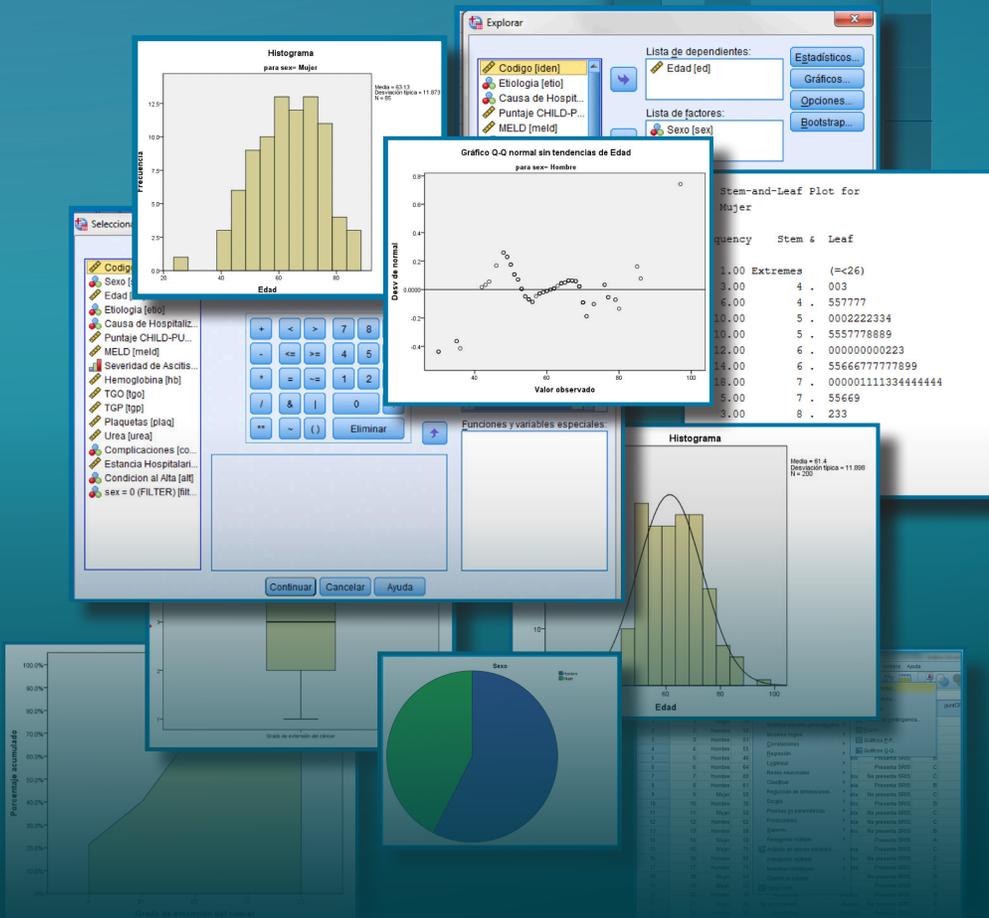


# BIOESTADÍSTICA EN LA INVESTIGACIÓN CLÍNICA: DE LA TEORÍA A LA PRÁCTICA

## análisis exploratorio univariado de datos



**Jorge Tarrillo-Purisaca**  
**P. Martín Padilla-Machaca**  
**Elsi Bazán-Rodríguez**



**FONDO EDITORIAL COMUNICACIONAL**

# **BIOESTADÍSTICA EN LA INVESTIGACIÓN CLÍNICA:**

DE LA TEORÍA A LA PRÁCTICA

análisis exploratorio univariado de datos



# BIOESTADÍSTICA EN LA INVESTIGACIÓN CLÍNICA:

## DE LA TEORÍA A LA PRÁCTICA

### análisis exploratorio univariado de datos

#### **Jorge Manuel Tarrillo-Purisaca**

*Médico epidemiólogo.*

*Egresado de la Maestría de Salud Pública, con mención en Epidemiología,  
y del Doctorado en Medicina.*

*Asesor en Salud y Epidemiología.*

#### **P. Martín Padilla-Machaca**

*Médico gastroenterólogo. Unidad de Hígado,*

*Hospital Nacional Guillermo Almenara Irigoyen, EsSalud.*

*Doctor en Salud Pública.*

*Maestro en Docencia Universitaria y Gestión Educativa.*

*Profesor de Medicina,*

*Universidad Nacional Mayor de San Marcos (UNMSM).*

*Docente de Maestría en Medicina,*

*Universidad de San Martín de Porres*

*Miembro del equipo de investigación Hepato, UNMSM.*

#### **Elsi Noemí Bazán-Rodríguez**

*Licenciada en Estadística en la Gestión de Servicios de Salud.*

*Maestra en Salud Pública.*

*Egresada de la Maestría en Salud Pública con mención en Epidemiología*

*Egresada del Doctorado en Salud Pública.*

*Docente de Medicina Humana. Universidad Privada San Juan Bautista (UPSJB).*

*Miembro del equipo de Investigación UPSJB.*



FONDO  
EDITORIAL  
COMUNICACIONAL

# Bioestadística en la investigación clínica:

DE LA TEORÍA A LA PRÁCTICA  
análisis exploratorio univariado de datos

---

## **Autores**

Jorge Manuel Tarrillo Purisaca

P. Martín Padilla-Machaca

Elsi Bazán Rodríguez

## **Editado por:**

Fondo Editorial Comunicacional  
Colegio Médico del Perú

Av, 28 de Julio 776,  
Miraflores. Lima 18, Perú  
Teléf. (01) 641-9847



1.ª edición  
Octubre de 2023

Tiraje: 500

HECHO EL DEPÓSITO LEGAL EN  
LA BIBLIOTECA NACIONAL DEL PERÚ N.º 2023-10488.

ISBN N.º 978-612-49026-8-0.

Todos los derechos reservados. Queda prohibida la reproducción total o parcial de esta obra protegida por los derechos de propiedad intelectual, o su uso en cualquier forma, o por cualquier medio, ya sea electrónico o mecánico, incluidos fotocopiado, grabación, transmisión o cualquier sistema de almacenamiento y recuperación de información, sin el permiso por escrito de los propietarios del copyright.

## **Asesoría editorial**

REP S.A.C.  
Miguel de Cervantes 485-502, San Isidro



## **Médico editor**

Jorge Candiotti Vera  
jcandiotti@revistasespecializadas.com  
999-65 85 31

## **Impresión**

Octubre de 2023  
Talleres gráficos de REP S.A.C.  
Emilio Althaus 355, Lima 14, Perú  
999-658531

## COMITÉ EJECUTIVO NACIONAL PERÍODO 2022-2024



- Decano
  - Vicedecano
  - Secretaria del Interior
  - Secretario del Exterior
  - Tesorero
  - Vocales
- Accesitarias
- Dr. José Raúl Urquiza Aréstegui
  - Dr. Alfredo Alonso Celis López
  - Dra. Wilda Cecilia Silva Rojas
  - Dr. Víctor Leonel Llaca Saravia
  - Dr. Wilder Alberto Díaz Correa
  - Dr. Herminio Renán Hernández Díaz
  - Dra. Amelia Cerrate Ángeles
  - Dr. César Augusto Portella Díaz
  - Dr. Leslie Marcial Soto Arquínigo
  - Dra. Celia Betzabet Moisés Alfaro
  - Dra. Milagros Dalila Sánchez Torrejón

### DECANOS DE LOS CONSEJOS REGIONALES DEL COLEGIO MÉDICO DEL PERÚ

- Decana Regional I - Trujillo
  - Decano Regional II - Iquitos
  - Decano Regional III - Lima
  - Decana Regional IV - Huancayo
  - Decano Regional V - Arequipa
  - Decana Regional VI - Cusco
  - Decano Regional VII - Piura
  - Decano Regional VIII - Chiclayo
  - Decano Regional IX - Ica
  - Decano Regional X - Huánuco
  - Decano Regional XI - Huaraz
  - Decano Regional XII - Tacna
  - Decano Regional XIII - Pucallpa
  - Decano Regional XIV - Puno
  - Decano Regional XV - San Martín
  - Decano Regional XVI - Ayacucho
  - Decana Regional XVII - Cajamarca
  - Decano Regional XVIII - Callao
  - Decano Regional XIX - Chimbote
  - Decano Regional XX - Pasco
  - Decano Regional XXI - Moquegua
  - Decano Regional XXII - Abancay
  - Decana Regional XXIII - Tumbes
  - Decano Regional XXIV - Huancavelica
  - Decano Regional XXV - Amazonas
  - Decano Regional XXVI - Madre de Dios
  - Decano Regional XXVII - Lima Provincias
- Dra. Elena Victoria Ríos de Edwards
  - Dr. Miguel Ángel Pinedo Saboya
  - Dr. Ildauro Aguirre Sosa
  - Dra. Armida Concepción Rojas Dávila de Izaguirre
  - Dr. Antony Gustavo Tohalino Meza
  - Dra. Eliana Janette Ojeda Lazo
  - Dr. Christian Yuri Requena Palacios
  - Dr. Ronald Jimmy Agüero Acuña
  - Dr. Luis Felipe Muñante Aparcana
  - Dr. Andrei Alekseevich Kochubei Hurtado
  - Dr. Alberto Fernando Del Valle Espejo
  - Dr. Jorge Eliseo López Claros
  - Dr. Carlos Abelardo Morales Hernández
  - Dr. Carlos Alberto Neira Ortega
  - Dr. Efraín Salazar Tito
  - Dr. Waldo Franz López Gutiérrez
  - Dra. Patricia Isabel Ocampo Quito
  - Dr. Luis Alberto Ortiz Pilco
  - Dr. Carlos Humberto Quiroz Urquiza
  - Dr. Manuel Alejandro Pomazono Goyas
  - Dr. Giancarlo Urquiza Pereira
  - Dr. José Luis Osorio Ticona
  - Dra. Lourdes Liliana Feijóo Oyola
  - Dr. Lino Elmer Rodríguez Julcamayán
  - Dr. Jorge Arturo La Torre y Jiménez
  - Dr. Elard Arturo Castor Cáceres
  - Dr. Juan Carlos Nicho Virú



