermeras Duranguenses. Miembro del sistema estatal de investigadora nivel 1 (Durango). Profesor de Tiempo Completo de la Facultad de Enfermería y Obstétrica de la Universidad Juárez del Estado de Durango. Enfermero General del Hospital General “Dr. Santiago Ramon y Cajal” ISSSTE Durango (Capítulo 16).

Cruz Quevedo Juana Edith

Académico de Carrera de Tiempo Completo Titular “C”, adscrita a la Facultad de Enfermería, Universidad Veracruzana, Región Veracruz. Licenciada en Enfermería por la Universidad Veracruzana. Maestría y Doctorado en Ciencias de Enfermería por la Universidad Autónoma de Nuevo León. Coordinadora del Programa de Doctorado en Ciencias de la Enfermería. Miembro de la Red de Enfermería en Salud del Adulto Mayor y de la Red de Automanejo en Enfermedades Crónicas. Responsable del CA Cuidado del Ser Humano (UV-

CA-357). Desarrolla la Línea de investigación de Funcionalidad en el adulto mayor (Capítulo 11).

Guerra Ordoñez Jesús Alejandro

Doctorado en Ciencias de Enfermería. Profesor de Tiempo Completo e Investigador de la Unidad Académica Multidisciplinaria Matamoros de la Universidad Autónoma de Tamaulipas. Sistema Nacional de Investigadores Nivel I. Perfil PRODEP. Miembro de la Red de Investigación en Sexualidad. Líder del Cuerpo Académico Consolidado Salud y Comportamiento Humano (Capítulo 3).

Delabra Salinas María Magdalena

Doctor en Ciencias de Enfermería por la Universidad Autónoma de Nuevo León. Profesor de tiempo completo en la Facultad de Enfermería Unidad Saltillo de la Universidad Autónoma de Coahuila. Miembro del NAB del programa de Maestría en Enfermería con atención al adulto mayor e integrante del Cuerpo Académico “Cuidado para el envejecimiento activo y saludable”. Miembro del SNI nivel 1, Sigma Theta Tau (Capítulo Tau Alpha), Red ESAM-México, Red CRONIFAM, Red Temática de Automanejo y Red Multidisciplinaria de cuidado, tecnología, conducta y salud”. Certificada como Máster training de automanejo de enfermedades crónicas (2016-2020). Certificación en competencias docentes (2022-2027). Trabajo en proyectos comunitarios y de prevención como riesgo de pie diabético, cuidado espiritual de enfermería, incontinencia urinaria con adultos mayores y sus cuidadores; fomento a la lactancia materna y parto respetado (Capítulo 6).

Duran Badillo Tirso

Doctor en Metodología de la Enseñanza, Profesor de Tiempo Completo y Jefe de la División de Estudios de Posgrado en la Unidad Académica Multidisciplinaria Matamoros de la Universidad Autónoma de Tamaulipas. Integrante del Cuerpo Académico “Salud y Comportamiento Humano”. Miembro de la Red ESAM México. Coordinador del Comité de formación de estudiantes de pregrado – GREEFES de la Red ESAM México. Fundador de los Grupos de Estudiantes de Enfermería para el Fomento del Envejecimiento Saludable (GREEFES). Perfil deseable PRODEP. Sistema Nacional de Investigadores Nivel I (Capítulo 3).

Domínguez-Chávez Claudia Jennifer

Doctora en Ciencias de Enfermería; Maestra en Ciencias de Enfermería; Licenciada en Enfermería. Profesora-Investigadora de Tiempo Completo de la Fa-

cultad de Enfermería de la Universidad Autónoma de Baja California. Miembro del Sistema Nacional de Investigadores. El área de interés en la investigación se centra en el envejecimiento, vejez y el estudio del declive cognitivo y motriz y uso de terapias complementarias en personas adultas mayores (Capítulo 9).

Esquivel Rubio Abraham Isaac

Doctor en Ciencias de Enfermería, Maestro en Enfermería, Licenciado en Enfermería y Especialista en Migración Internacional. Profesor-Investigador en la Facultad de Enfermería de la Universidad Autónoma de Baja California y miembro del Sistema Nacional de Investigadores, Nivel 1. Producción académica y científica en artículos y capítulos de libro en las temáticas de Salud intercultural, migración y prevención de VIH/ITS en grupos vulnerables (Capítulo 9).

Gómez-Melasio Dafne Astrid

Docente y coordinadora del Departamento de Servicio Social de la Facultad de Enfermería Unidad Saltillo de la Universidad Autónoma de Coahuila. Maestra en Enfermería por la Facultad de Enfermería de la UAdeC. Doctora en Ciencias de Enfermería por la Facultad de Enfermería de la Universidad Autónoma de Nuevos León. Integrante del cuerpo académico Prevención de Conductas Adictivas, y colaboradora del cuerpo académico Investigación Educativa en Enfermería y Salud de la UAdeC. Miembro nivel 1 del Sistema Nacional de Investigadores del CONAHCYT (Capítulo 12).