**Dr. OSCAR PAMO REYNA**

*Director del FEC*

*Médico internista, Hospital Nacional Arzobispo Loayza*

*Universidad Peruana Cayetano Heredia*

*Académico de número, Academia Nacional de Medicina*

**Dr. JORGE GONZÁLEZ MENDOZA**

*Secretario del FEC*

*Médico infectólogo*

*Moderador, Programa para la Monitorización de Enfermedades*

*Emergentes (ProMED) Sociedad Internacional de Enfermedades Infecciosas*

**Dr. CIRO MAGUIÑA VARGAS**

*Miembro del FEC*

*Médico infectólogo tropicalista dermatólogo*

*Univesidad Peruana Cayetano Heredia*

*Académico de número, Academia Nacional de Medicina*

**Dr. RICARDO IVÁN ÁLVAREZ CARRASCO**

*Miembro del FEC*

*Instituto Nacional Materno Perinatal*

*Asociación Médica Peruana de Patología Clínica*

*Academia Panamericana de Historia de la Medicina*

*Asociación de Historia de la Medicina Peruana*

**Dr. ALBERTO EMILIO ZOLEZZI FRANCIS**

*Director del Comité Editorial de la Revista Acta Médica Peruana*

*Miembro del FEC*

*Médico gastroenterólogo*

*Hospital Nacional María Auxiliadora*

*Univesidad Ricardo Palma*

**Dr. HORACIO VARGAS MURGA**

*Miembro del FEC*

*Médico psiquiatra*

*Docente principal de la UPCH*

*Instituto Nacional de Salud Mental Honorio Delgado - Hideyo Noguchi*

## Presentación del CMP

*El Fondo Editorial Comunicacional del CMP (FEC) tiene la satisfacción de presentar esta importante publicación de los doctores Jorge Manuel Tarrillo-Purisaca y P. Martín Padilla-Machaca y la licenciada Elsi Bazán-Rodríguez.*

*Esta valiosa obra, titulada Bioestadística en la investigación clínica: de la teoría a la práctica, cuenta con el auspicio editorial y el financiamiento del FEC y el CMP.*

*El presente libro permite que los médicos en general se familiaricen con la teoría y la práctica de la bioestadística. Este es el gran aporte de este texto.*

*El CMP se siente complacido de publicar a través del FEC esta obra, ya que no solo revela un gran esfuerzo del autor por publicarlo, sino que se constituye en una fuente de consulta básica sobre la bioestadística y su entorno necesario de conocer por los miembros de la orden médica y los estudiantes de pregrado.*

Raúl Urquizo Aréstegui  
Decano nacional  
Colegio Médico del Perú

Oscar Pamo Reyna  
Director  
Fondo Editorial Comunicacional

*A nuestros estudiantes de  
pregrado y posgrado  
en Ciencias de la Salud*

## **Agradecimiento**

*Al Fondo Editorial Comunicacional del Colegio Médico del Perú (FEC-CMP), por su valioso apoyo en la difusión de esta obra.*

*Al equipo revisor por pares del FEC-CMP.*

*A la doctora Liliana Gonzales Hamada, profesora del curso Proyecto de Investigación de la maestría de la Universidad de San Martín de Porres, por la revisión del manuscrito y sugerencias que contribuyeron a mejorar nuestra obra.*



# Contenido

<b>Prólogo</b> .....	13
<i>Dr. Pedro Miguel Palacios Celi</i>	
<b>Prefacio</b> .....	15
<b>Análisis exploratorio univariado de datos</b> .....	17
CAPÍTULO 1. Conceptos básicos .....	19
CAPÍTULO 2. Las variables de investigación .....	23
CAPÍTULO 3. Elaboración de la base de datos .....	29
CAPÍTULO 4. Distribución normal.....	33
CAPÍTULO 5. Análisis de variables cuantitativas.....	37
CAPÍTULO 6. Análisis de variables cualitativas .....	55
CAPÍTULO 7. Intervalos de confianza .....	63
<b>Anexos</b> .....	69
ANEXO 1. Diseño de base de datos.....	71
ANEXO 2. Análisis de variable cuantitativa .....	89
ANEXO 3. Análisis de variable ordinal.....	99
ANEXO 4. Análisis de variable nominal .....	107
ANEXO 5. Crear una nueva variable: comando calcular variable .....	113
ANEXO 6. Recodificación de una variable .....	117
ANEXO 7. Selección de casos .....	125
ANEXO 8. Comando explorar .....	129
ANEXO 9. Base de datos ficticia de cirrosis .....	135
ANEXO 10. Base de datos Cáncer.....	139
ANEXO 11. Base de datos IMC.....	141
<b>Bibliografía</b> .....	143



## Prólogo

*El Perú, a pesar de algunos avances, sigue teniendo una brecha importante en cuanto a la investigación y la publicación en cantidad y calidad. Existen múltiples causas de este déficit, una de ellas es la falta de preparación de los investigadores, especialmente en el área de la bioestadística. Esta es una de las ramas de la estadística que se encarga de la recolección, el análisis y la presentación de datos, que con el avance de la ciencia y la tecnología se ha convertido en una herramienta esencial para la investigación clínica.*

*En este sentido, el doctor P. Martín Padilla, acompañado de la experticia del doctor Jorge Manuel Tarrillo Purisaca, destacado epidemiólogo y de la maestra Elsi Bazán Rodríguez, licenciada en Estadística en la gestión de servicios de salud, nos presentan el libro Bioestadística en la investigación clínica: de la teoría a la práctica, que llega a nosotros con el propósito de ayudar a los jóvenes investigadores a comprender el uso de la bioestadística en la investigación clínica. El doctor Padilla cuenta con una experiencia de más de 20 años dedicadas a la enfermedades del hígado y trasplantes, en el Hospital Guillermo Almenara de EsSalud, es docente Renacyt de la UNMSM, con maestría en la USMP.*

*El contenido de este libro a lo largo de sus 144 páginas consta de 7 capítulos y 11 anexos, que incluyen los conceptos básicos de la bioestadística, desde su teoría hasta su aplicación práctica; desde la selección de variables de investigación hasta la elaboración de bases de datos. Así también, aspectos más complejos, como el análisis de variables cuantitativas, cualitativas, y los intervalos de confianza. En los anexos también se incluyen bases de datos ficticias para que los lectores puedan poner en práctica los conceptos de bioestadística.*

*Este libro constituye una valiosa herramienta de apoyo para la educación y el entrenamiento permanente de los jóvenes investigadores, estudiantes y profesionales de la salud, con el fin de mejorar la calidad de la investigación clínica en el Perú.*

*Estamos seguros de que este libro ayudará a los lectores a comprender los principios básicos de la bioestadística, desde la teoría hasta la práctica, a desarrollar las habilidades necesarias que les permita aplicar de forma segura y estructurada en su propia investigación clínica, logrando una investigación científica de calidad.*

Dr. Pedro Miguel Palacios Celi

Past decano nacional del Colegio Médico del Perú



## Prefacio

*Cuando realizamos una investigación como profesionales de la salud, nos viene el recuerdo de nuestra tesis de grado de bachiller, del título universitario y, con el tiempo, los grados de maestría o doctorado: ¡Qué parte más complicada la de los resultados y la estadística! Normalmente entregamos al profesional de estadística para que realice todos los análisis de acuerdo a los objetivos de nuestra investigación, recibimos como productos una serie de cuadros, gráficas y resultados de pruebas estadísticas, que a criterio del profesional de estadística se ajustan a nuestros datos.*

*A veces pensamos que la estadística es un terreno exclusivo de los estadísticos, mantener este pensamiento significa seguir restando las posibilidades del profesional de salud en el desarrollo de su curiosidad exploratoria con los datos y resultados de sus investigaciones. Consecuentemente, se pierde la oportunidad de descubrir el comportamiento y las relaciones de las variables que inicialmente no se habrían sospechado ni haberse planteado en los estudios de investigación. Por consiguiente, es conveniente que el profesional de salud desarrolle competencias para utilizar adecuadamente la estadística en el proceso de la investigación. No significa remplazar el aporte del profesional en estadística, sino tener las competencias para entender el lenguaje estadístico y, así, poder explorar minuciosamente las variables de estudio.*

*Visto como un enfoque integral a la investigación en salud, es necesario que esta sea asumida como proceso, los hallazgos de una investigación deben llevar consecuentemente a nuevas investigaciones, de todo esto resulta una espiral ascendente en la generación de conocimiento. En este proceso, es importante incorporar la visión del estadístico, del metodólogo y, por supuesto, del epidemiólogo, para lograr un adecuado nivel de rigor científico. Enfocar la investigación como proceso y resultado de la intervención de un equipo multidisciplinario, contribuirá a elevar el nivel científico de nuestros alumnos de pregrado y posgrado.*

*Iniciarse en la investigación conlleva a adquirir una adecuada comprensión de la estadística, conocer las metodologías de investigación y su correcta utilización, así como lograr la interiorización del enfoque y las técnicas epidemiológicas. Una vez que se han adquirido estas competencias, podrá integrarse eficazmente a cualquier equipo de investigación en el campo de la salud.*

*Es importante, con respecto a la bioestadística, señalar que permite lo siguiente:*

- 1. Aprender a gestionar de modo correcto una base de datos.*
- 2. Realizar una exploración completa de las variables del estudio o de la base de datos.*
- 3. Realizar, en la exploración de datos, los tres niveles de análisis: univariado, bivariado y multivariado.*

*Aquí es donde descubrimos el objetivo de este libro: presentar las herramientas básicas en el diseño de una base de datos y la realización de una exploración de datos, tanto de variables cualitativas como cuantitativas. Este primer objetivo es el más importante. Por esto, expondremos los conceptos básicos, la aplicación tipo para fijar los conceptos y, para efectos prácticos, la utilización del paquete estadístico SPSS, el cual estamos familiarizados la mayoría de nosotros.*

*Otros de los objetivos de nuestra obra es el autoaprendizaje, en este sentido le sugerimos que complete el ejercicio realizando el análisis del resto de variables; si desea ir más allá en su entrenamiento, podría realizar una selección de resultados que deberían ir en supuestos informes de investigación y un listado de hipótesis generado a partir del análisis del comportamiento de las variables de esas bases con datos ficticios.*

*En esta obra, no presentaremos cálculos estadísticos tediosos, por lo que centraremos nuestra atención en aprender a utilizar los conceptos de la estadística y la interpretación de los resultados que nos ofrecen los paquetes estadísticos, en particular el SPSS. Para una mayor profundización matemática o estadística, recomendamos autores en las referencias bibliográficas incluidas al final de este libro.*

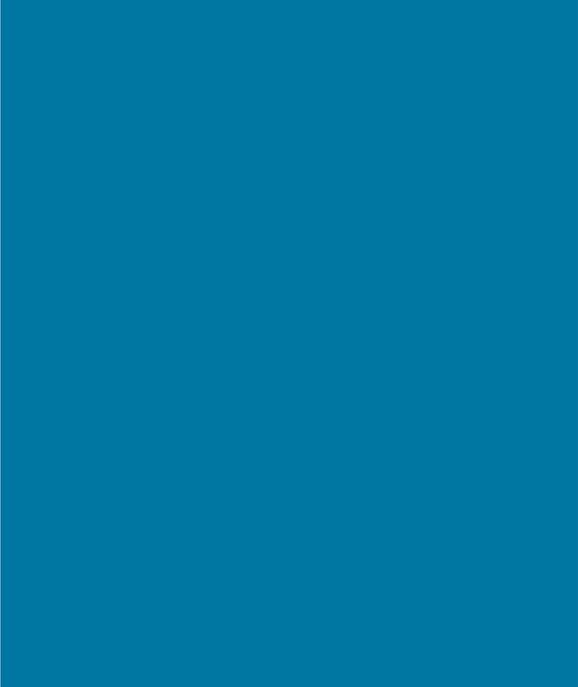
*El análisis de las variables de estudio, mediante la utilización de paquetes estadísticos, conlleva a la generación de una serie de cuadros y gráficos que no necesariamente se incorporarán a los informes de la investigación; pero, sin embargo, son insumos para la reflexión y el análisis del investigador. Esta es la oportunidad del investigador con las competencias estadísticas para detectar comportamientos peculiares y relaciones especiales entre las variables que merecerían un tratamiento diferenciado, para poder confirmar o descartar nuestras hipótesis de trabajo.*