Medina Fernández Josué Arturo

Licenciado en Enfermería, Maestro en Enfermería y Doctor en Salud Pública. Profesor e Investigador de Tiempo Completo en el Departamento de Ciencias de Enfermería de la División de Ciencias de la Salud de la Universidad Autónoma del Estado de Quintana Roo. Integrante del cuerpo académico “Enfermería, educación y salud”. Estudia el comportamiento del cuidado del adulto mayor y sus enfermedades crónicas. Cuenta con el Reconocimiento al Perfil Deseable (PRODEP) (Capítulo 10).

Medina Fernández Isaí Arturo

Doctor en Salud Pública por la UNICLA, Maestro en Enfermería por la UAdeC y Licenciado en Enfermería por la UADY. Profesor de Tiempo Completo de la Facultad de Enfermería Unidad Saltillo de la Universidad Autónoma de Coahuila. Integrante del CA Cuidado para el envejecimiento activo y saluda-

ble”. Es autor de artículos científicos indexadas en revistas nacionales e internacionales, así como autor de diversos capítulos de libros, director de tesis de licenciatura y posgrado. Cuenta con perfil PRODEP, certificado ante COMCE e investigador Junior por el COECYT (Capítulo 12).

Miranda Félix Patricia Enedina

Profesora investigadora perfil PRODEP, Facultad de Enfermería Mochis, Universidad Autónoma de Sinaloa, Doctora en ciencias de Enfermería. Miembro del Sistema Nacional de investigadores (CONAHCYT), Líder del grupo de investigación; Estados Crónicos y Entorno Biopsicosocial, presidenta del Comité de Ética en investigación. Participa en redes de trabajo colegiado nacional e internacional (Sigma, ARFE, Red de Docentes de América Latina y del Caribe-RedDOLAC), autor y coautor en diversas publicaciones en revistas indizadas y JCR de circulación nacional e internacional (Capítulo 15).

Navarro Oliva Edna Idalia Paulina

Licenciada en Enfermería, Maestra y Doctora en Ciencias de Enfermería y Postdoctoranda en Enfermería Psiquiátrica. Profesor e Investigador de Tiempo Completo de la Facultad de Enfermería “Dr. Santiago Valdés Galindo” de la Universidad Autónoma de Coahuila. Líder del Cuerpo Académico “Prevención de Conductas Adictivas”. Pertenecer al Sistema Nacional de Investigadores (SNI) Nivel I del Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT) y al Sistema Estatal de Investigadores en el Nivel Émerito del Consejo Estatal de Ciencia y Tecnología de Coahuila (COECyT). Cuenta con el Reconocimiento al Perfil Deseable (PRODEP) (Capítulo 14).

Nuncio Domínguez José Luis

Doctor en Ciencias Sociales, Maestro en Administración de la Atención de Enfermería, Profesor de Tiempo Completo con Perfil Deseable PRODEP, Catedrático Investigador, Miembro del Cuerpo Académico de Administración en Servicios de Enfermería y Salud, colaborador en el Cuerpo Académico Cuidado Integral para la Atención al Adulto Mayor, Facultad de Enfermería “Dr. Santiago Valdés Galindo” Universidad Autónoma de Coahuila. Estudiante de Doctorado en Ciencias Sociales con acentuación en grupos vulnerables, Facultad de Trabajo Social de la Universidad Autónoma de Coahuila, Certificación Nacional como Enfermero Docente por el Consejo Mexicano de Certificación de Enfermería, Miembro Red ESAM-México, Miembro del Colegio de Actualización de Enfermería de Coahuila (Capítulo 2).

Ortiz Félix Rosario Edith

Profesora investigadora perfil PRODEP, Facultad de Enfermería Mochis, Universidad Autónoma de Sinaloa, Doctora en Ciencias de Enfermería. Miembro del Sistema Nacional de investigadores (CONAHCYT) Nivel, coordinadora del programa de Maestría y doctorado en Ciencias de Enfermería, miembro del grupo de investigación; Estados Crónicos y Entorno Biopsicosocial, participa en redes de trabajo colegiado nacional e internacional (Sigma, ARFE, Red de Docentes de América Latina y del Caribe-RedDOLAC), autor y coautor en diversas publicaciones en revistas indizadas y JCR de circulación nacional e internacional (Capítulo 15).

Orozco Gómez Claudia

Doctora en Ciencias de Enfermería por la Universidad Autónoma de Nuevo León. Profesora de tiempo completo en la Facultad de Enfermería y Nutriología de la Universidad Autónoma de Chihuahua. Miembro de la Red Temática de Investigación en Sexualidad Responsable y del Cuerpo en formación “Respuestas humanas a la salud sexual y reproductiva. Miembro del Sistema Nacional de Investigadores nivel candidata (Capítulo 4).