*Acompañamos a este libro tres bases de datos ficticias. Una de ellas presenta pacientes con cirrosis hepática, la segunda a pacientes con cáncer y la tercera (IMC) tiene personas con peso y talla, con ellas realizamos los ejercicios tipo del manual.*

*Es muy importante para la consolidación de las competencias que realice el análisis de variables, ahora construyendo su propia base de datos y siguiendo las recomendaciones del manual. De tal manera que pueda generar con sus resultados el informe de una investigación y un listado de nuevas investigaciones motivadas por el análisis que ha realizado de las variables de su base de datos.*

*Esperamos que este libro pueda lograr su objetivo central: contribuir en la adquisición de competencias para realizar un correcto análisis exploratorio univariado de datos. Esto se pondrá en evidencia si usted, a partir de una base de datos, genera e interpreta todas las medidas de resumen y las gráficas de las variables cualitativas, ordinales y cuantitativas; y, en el caso de las variables cuantitativas, realiza también la evaluación de la normalidad. Asimismo, esperamos que este libro le sea de mucha utilidad en el maravilloso y largo camino de la investigación clínica aplicada.*

Los autores



*ANÁLISIS EXPLORATORIO  
UNIVARIADO DE DATOS*





# Conceptos básicos

# 1

Es necesario, brevemente, puntualizar algunos conceptos sobre términos que serán utilizados frecuentemente.

## ESTADÍSTICA

Genéricamente, la estadística es una ciencia formal orientada a la recolección, organización, procesamiento, análisis e interpretación de datos. Estos datos son de variables o características procedentes de una población o muestra, los cuales, al ser procesados y resumidos, generan información que sirve para describir el comportamiento de las variables en estudio y/o explicar relaciones entre esas variables. Cuando esta información procede de una muestra se puede realizar inferencias hacia la población de donde se extrajo la muestra estudiada, mediante la aplicación de pruebas estadísticas. Finalmente, esta información permite tomar decisiones o incrementar el conocimiento acerca de la realidad de donde se extrajo los datos.<sup>1-3</sup>

## POBLACIÓN

En estadística, una población es el conjunto finito o infinito de sujetos o individuos de la que se obtiene la muestra o participantes en un estudio y a la que se quiere extrapolar los resultados de ese estudio. La población está definida por el interés particular de la investigación, se ubica en un lugar y un tiempo específicos.<sup>1,2</sup>

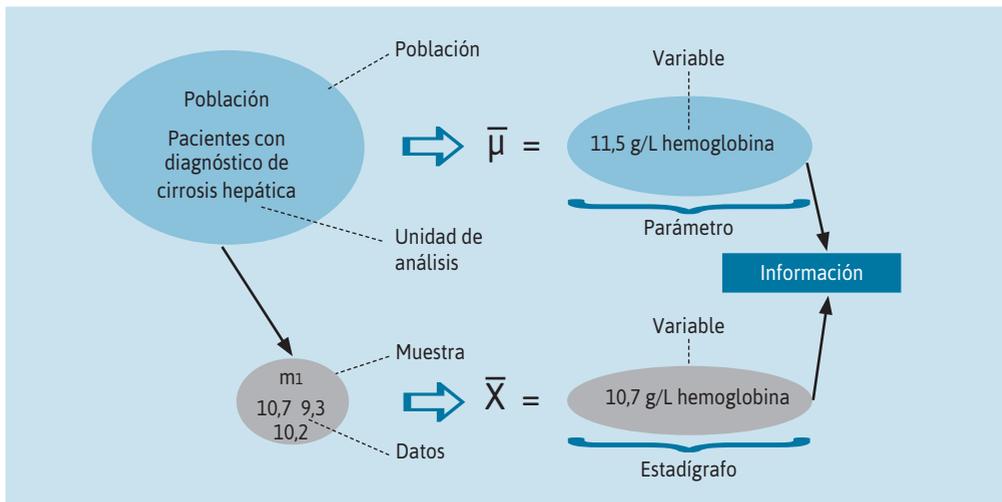


FIGURA 1.1. Esquema de un hipotético caso de estudio de hemoglobina en pacientes con cirrosis hepática.

**POR EJEMPLO**, en la Figura 1.1, se ha esquematizado un hipotético caso de estudio de hemoglobina en pacientes con diagnóstico de cirrosis hepática. En este ejemplo, la población de interés está constituida por todos los pacientes con diagnóstico de cirrosis hepática que acuden al hospital BN en un tiempo determinado.

### **MUESTRA**

Parte de la población seleccionada mediante algún procedimiento de muestreo. Para elegir el tipo de muestreo, se debe considerar el objetivo que se plantee en el estudio.<sup>1,2</sup>

### **UNIDAD DE ANÁLISIS**

Es el objeto o elemento indivisible del cual se obtiene la información que requiere el estudio. En el caso del ejemplo anterior, la unidad de análisis es cada uno de los pacientes con diagnóstico de cirrosis del hospital en referencia.<sup>1,2</sup>

### **VARIABLE**

Es la característica de estudio y que tal como se observa, toma diferentes valores entre las unidades de análisis. En el siguiente capítulo, se explicará en extenso sobre las variables. En el ejemplo presentado, la variable de estudio es la hemoglobina.<sup>1,2</sup>

### **DATO**

El dato es el valor que adquiere la variable en las unidades de estudio. El tipo de valor que adquiere depende de la escala en la que se mide la variable y del instrumento que se utiliza para obtener esta medida. En el caso de la hemoglobina serán cada uno de los valores en mg de hemoglobina por mL de sangre que tiene cada paciente con cirrosis, incluidos en el estudio.<sup>1,4</sup>

### **FUENTE DE DATOS**

Las fuentes de donde se obtienen los datos pueden ser primarias o secundarias. En las fuentes primarias, los datos son obtenidos directamente por el investigador en el estudio o investigación; en el caso de las fuentes secundarias, el investigador obtiene los datos de fuentes externas a la investigación o registrados de modo rutinario.<sup>1,4</sup>

### **ESTADÍGRAFO**

Es la medida de resumen del conjunto de datos de una muestra. En el caso del ejemplo, se tiene el promedio o media de hemoglobina en los pacientes con cirrosis hepática considerados en la muestra, 10,7 mg/mL. El tipo de estadígrafo depende del tipo de variable y del objetivo de estudio. Estadísticamente, se puede inferir el valor del estadígrafo a la población de donde se extrajo la muestra, para esto, se debe cumplir los supuestos de la estadística inferencial.<sup>5</sup>

### **PARÁMETRO**

Es la medida de resumen del conjunto de datos de la población. En el ejemplo, la media de hemoglobina en los pacientes con cirrosis hepática del hospital en referencia es de 11,5 mg/mL.<sup>5</sup>

## INFORMACIÓN

En la perspectiva estadística, información es el resultado de los datos que han sido procesados, analizados e interpretados mediante técnicas estadísticas. La variedad de información depende del tipo de dato, del objetivo de estudio y de los instrumentos utilizados en la obtención de datos.<sup>6</sup>

## BIOESTADÍSTICA

Es la aplicación de los conceptos y herramientas de la estadística al estudio de los datos que proceden de las ciencias biológicas o médicas.

## EXPLORACIÓN DE DATOS

Los datos que se obtienen de una investigación o estudio se recogen en una base de datos para luego aplicar las diversas pruebas estadísticas y obtener los resultados que se esperan. Antes de aplicar cualquier técnica o prueba estadística es conveniente examinar los datos, con el propósito de poder entender: El comportamiento de cada variable a partir de las medidas de resumen y la relación entre las diversas variables en estudio. Esta etapa es conocida como análisis exploratorio de datos (AED).

En el AED se realiza una revisión de la base de datos con el propósito de detectar errores en el diseño y recojo de los datos, analizar los datos perdidos o ausentes y darle un tratamiento especial para no fomentar errores o sesgos en los resultados, detectar los datos atípicos o extremos y tomar una determinación racional en su manejo y, finalmente, poder realizar la comprobación de los supuestos que exigen las diversas pruebas estadísticas como son la normalidad, la linealidad y la homocedasticidad.

Realizar un cuidadoso análisis exploratorio de datos, permitirá detectar no solo errores, sino también indicios de comportamientos o relaciones de las variables no previstas o que sería necesario explorar en estudios diseñados especialmente para ese propósito. Visto de este modo, el AED es una fuente constante de interrogantes que pueden llevar a mejorar o replantear estrategias en la investigación.

Desde la perspectiva de este libro, se presentan los diversos conceptos y herramientas estadísticas para realizar el AED considerando las variables individualmente; concretamente, realizar un análisis univariado.



En el instrumento de recojo de información del estudio, se puede observar una serie de datos que se tienen que registrar por cada paciente. Estos datos corresponden a variables que se intentan conocer y que tienen relación con el objetivo del estudio.

## Concepto de variable

Una variable se la puede definir como una propiedad, característica o atributo identificado en un sujeto u objeto de estudio, que puede variar y que es susceptible de medición. La medición se puede realizar a toda la población o a una muestra extraída de la población.

### CARACTERÍSTICAS DE LAS VARIABLES

Las variables tienen tres características, exhaustiva, excluyente y discriminante.<sup>1,7,5</sup>

- **PROPIEDAD EXHAUSTIVA.** Una variable contiene todas alternativas posibles de su valor, de tal modo que se incluya todas las respuestas posibles de la población.
- **PROPIEDAD EXCLUYENTE.** Los valores o categorías de las variables son excluyentes entre sí.
- **PROPIEDAD DISCRIMINANTE.** La variable tiene categorías concretas, por lo que no existe la posibilidad de que las categorías estén contenidas una en otra.

## Tipos de variables

Las variables se las pueden clasificar de acuerdo a su naturaleza, por la función que cumplen en el análisis y por su estructura.

### SEGÚN SU NATURALEZA

- **VARIABLES CUALITATIVAS**

Estas variables solo consideran la ausencia o presencia de una característica, sin tener valor numérico alguno.<sup>1,2,5</sup>

**POR EJEMPLO,** la variable sexo, refiere la presencia del sexo masculino o femenino en cada sujeto que se estudia.

- **VARIABLES CUANTITATIVAS**

La variable es cuantitativa cuando se puede medir o contar su cantidad en valores numéricos. Estos valores pueden ser continuos o discretos. Un valor es continuo cuando sus cifras se pueden fraccionar en valores infinitos entre dos intervalos; en el caso del valor discreto no

es posible fraccionarlo entre dos intervalos. Una variable que admite medición tiene valores continuos (la talla o el peso), mientras que la variable que se cuenta, tiene valores discretos (el número de hijos, el número de síntomas, etc.).<sup>1,5</sup>

## SEGÚN EL NIVEL DE MEDICIÓN O SU ESCALA DE MEDIDA

### ● VARIABLE NOMINAL

La escala nominal hace referencia a que las categorías de la variable son independientes y se miden tomando como referencia la presencia o ausencia de la característica sin que exista una implicancia jerárquica. Puede ser dicotómica (cuando solo tienen dos valores o características) o politómica (cuando tienen más de dos opciones, valores o características).

**POR EJEMPLO**, una variable nominal dicotómica es la variable sexo, con dos opciones, masculino y femenino.

**POR EJEMPLO**, una variable nominal politómica es el tipo de seguro, que puede ser SIS, EsSalud, privado, sin seguro.

Las operaciones que se pueden aplicar son = (igual) o  $\neq$  (diferente).<sup>1,5</sup>

### ● VARIABLE ORDINAL

En la escala ordinal, además de considerar la presencia o ausencia de la característica de estudio, posee grados que pueden relacionarse de menor a mayor o viceversa. Este ordenamiento obedece al arbitrio o convencionalismo basado en la experiencia o el conocimiento aceptado en el área en que se realiza el estudio.

**POR EJEMPLO**, la variable estado nutricional, que puede ser bajo peso, peso normal, sobrepeso y obesidad.

Se pueden aplicar las operaciones = (igual),  $\neq$  (diferente), < (menor que), > (mayor que).<sup>1,5</sup>

### ● INTERVALO

La escala de intervalo incluye las características de las anteriores, pero sí se puede medir las distancias entre categorías. Sin embargo, en este tipo de escala, el punto cero y la unidad de medición son arbitrarios, porque adquieren sentido solo cuando son comparados con otros valores que están en la categoría. Esta escala no tiene el cero absoluto.

**POR EJEMPLO**, las variables temperatura y frecuencia cardíaca.

Las operaciones que se pueden aplicar son = (igual),  $\neq$  (diferente), < (menor que), > (mayor que) + (suma), - (resta).<sup>1,5</sup>

### ● RAZÓN

La escala de razón posee un cero absoluto y sus mediciones son realizadas en un sistema isométrico.

**POR EJEMPLO**, las variables peso y talla.

Las operaciones que se pueden aplicar son = (igual),  $\neq$  (diferente), < (menor que), > (mayor que) + (suma), - (resta),  $\times$  (multiplicación) y  $\div$  (división).<sup>1,5</sup>

### SEGÚN LA FUNCIÓN QUE CUMPLEN EN EL ANÁLISIS

- **VARIABLE INDEPENDIENTE O CAUSAL**

Aquella que influye en la variable dependiente. Llamada también de exposición.

- **VARIABLE DEPENDIENTE, DE EFECTO O RESULTADO**

Aquella que depende o es el efecto de la variable independiente. Muchas veces se refiere a esta variable como la de interés, de criterio o principal de la investigación.

- **VARIABLES INTERVINIENTES (VARIABLES CONFUSORAS O DE INTERACCIÓN)**

Son las variables relacionadas con el perfil de los sujetos en estudio.

En los estudios en los que se evalúa la relación entre una variable independiente y otra dependiente, se pueden encontrar variables que pueden enmascarar o modificar el efecto de la relación. Son las variables o factores de confusión, que son variables externas a la relación que se estudia.

Estas variables se caracterizan porque en el tiempo se presentan en el sujeto con anterioridad a la variable independiente (exposición) y, además, están relacionadas con ambas variables, independiente y dependiente. Las variables de confusión producen sesgo en la relación que se estudia, por lo que requieren un tratamiento especial.<sup>8</sup>

Las variables de interacción no tienen relación con las variables de estudio, pero modifican la intensidad de la relación entre estas variables.

El manejo de estas variables depende del criterio del investigador y el objetivo de la investigación. Se pueden eliminar en la fase de diseño del estudio, mediante la restricción, el emparejamiento o la asignación aleatoria (randomización). Si se desea evaluar el rol de estas variables, entonces no se eliminan. En todo caso, en el momento del análisis, se procede a realizar un análisis estratificado y multivariado.<sup>8</sup>

### SEGÚN SU ESTRUCTURA

- **VARIABLES SIMPLES**

Aquellas variables que tienen un solo indicador o unidad de medida.

**POR EJEMPLO**, las variables sexo y edad.

- **VARIABLES COMPLEJAS**

Variables que se constituyen en dimensiones y poseen indicadores para las respectivas dimensiones.

**POR EJEMPLO**, cuando se estudian actitudes o conductas, estas variables tienen varias dimensiones; en ocasiones, se constituyen variables complejas cuando se evalúan el perfil clínico, el perfil sociodemográfico, etc.

**Tabla 2.1.** FORMATO DE OPERACIONALIZACIÓN DE LAS VARIABLES

Variable	Tipo-indicadores	Escala de medición	Valores de medición
• Cirrosis hepática	Variable categórica: Dicotómica Presente/ausente	Nominal	N.º, %, prevalencia
• Edad	Variable numérica: años exactos	De razón	x, DE

Fuente: Formato de operacionalización de variables de estudio ficticio.

## Operacionalización de las variables

Es el procedimiento a través del cual las variables se analizan y se determinan las formas de realizar su medición. En un estudio, con respecto a las variables, se recomienda realizar las siguientes etapas: identificar las variables del estudio; conceptualizar las variables teóricas; operacionalizar las variables empíricas; elaborar o determinar las escalas de medición.

Todas estas variables, antes de ser plasmadas en la ficha o instrumento de recojo de datos, han tenido que ser conceptualizadas y operacionalizadas. Es decir, se ha definido cada una de las variables para que puedan ser medibles y manejables.

Existen varias formas o formatos de presentación de la operacionalización de variables. En la Tabla 2.1, se tiene una presentación muy utilizada.

En la columna ‘Variable’, se asigna el nombre de cada una de las variables determinadas por las diversas dimensiones, componentes o factores que se han considerado en la conceptualización.

En la columna ‘Tipo-indicadores’, se debe indicar el tipo de variable, su indicador o indicadores y sus diversos niveles, categorías o valores posibles.

En la columna ‘Escala de medición’, se especifica en qué escala se medirá la variable. Es importante esta información, porque, de acuerdo con la escala, se podrá realizar determinados análisis estadísticos a nivel univariado, bivariado o multivariado.

En la columna valores de medición se deben indicar las medidas de resumen para cada variable consideradas en el estudio.

## Análisis estadístico de las variables

En el Tabla 2.2, se pueden visualizar todas las medidas de resumen de las variables de acuerdo con su tipo. A este análisis variable por variable se le denomina análisis univariado. Todas estas medidas serán descritas en los siguientes capítulos.

TABLA 2.2. ANÁLISIS UNIVARIADO DE LAS VARIABLES

Análisis	Nominal	Ordinal	Intervalo / razón
<ul style="list-style-type: none"> <li>Medidas de resumen</li> </ul>	Distribución de frecuencias Moda	Distribución de frecuencias Moda Mínimo Máximo Mediana Cuartiles Percentiles Rango intercuartílico	Distribución de frecuencias Moda Mediana Media Mínimo Máximo Rango Varianza Desviación típica Cuartiles Rango intercuartílico Percentiles Coeficiente de variación Asimetría Curtosis
<ul style="list-style-type: none"> <li>Gráficos</li> </ul>	Diagrama de sectores o de barras	Diagrama de sectores o de barras Gráfico de caja y bigotes	Histograma Gráfico de caja y bigotes Gráfico de tallo y hojas

Fuente: Resumen elaboración propia de la exposición del análisis univariado de las variables



## Diseño de base de datos

Se recomienda que, posteriormente a la operacionalización de las variables y antes de aplicar el instrumento que permite obtener los datos, se diseñe la base de datos en un programa informático para tal fin. Cualquier programa que se utilice puede ser adecuado, porque los paquetes o programas estadísticos que se usan en la actualidad para realizar los análisis permiten la captura de las bases de datos sin mucha dificultad, sea cual sea el programa en que originalmente se construyó. Lo importante es tener en cuenta recomendaciones genéricas que hacen más fácil el traslado de esta información hacia estos programas estadísticos.<sup>1,9</sup>

## Recomendaciones para la elaboración de la base de datos para SPSS

### LISTADO DE TODAS LAS VARIABLES DE ESTUDIO CON ESPECIFICACIÓN DE LAS CARACTERÍSTICAS A TENER EN CUENTA EN EL DISEÑO DE LA BASE

Una recomendación importante es que se construya un listado de todas las variables de estudio con especificación de las características que se deben tener en cuenta en el diseño de la base (Tabla 3.1).<sup>5,10</sup> Estas características son número, nombre, tipo, etiqueta, valores y escala.

- **NÚMERO**

Colocar el número correlativo de los casos (objetos, sujetos de estudio) que se van agregando en el listado.

**TABLA 3.1.** LISTADO DE VARIABLES Y SUS CARACTERÍSTICAS, PARA EL INGRESO DE DATOS EN SPSS

Número	Nombre	Tipo	Etiqueta	Valores	Escala
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....

Fuente: Diseño construido a partir de los requerimientos de SPSS en la creación de variables

- **NOMBRE**

A cada una de las variables se le debe asignar un nombre corto y en minúscula. Por ejemplo, ‘sex’, para sexo; ‘dx’, para diagnóstico; ‘etiolog’, para etiología, etc. Se recomienda básicamente que el nombre no sea muy largo, sea entendible y, en la medida de posible, relacionado con el nombre completo de la variable. También como nombre podría asignarse números correlativos a las variables.

- **TIPO**

Es importante, desde un inicio, tener claro qué tipo de registro tendrá cada variable, para que cuando la variable sea exportada a un *software* estadístico, esta sea reconocida adecuadamente para los análisis respectivos. Se puede elegir el tipo numérico en caso de que la variable tenga registro en número; tipo fecha, en caso de ser un registro relacionado con fechas calendario; tipo cadena, en caso de ser texto; otros tipos como moneda, notación científica, en caso de necesidad, aunque su uso no es frecuente.<sup>10</sup>

- **ETIQUETA**

Se debe colocar el nombre completo de la variable.

- **VALORES**

En esta columna se debe especificar los niveles, categorías o valores de la variable y asignarles un número, para que los paquetes estadísticos puedan reconocer los diversos valores y poder realizar los análisis respectivos sin contratiempos. Por ejemplo, la variable ‘sex’ (sexo) tiene dos categorías, masculino y femenino; se pueden asignar los números 1 y 2, respectivamente, o 0 y 1, si se prefiere).