Rios Willars Ernesto

Doctor en Biotecnología y Máster en Informática con acentuación en Sistemas de Información por la Universidad Autónoma de Coahuila. Máster en Educación por la Universidad Interamericana para el Desarrollo. Profesor de asignatura en Sistemas bioinformáticos y taller de investigación en la Facultad de Sistemas de la UAdeC. Profesor de medio tiempo en la especialidad de electrónica y comunicaciones en el Centro de Bachillerato Tecnológico y de Servicios (CBTiS 235). Solicitud de registro de modelos de innovación gerontecnológica y de modelos hidropónicos. Miembro de la Sociedad Mexicana de Ingeniería Biomédica, Red Multidisciplinaria de cuidado, tecnología, conducta y salud. Trabajo en proyectos de desarrollo de algoritmos genéticos de alto rendimiento; proyectos gerontecnológicos, electrónicos y matemáticos (Capítulo 6).

Sifuentes Leura Daniel

Doctor en Ciencias de Enfermería. Máster en Ciencias de Enfermería. Profesor de Tiempo Completo, Facultad de Enfermería “Dr. Santiago Valdés Galindo” Universidad Autónoma de Coahuila. Unidad Saltillo. Saltillo, Coahuila, 25000, México. Perfil deseable PRODEP. Certificación Docente COMCE. Candidato

a Investigador del Sistema Nacional de Investigadores. Miembro de la Red de Enfermería en Salud del Adulto Mayor-México (Red-ESAM). Miembro Red CRONIFAM. Integrante del Cuerpo académico “Cuidado para el envejecimiento activo y saludable” (Capítulo 5).

Tenorio Gómez Marcela

Académico de Carrera de Tiempo Completo Titular “C”, adscrita a la Facultad de Enfermería, Universidad Veracruzana, Región Veracruz. Licenciada en Enfermería con Maestría en Enfermería (grado obtenido con Mención Honorífica), ambos por la Universidad Veracruzana. Doctorado en Ciencias Jurídicas, Administrativas y de la Educación. Miembro del Cuerpo Académico Cuidado del Ser Humano UV-CA 357. Desarrolla la línea de investigación de Funcionalidad en el adulto mayor (Capítulo 11).

Torres Obregón Reyna

Doctora en Ciencias de Enfermería por la UANL, Profesora e investigadora de Tiempo Completo de la Facultad de Enfermería Unidad Saltillo de la Universidad Autónoma de Coahuila. Miembro del Sistema Nacional de Investigadores (SNI) del CONAHCYT, Coordinadora del Departamento de Posgrado e Investigación de la Facultad de Enfermería U.S. y Coordinadora del Programa de Maestría en Enfermería (PNPC, 2454). Integrante del CA Cuidado para el envejecimiento activo y saludable”. Es autora de artículos científicos indexadas en revistas nacionales e internacionales, así como autor de diversos capítulos de libros, director de tesis de licenciatura y posgrado. Cuenta con perfil PRODEP, certificado ante COMCE e investigador emérito por el COECYT (Capítulo 13).

Torres Soto Nissa Yaing

Licenciada en Psicología Social Comunitaria, Maestra en Innovación Educativa y Doctora en Ciencias Sociales. Profesora e Investigadora de Tiempo Completo en la Universidad Autónoma del Estado de Quintana Roo (UQROO). Directora de la Secretaría Técnica de Investigación y Extensión (STIE). Es integrante del miembro del Comité de Ética en Investigación en la DCS UQROO. Cuenta con Perfil PRODEP y forma parte del Sistema Nacional de Investigadores Nivel 1.

Bioestadística:
Guía para un programa académico

se terminó de editar en Grupo Editorial Biblioteca, S.A. de C.V.
ubicados en Manantiales 29, Colonia Chapultepec,
Cuernavaca, Morelos, C.P. 62450
en el mes de diciembre de 2023.

Su edición consta de 600 ejemplares.

La presente obra aborda la crisis de replicabilidad en la ciencia, destacando la preocupación por la dificultad de reproducir resultados de estudios científicos. Se señalan dos razones clave detrás de esta crisis: la interpretación inadecuada de pruebas estadísticas y la tendencia a publicar resultados positivos, generando una visión sesgada de la realidad. En respuesta a esta problemática, el análisis de datos se presenta como una herramienta esencial, especialmente en el campo de la salud. El libro "Análisis Estadístico: Fundamentos y Aplicaciones" reúne la experiencia de destacados profesionales, abordando desde conceptos básicos hasta temas específicos como pruebas de hipótesis y análisis de regresión, proporcionando una guía integral para el análisis estadístico. La obra no solo destaca por su contenido sustantivo, sino también por su enfoque pedagógico, convirtiéndola en una herramienta educativa efectiva para estudiantes y profesionales.

Este libro no solo es un recurso de referencia, sino también un compañero invaluable en el viaje hacia la comprensión profunda y aplicada de la estadística. Cada capítulo contribuye a la construcción de una base sólida para abordar los desafíos y aplicaciones prácticas de la estadística, y su enfoque pedagógico lo hace accesible a una amplia audiencia. En resumen, "Análisis Estadístico: Fundamentos y Aplicaciones" no solo proporciona teorías abstractas, sino que también se centra en su aplicabilidad práctica, guiando a los lectores a través de la complejidad del análisis estadístico y equipándolos con las herramientas necesarias para enfrentar los desafíos estadísticos con confianza y competencia.



9 786076 959145 >



LA BIBLIOTECA