En caso de una variable politómica como ‘tipseg’ (tipo de seguro), que puede tener los ítems EsSalud, Minsa y privado, se pueden asignar los números 1, 2 y 3, respectivamente, para designar a los tipos de seguro que se pueden consignar.

En caso de una variable ordinal, es necesario respetar el orden o jerarquía en el momento de asignar los números.

**POR EJEMPLO**, en la variable ‘imc’ (índice de masa corporal), se tienen los niveles de bajo peso, normal, sobrepeso y obesidad a los cuales se deben asignar los números 1, 2, 3 y 4, respectivamente; en caso de colocar el orden normal, bajo peso, sobrepeso y obesidad, se está cometiendo un error de orden y, por consiguiente, de acuerdo al orden que se ha asignado, la valoración en números de 1, 2, 3 y 4 no es congruente con el comportamiento jerárquico de la variable ordinal.

En caso de variables numéricas, no se pueden asignar valores, se deben registrar todos los valores tal y conforme se presentan. Solo en caso de que una variable numérica sea transformada en ordinal o nominal, se deben asignar valores según lo anteriormente expresado.<sup>10</sup>

Una observación particular es el caso de las variables nominales, que son referidas como ausencia o presencia de una característica. Se pueden asignar el número 0 para ausencia y el 1 para presencia de la característica en estudio.

**POR EJEMPLO**, a la variable ‘diagnóstico de cirrosis’ que tiene dos categorías, “sin diagnóstico de cirrosis” y “con diagnóstico de cirrosis”, se le pueden asignar 0 y 1, respectivamente.

### ● ESCALA

Se debe asignar el tipo de escala de la variable: nominal, ordinal o escalar, tanto para variables discretas y continuas.<sup>10</sup>

### RECOMENDACIÓN SOBRE ALGUNAS CONSIDERACIONES EN EL MOMENTO DE DISEÑO DE LA BASE DE DATOS

Se pueden usar los aplicativos para procesar texto, hojas de cálculo o de programación, pero en todos ellos se deben transferir, en la medida de lo posible, todas las especificaciones de las características de las variables que se han designado en el paso anterior.<sup>10</sup>

En la Figura 3.1 se presenta un ejemplo de una base diseñada en una hoja de cálculo. El caso 1 (iden), se trata de una persona de sexo masculino (sexo: 1), que procede de Lima Cercado (procedencia: 1), afiliada a EsSalud (tipo de seguro: 1), con sobrepeso por el IMC (índice de masa corporal: 3). En este caso el registro se realiza en valores numéricos y las etiquetas indican la categoría o nivel de cada variable. Esta forma de diseñar una base de datos facilitará su exportación a cualquier *software* estadístico.

Para mayores especificaciones en el diseño de una base en una hoja de cálculo, se puede revisar el Anexo 1.

	A	B	C	D	E	F	G
1	iden	sex	proced	tipseg	imc	Índice de Masa Corporal:	
2	1	1	1	1	3	1 Bajo peso	
3	2	1	7	2	3	2 Normal	
4	3	2	4	2	4	3 Sobrepeso	
5	4	2	3	2	4	4 Obesidad	
6	5	2			1		
7	6	2			1		

**Procedencia:**

- 1 Lima cercado
- 2 Rimac
- 3 Los Olivos
- 4 Carabayllo
- 5 Comas
- 6 San Martín de Porras
- 7 Jesús María
- 8 Lince

**Tipo de seguro:**

- 1 EsSalud
- 2 Minsa
- 3 Privado

FIGURA 3.1. Base de datos y las variables de estudio.

### RESTRICCIONES EN EL INGRESO DE DATOS

Incluir restricciones al ingreso de datos, para evitar valores incorrectos. Ver Anexo 1.

### DATO PERDIDO

Se debe reconocer y dar un valor para los casos en los cuales no existe el dato, llamado también dato perdido. Ninguna celda debe quedar sin valor.

### CREACIÓN DE LAS VARIABLES DE ESTUDIO CON TODAS LAS CARACTERÍSTICAS ASIGNADAS

En caso de usar SPSS para el diseño de una base de datos, inicialmente, se debe crear las variables del estudio considerando todas las características que se han asignado en el listado de las variables. Ver el Anexo 1, para el diseño de una base de datos en SPSS.

### Uso de los paquetes estadísticos

Diseñada la base de datos y registrados todos los casos del estudio, se procede a la realización del análisis univariado, bivariado o multivariado. Para esto, se puede utilizar cualquier *software* estadístico. Por su aplicabilidad, específicamente para el campo de la salud, se recomienda SPSS, STATA, EpiInfo, EpiDat.

Cada programa estadístico tiene su comando para exportar una base de datos diseñada en una hoja de cálculo. Se puede utilizar también los comandos de copiado y pegado después de haber seleccionado las variables o casos que se desean trasladar o exportar.<sup>4,11</sup>

Antes de realizar cualquier análisis estadístico, es conveniente tener muy claro el concepto de la distribución normal, para poder explicar y relacionarla con las distribuciones de frecuencias de las variables cuantitativas que se analizan.<sup>1,9,12</sup>

La importancia de determinar si las variables cuantitativas del estudio siguen una distribución normal es porque permitirá utilizar determinados análisis estadísticos. Así, por ejemplo, si se concluye que la variable cuantitativa de estudio tiene una distribución normal, se utilizarán las pruebas paramétricas para su análisis; en caso contrario, es decir, que no sigue una distribución normal, se utilizarán las pruebas no paramétricas para su análisis respectivo.

## Representación de la distribución normal de los valores de una variable cuantitativa

La distribución normal de los valores de una variable cuantitativa se representan en una gráfica (Figura 4.1). Los valores de esta variable se reparten en valores bajos, medios y altos, los cuales van creando o perfilando una gráfica de forma acampanada y simétrica, llamada curva de Gauss o campana de Gauss. El punto máximo de altura de esta curva indica el promedio o media y tiene a su costado derecho e izquierdo, puntos de inflexión simétricos.

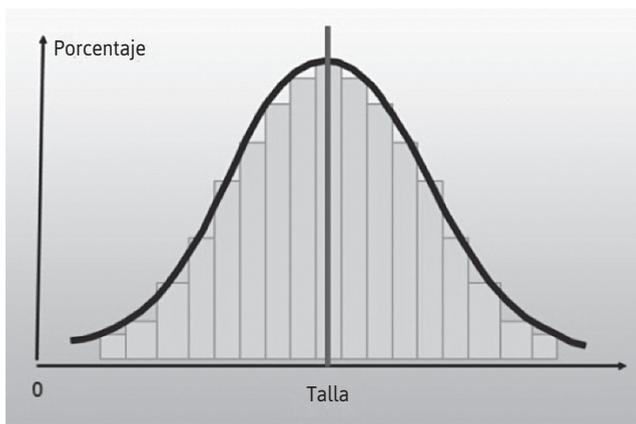
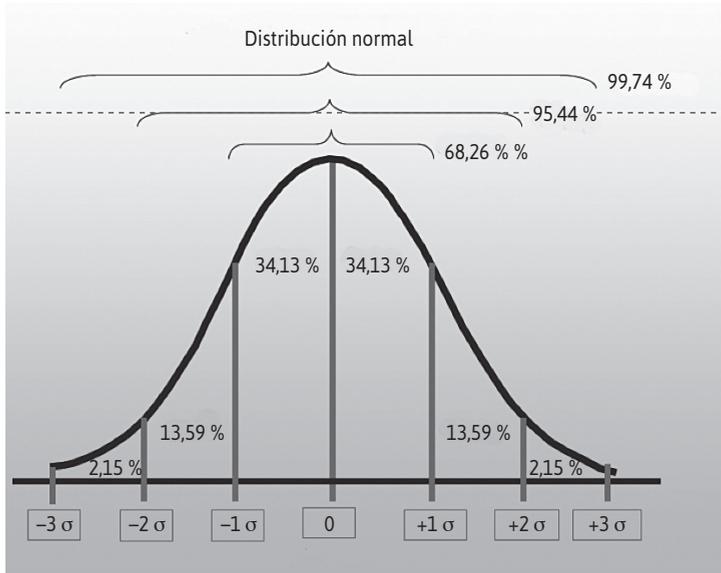


FIGURA 4.1. Gráfica que insinúa una distribución normal.



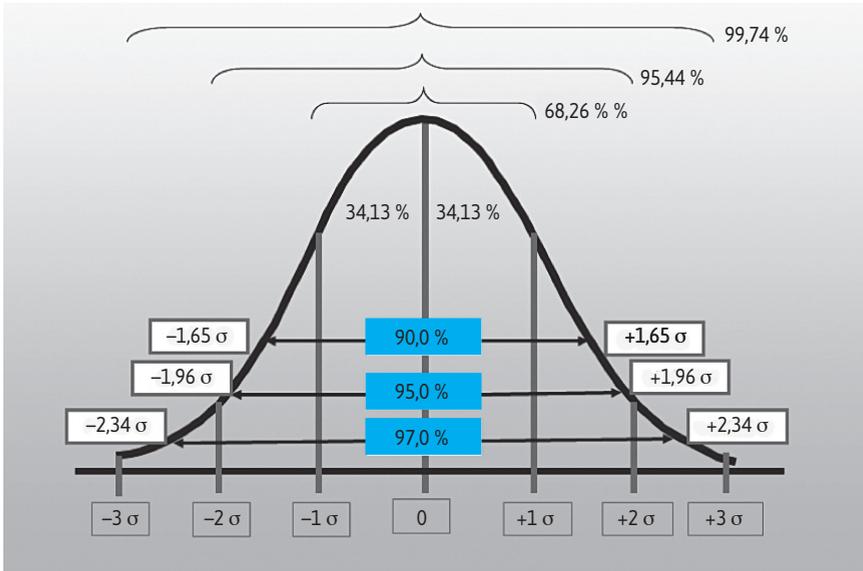
**FIGURA 4.2.** Distribución normal, desviaciones estándar y porcentajes comprendidos entre las tres desviaciones estándar hacia la derecha e izquierda.

Estadísticamente, la distribución normal estándar tiene una media de 0 y una desviación estándar de 1. La curtosis y la asimetría dan información sobre la desviación de una curva con respecto a la normal. En la curva normal, la media coincide en valor con la mediana y la moda.<sup>12</sup>

En la Figura 4.2, se observa cómo, en una curva normal, los valores se desvían de la media tanto a la derecha como a la izquierda en tres desviaciones estándar y el porcentaje de casos que agrupa en cada uno de ellos; de la media hacia la izquierda la primera desviación ( $-1$ ) agrupa a un 34,13 % de los casos, lo mismo sucede a la derecha de la media con la primera desviación ( $+1$ ) que también agrupa al 34,13 %, por lo que el 68,26 % de los casos se agrupan alrededor de la primera desviación estándar de izquierda a derecha. Si se considera la segunda desviación estándar, esta incorpora el 13,59 % de los casos a ambos lados, por lo que, si se considera el total de casos a dos desviaciones estándar tanto de derecha e izquierda, se agrupa un total de 95,44 %. Finalmente, si se consideran las tres desviaciones estándar de derecha e izquierda, se agrupan al 99,74 % de casos de la población en estudio.

Si la curva de la Figura 4.2 fuera la curva de distribución de la talla en sujetos de estudio, se observaría que desde la media hacia la derecha se tiene una talla mayor que la media; lo contrario se podría decir hacia la izquierda, se tiene una talla menor que la media.

Otros porcentajes bajo la curva normal que tienen uso en estadística y sus respectivas desviaciones estándar se observan en la Figura 4.3.



**FIGURA 4.3.** Distribución normal, desviaciones estándar, porcentajes comprendidos entre las tres desviaciones estándar hacia la derecha e izquierda y valores de Z más usados.

Las llaves en la Figura 4.3 señalan los porcentajes bajo la curva normal de acuerdo con las desviaciones estándar. Las flechas horizontales de la misma figura señalan los porcentajes bajo la curva normal que más se utilizan y expresan sus respectivos valores de las desviaciones estándar: al 90 %, al 95 % y al 99 % bajo la curva le corresponden un valor de  $\pm 1,65$ ,  $\pm 1,96$  y  $\pm 2,34$ , respectivamente. A este valor también se le llama valor Z.

#### CARACTERÍSTICAS DE LA CURVA O DISTRIBUCIÓN NORMAL

- **DE DISTRIBUCIÓN ORDENADA.** Todos los casos ordenados de acuerdo con el valor de las medidas de los respectivos casos, lo que origina el agrupamiento de los casos por valores idénticos (frecuencias).
- **ES SIMÉTRICA.** La curva tiene la misma configuración hacia la derecha y la izquierda de la media. La mediana coincide con la media.
- **ES UNIMODAL.** Solo tiene un valor que posee la mayor frecuencia de casos. La moda coincide con la media y la mediana.
- **DE FORMA ACAMPANADA.** El perfil de la curva se insinúa como una campana.
- **ES ASINTÓTICA.** Los valores de los extremos de la curva se extienden indefinidamente en ambas direcciones. La curva no toca el eje de las abscisas (eje horizontal).



Es necesario precisar que cuando se realiza la medición de una variable cuantitativa en una muestra grande o pequeña, se tendrá al final 'n' datos individuales, los cuales van a requerir una suerte de síntesis para poder ser presentados como información del comportamiento de la variable en el grupo de estudio. En ese sentido, la estadística provee una serie de herramientas para ese análisis y que, al final, reportan las denominadas medidas de resumen.

Las medidas de resumen se clasifican en medidas de posición, medidas de dispersión y medidas de forma.<sup>4,2,13</sup>

## Medidas de posición

Estas medidas están referidas a los valores de la variable relacionadas con la posición que ocupan dentro del rango de valores que tiene. Estas medidas se clasifican en medidas de tendencia central y medidas de posición no central.<sup>4</sup>

### MEDIDAS DE TENDENCIA CENTRAL

Son la media, la mediana y la moda. Estas tres medidas ya se vieron en la distribución normal.<sup>14</sup>

- **MEDIA (O PROMEDIO)**

Es el valor que se obtiene cuando se divide la suma de todos los valores de las 'n' observaciones o casos entre el número de estos. De este modo, el valor de la variable se distribuye en partes iguales entre cada una de las observaciones o casos.

Esta medida está en relación con los valores que tiene la variable en cada observación o caso.<sup>4</sup>

- **MEDIANA**

Es el valor de la variable que se ubica en la posición central del conjunto de los valores de las 'n' observaciones o casos. Esta medida permite dividir el conjunto de observaciones en dos partes iguales, un 50 % de las observaciones tienen un valor por debajo de esta medida y el otro 50 % tienen valores por encima de la mediana.<sup>4</sup>

La mediana es el valor de la variable, pero en relación con la población de las observaciones.

- **MODA**

Es el valor que se repite con mayor frecuencia en las 'n' observaciones o casos.

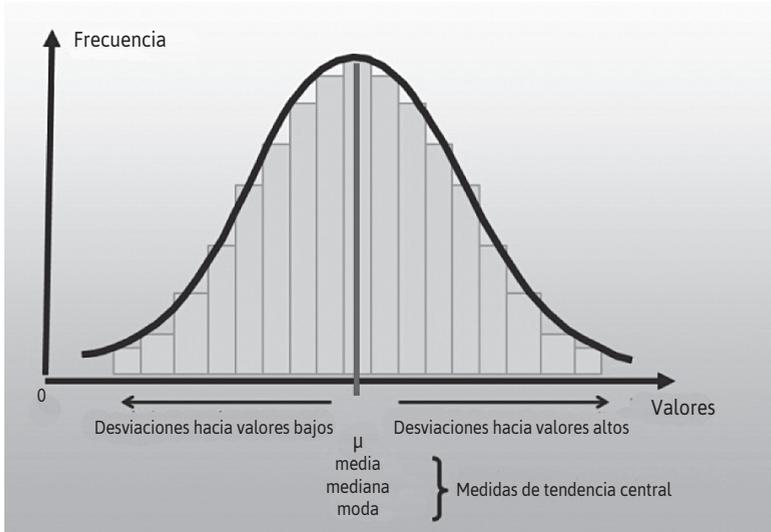


FIGURA 5.1. Curva normal y medidas de tendencia central.

Como ya se sabe, en una distribución normal estas tres medidas coinciden en cuanto a sus valores.<sup>4</sup>

Estas tres medidas, la media, la mediana y la moda, son importantes y cada una de ellas tiene sus limitaciones y su valor en sí misma.

- La media es una medida afectada fuertemente por los valores extremos que puede tener la muestra o población y, por consiguiente, oculta extremos o desigualdades.

**POREJEMPLO,** se informa que la media de dosaje de alfafetoproteína en un grupo de 25 pacientes es de 135,92 ng/mL, si solo se toma en cuenta esta información se podría pensar que se trata de un grupo de pacientes que, además de la cirrosis, tiene cáncer de hígado, sin embargo, cuando se observan los resultados paciente por paciente, se tienen los siguientes valores:

12, 11, 17, 15, 19, 20, 21, 20, 23, 14, 10, 17, 16, 18, 17, 21, 22, 13, 11, 19, 10, 15, 25, 12, 3 000

Entonces, se observa que en realidad se trata de un grupo de pacientes con cirrosis y que solo uno de ellos tiene un valor extremo de 3 000 ng/mL y que, probablemente, sea el único caso con cáncer de hígado.

Cuando se ordena el valor de todos los pacientes, se obtiene lo siguiente:

10, 10, 11, 11, 12, 12, 13, 14, 15, 15, 16, 17, 17, 17, 18, 19, 19, 20, 20, 21, 21, 22, 23, 25, 3 000

En el ejemplo, se observa que la mediana es de 17 ng/mL. La mediana informa que el 50 % de los pacientes tiene un valor inferior a 17 ng/mL y que el otro 50 % restante tiene valores por encima de 17 ng/mL.

En el ejemplo, la moda tiene un valor de 17 ng/mL. Indica que el valor más frecuente en el grupo de pacientes del ejemplo es 17 ng/mL.

Si se elimina el caso con valor extremo (3 000), porque es el único caso con cáncer, los resultados son los siguientes: la media es de 16,58 ng/mL; la mediana, de 17 ng/mL y la moda, de 17 ng/mL.

En este caso, la mediana y la moda no han sido afectadas por los valores extremos.

- Dependiente del objetivo del análisis, la mediana puede ser más informativa que la media.
- En otras circunstancias, la moda tiene importancia en salud, cuando se desea saber cuál es el grupo que se ve más afectado por algún problema de salud o, en sentido positivo, tiene mejores condiciones de salud.

### MEDIDAS DE POSICIÓN NO CENTRAL

#### • CUARTILES

Son los tres valores de la variable (Q1, Q2 y Q3) que dividen al conjunto de observaciones o casos en cuatro partes iguales. Q1, Q2 y Q3 determinan los valores correspondientes al 25 %, al 50 % y al 75 % de los datos (Figura 5.2).

El valor de Q2 coincide con la mediana.

Los cuartiles se pueden representar en un histograma de frecuencias o en la gráfica de caja y bigotes (Figura 5.3).<sup>1</sup>

Los cuartiles representados en una gráfica de curva normal y en una gráfica de caja y bigotes se presentan en la Figura 5.4.

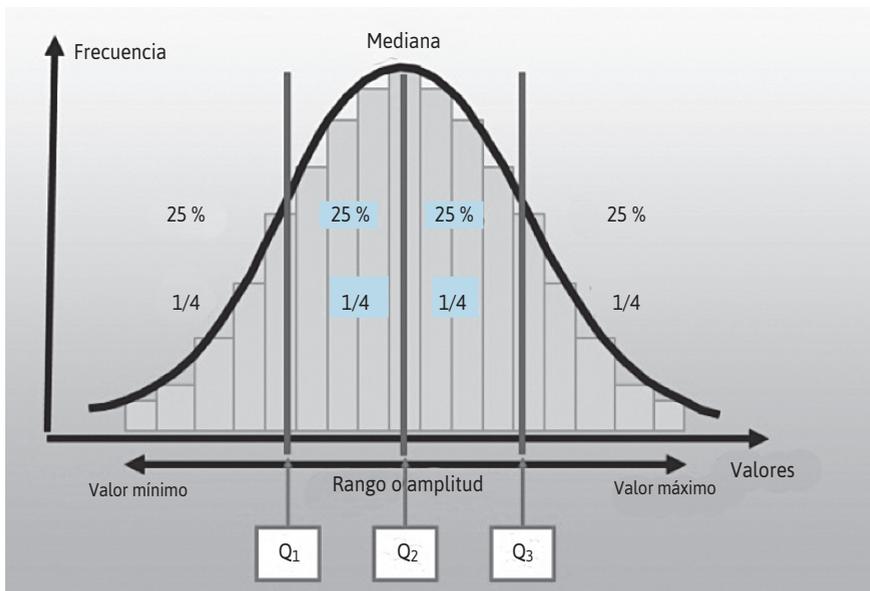


FIGURA 5.2. Ubicación de los cuartiles en la curva normal.

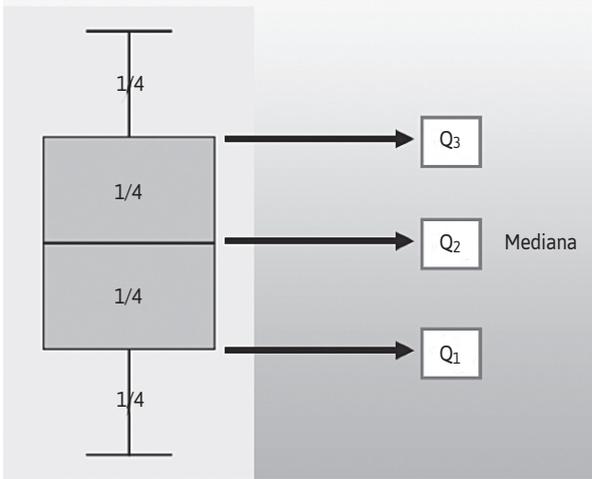


FIGURA 5.3. Medidas de posición no central. Cuartiles en la gráfica de caja y bigotes.

● **DECILES**

Son los nueve valores (D1, D2, D3, ..., D10) que dividen al conjunto de observaciones o casos en diez partes iguales. Dan los valores correspondientes al 10 %, al 20 %, al 30 %... y al 90 % de las observaciones, respectivamente.

El valor de D5 coincide con la mediana.

● **PERCENTILES**

Son los 99 valores (P1, P2, ... P99) que dividen la serie de datos en 100 partes iguales. Dan los valores correspondientes al 1 %, al 2 %... y al 99 % de los datos.

El valor de P50 coincide con la mediana.

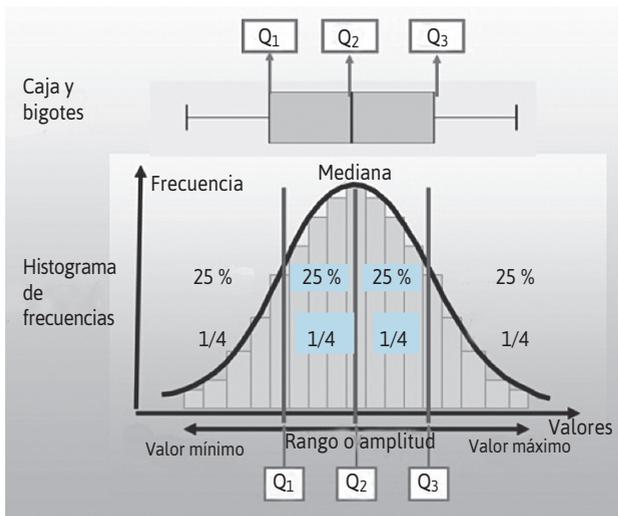


FIGURA 5.4. Medidas de posición no central. Gráfica de la curva normal y la gráfica de caja y bigotes.

**PRESENTACIÓN DE UN EJEMPLO**

Para el análisis de estas medidas de posición, se utilizará una base ficticia sobre cirrosis, con la utilización de dos variables, la edad y el puntaje MELD (siglas del inglés *Model for End Stage Liver Disease*).

El puntaje MELD es un modelo matemático de predicción de la supervivencia de un paciente con cirrosis basado en simples valores de laboratorio rutinarios, bilirrubina, INR y creatinina. El puntaje MELD es más objetivo y preciso que la clasificación de Child-Pugh).<sup>15,16</sup>

Los resultados se presentan en la Tabla 5.1.

- **EDAD.** La media o promedio es de 61,4 años, se trata de una población predominantemente de la tercera edad. La mediana tiene un valor de 61 años, es decir que el 50 % de la población tiene una edad por debajo de 61 años y el otro 50 % tiene una edad por encima de 61 años. La edad más frecuente en esta muestra es 60 años. Las tres medidas no son iguales, pero tienen cierta proximidad.
- **MELD.** La media o promedio tiene una puntuación de 15,45. La mediana tiene un valor de 14 de puntuación, es decir, que el 50 % de la población tiene un puntaje de MELD por debajo de 14 de MELD y el otro 50 % tiene un MELD por encima de 14. El puntaje de MELD más frecuente en esta muestra es 14. La mediana y la moda son iguales, pero están distantes de la media.

En el análisis se ha descrito cada una de estas medidas en cada una de las variables, pero no se han comparado, porque las dimensiones en que han sido medidas no son comparables. No se pueden comparar años con puntaje MELD.<sup>15</sup> Se podrían comparar estas medidas si se tienen los resultados de dos estudios realizados en diferentes poblaciones de pacientes, pero se debe tener en cuenta que el análisis no solo debe considerar estas medidas, sino también las medidas de dispersión y de forma para evaluar si son semejantes o diferentes entre sí.

En la Tabla 5.2, se presentan la información de los cuartiles, deciles y percentiles para las mismas variables de edad y MELD.

- **EDAD**
  - **CUARTILES.** Hasta el 25 % de la población (primer cuartil o Q1) tiene una edad de 53 años; hasta el 50 % de la población, una edad de 61 años o el 50 % de la población tiene una edad por encima de 61 años (segundo cuartil o Q2 o mediana). Hasta el 75 % de la población tiene una edad de 70 años o que un 25 % de la población tiene una edad mayor de 70 años.

**TABLA 5.1.** MEDIDAS DE TENDENCIA CENTRAL PARA LAS VARIABLES: EDAD Y MELD

Medida	Edad	MELD
• Media	61,40	15,45
• Mediana	61,00	14,00
• Moda	60,00	14,00

Fuente: Base de datos ficticia de cirrosis (Anexo 9)

**TABLA 5.2.** CUARTILES, DECILES Y PERCENTILES PARA LAS VARIABLES: EDAD Y MELD

Medida		Edad	MELD
• Cuartiles	Q1	53	11,00
	Q2	61	14,00
	Q3	70	18,00
• Deciles	10	47,10	9,00
	20	52,00	10,00
	30	55,00	11,00
	....	....	....
	80	70,80	19,00
• Percentiles	90	76,00	25,90
	1	30,00	7,00
	5	42,05	8,00
	15	50	10,00
	35	56,35	12,00
	...	...	...
	95	79,95	28,00

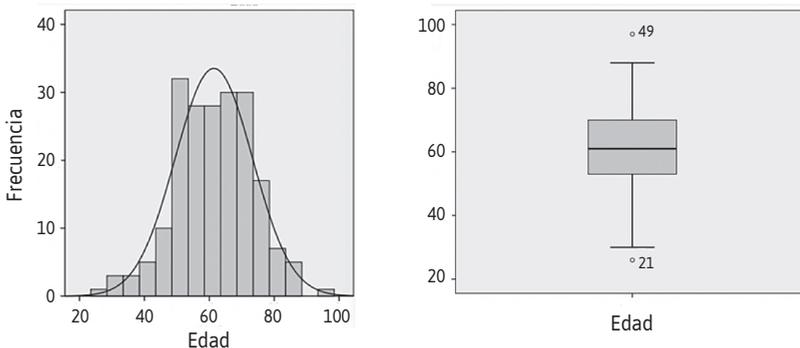
Fuente: Base de datos ficticia de cirrosis (Anexo 9)

- **DECILES.** En este caso, los valores de la edad se dan por grupo de edad que se van integrando en porcentajes de 10 en 10. Por ejemplo, si se interpreta el decil 30, se puede expresar que un 30 % de la población tiene una edad hasta 55 años (o es menor de 56 años) o que el 70 % de la población tiene una edad por encima de los 55 años.
- **PERCENTILES.** Al igual que en los deciles, se puede elegir un punto de corte de la población y expresar el valor de la variable por debajo o por encima de este punto. Por ejemplo, se puede decir que el 15 % de la población tiene una edad de hasta 50 años o que el 85 % de la población tiene una edad por encima de los 50 años.

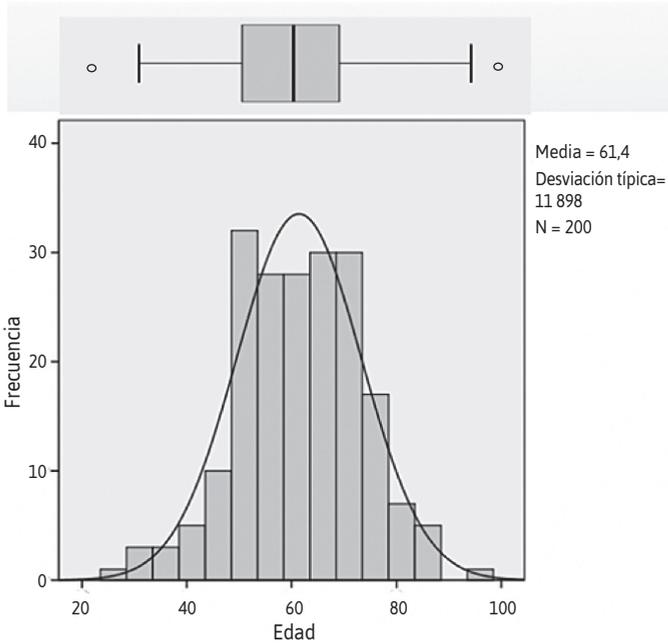
● **MELD**

Una interpretación parecida se puede realizar en función de la variable MELD.

La Figura 5.5 presenta las gráficas generadas por el software estadístico y la Figura 5.6, la gráfica de la relación de ambas gráficas.



**FIGURA 5.5.** Gráficas de la distribución de los datos comparativamente para la variable edad. A) Curva normal. B) Gráfica de caja y bigotes.

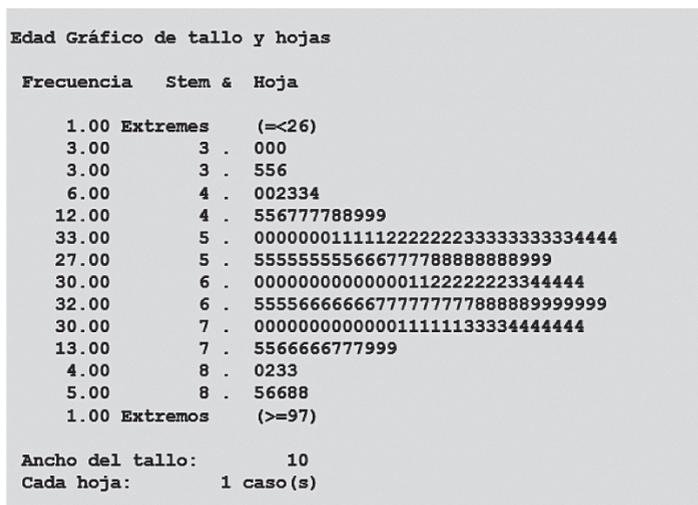


**FIGURA 5.6.** Correlación de la gráfica de la distribución de los datos y la gráfica de caja y bigotes en la variable edad.

En la gráfica de caja y bigotes se observa que algunos valores se comportan como valores extremos, más adelante se explicará este tipo de datos y qué medidas tomar al respecto.

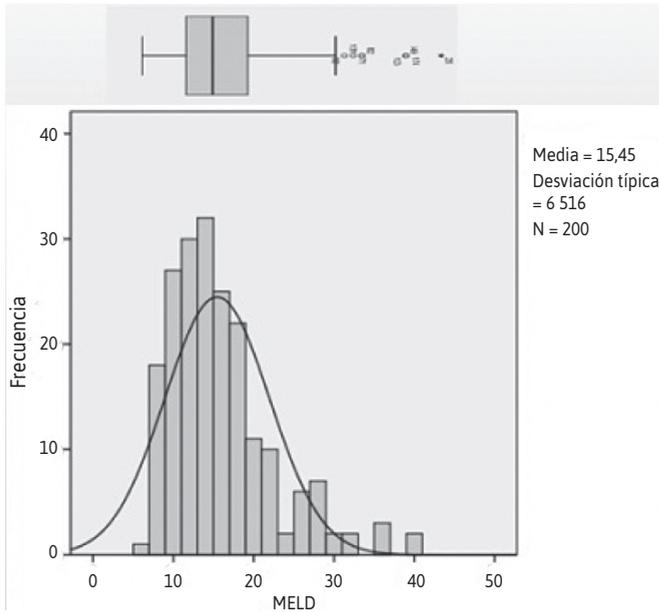
Para visualizar la variable edad, también se podría utilizar la gráfica de tallos y hojas (Figura 5.7).

Esta gráfica tiene tres columnas, frecuencia, stem y hoja. En la columna frecuencia se presenta el número de veces que se repite el valor del stem; en la columna stem, los valores por medias decenas y en la columna hoja, los valores de la variable (el valor se considera en unidades).



**Figura 5.7.** Gráfica de tallos y hojas en la variable edad.





**FIGURA 5.10.** Comparación de la gráfica de la distribución de los datos y la gráfica de caja y bigotes en la variable MELD.

## Medidas de dispersión

Estas medidas tienen como objetivo resumir la heterogeneidad de los datos, es decir, lo separado que estos están entre sí y con respecto de las medidas de posición central.

Las medidas de dispersión contribuyen a configurar la curva de distribución de los datos; con base en esto, se podrá conjeturar que los datos del conjunto, muestra o población en estudio, tienen una aproximación o no a la curva de distribución normal. También permitirá comparar el comportamiento de la variable en estudio en diversos conjuntos de datos, muestras o poblaciones.

Se clasifican en medidas de dispersión absoluta y medidas de dispersión relativa, que se verán a continuación. Solo se incluyen las fórmulas en las medidas que no son calculadas por el SPSS.

### MEDIDAS DE DISPERSIÓN ABSOLUTA

Estas medidas están dadas en las mismas unidades en las que se mide la variable: recorridos, desviaciones medias, varianza, desviación típica (estándar) y desviación media.

#### ● RECORRIDO O RANGO

Es el valor que resulta de la diferencia entre los valores máximo y mínimo observados. Es la medida de dispersión más sencilla y la que proporciona menos información. Tiene el inconveniente de ser susceptible de error, por cuanto solo toma en cuenta dos valores del total de la serie de casos u observaciones, lo que puede provocar una distorsión de la realidad. El riesgo se presenta, sobre todo, si en la serie de casos se tiene valores extremos (*outliers*).

El rango o recorrido también pueden calcularse en los cuartiles, deciles o percentiles:

$$Q_{1-3} = Q_3 - Q_1$$

$$P_{10-80} = P_{80} - P_{10}$$

También pueden calcularse los rangos semicuartílico o semipercentílico.

$$SQ_{1-3} = \frac{Q_3 - Q_1}{2}$$

$$SQ_{1-3} = \frac{P_{80} - P_{10}}{2}$$

- **DESVIACIÓN MEDIA**

Es una medida que valora el promedio de las desviaciones de cada uno de los valores con respecto a la media del conjunto de datos.

Para el cálculo de la desviación media entre el valor específico de la variable estadística y la media aritmética, se usa la siguiente fórmula:

$$D_1 = \text{valor específico} - \text{media}$$

Para el cálculo de la desviación media de los valores absolutos de las desviaciones respecto a la media, se usa la siguiente fórmula:

$$D_x = \frac{(x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x})}{N}$$

$$D_x = \frac{\sum (x_1 - x)}{N}$$

Cuando se suman las diferencias de todos los datos no se debe tomar en cuenta el signo. Como se observa, esta medida es una media aritmética.

- **VARIANZA**

Matemáticamente, se podría decir que la varianza es la media aritmética del cuadrado de las desviaciones respecto a la media de una variable de un conjunto de datos. Si la variable, por ejemplo, utiliza como unidad el metro, la varianza se expresa en metros cuadrados; en caso de la edad, la varianza se expresa en años cuadrados. Esto imposibilita su utilidad, por lo que se recurre a la desviación estándar.

- **DESVIACIÓN ESTÁNDAR (desviación típica)**

Es la raíz cuadrada de la varianza. Informa de la media de distancias que tienen los datos con respecto a la media de la variable. Se asume que todos los datos o valores de la variable se encuentran a tres desviaciones de la media, tanto por encima como por debajo de la misma.

La medida de dispersión absoluta más utilizada es la desviación estándar, que informa de la variabilidad absoluta de los valores con respecto a la media. Como se veía en la curva normal, estas desviaciones incluyen un determinado porcentaje de la población con determinados valores.

TABLA 5.3. MEDIDAS DE DISPERSIÓN PARA LA VARIABLE EDAD

Medida	Edad
• Desviación estándar	11,898
• Varianza	141,558
• Mínimo	26,00
• Máximo	97,00
• Rango o recorrido	71,00

Fuente: Base de datos ficticia de cirrosis (Anexo 9)

**POR EJEMPLO**, en la Tabla 5.3, se presentan ilustrativamente los resultados del análisis de frecuencias para la variable edad, según la base de datos ficticia de cirrosis ya utilizada.

A continuación, se analizan las medidas de dispersión absoluta para la variable edad.

- **RANGO O RECORRIDO.** En esta muestra, los pacientes incluidos están comprendidos entre las edades de 26 años a 97 años, con un rango o recorrido de 71 años. Esta medida permitiría comparar inicialmente si dos grupos o poblaciones son semejantes por su rango.
- **DESVIACIÓN ESTÁNDAR.** Su valor es 11,89 años.

Con los valores de la media y la desviación estándar se podría determinar la concentración porcentual de personas por edad de acuerdo a la desviación estándar. A partir de la media, se van agregando a cada lado 11,89 años.

En la Tabla 5.4, se visualiza el porcentaje de población comprendida en el rango comprendido entre determinadas desviaciones estándar.

TABLA 5.4. VALORES Y PORCENTAJES DE ACUERDO A LA DESVIACIÓN ESTÁNDAR (DE) PARA LA VARIABLE EDAD

	-3 DE	-2 DE	-1 DE	Media	+1 DE	+2 DE	+3 DE
• Edad	25,70	37,60	49,50	61,40	73,30	85,20	97,10
• % de población incluida a ambos lados de la media	2,15	13,59	34,13		34,13	13,59	2,15
			68,26				
			95,44				
			100,00				
• % de población incluida cuando se consideran valores a cada lado de la media			34,13		34,13		
		47,72			47,72		
		50,00			50,00		

Fuente: Base de datos ficticia de cirrosis (Anexo 9).

## Medidas de dispersión relativa

Estas medidas informan de la dispersión en forma relativa, como un porcentaje. Permiten una sencilla comparación entre la dispersión de distintas distribuciones. Matemáticamente es la división de cualquier medida de dispersión entre cualquier medida de tendencia central.

Incluyen el coeficiente de variación, el coeficiente de apertura, los recorridos relativos y el índice de desviación respecto de la mediana. El más utilizado es el coeficiente de variación.

### COEFICIENTE DE VARIACIÓN

Llamado también coeficiente de variación de Pearson. Su fórmula expresa la desviación estándar como porcentaje de la media aritmética. Muestra una mejor interpretación porcentual del grado de variabilidad que la desviación típica o estándar. Permite comparar distribuciones con distintas unidades y de diferentes muestras o poblaciones. Matemáticamente, es el cociente entre la desviación estándar y la media.

Fórmula para obtener el coeficiente de variación (CV)

$$CV = \frac{\text{desviación estándar}}{\text{media}} \times 100$$

Se podría comparar la variabilidad de los valores de edad en dos estudios.

**POR EJEMPLO**, dos investigadores realizan un estudio sobre hepatocarcinoma en sus respectivos hospitales y tienen los siguientes resultados:

- Estudio 1: media = 61,4 años y desviación estándar = 11,89 años.
- Estudio 2: media = 50 años y desviación estándar = 11,89 años.

La primera impresión es que en ambos estudios la variabilidad es la misma, pero según el cálculo del coeficiente de variación se tiene lo siguiente:

- Estudio 1: CV = 19,36 %
- Estudio 2: CV = 23,78 %

Se concluye que en el segundo estudio la variabilidad es más amplia.

### COEFICIENTE DE APERTURA

Matemáticamente es el cociente entre los valores extremos de la distribución de datos.

### RECORRIDOS RELATIVOS

Matemáticamente es el cociente entre el recorrido y la mediana. Estos recorridos pueden ser:

- **RECORRIDO INTERCUARTÍLICO RELATIVO (RIR)**. Es el cociente entre la diferencia del cuartil 3 (Q3) y el cuartil 1 (Q1) y la mediana o cuartil 2 (Q2)
- **RECORRIDO SEMICUARTÍLICO RELATIVO**. Es el cociente entre la diferencia del cuartil 3 (Q3) y la sumatoria de los mismos cuartiles.

### ÍNDICE DE DESVIACIÓN RESPECTO DE LA MEDIANA

Matemáticamente es el cociente entre la desviación media con respecto a la mediana y la mediana.

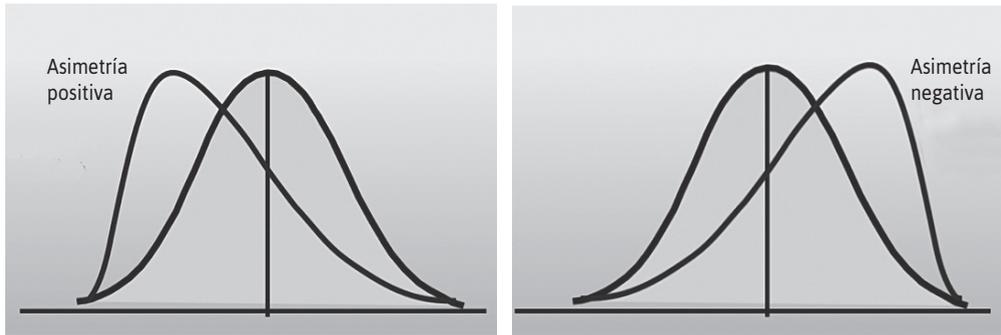


FIGURA 5.11. Tipos de asimetría. A. Asimetría positiva. B. asimetría negativa.

## Medidas de forma

Estas medidas informan sobre el aspecto que tiene la gráfica de la distribución. Son el coeficiente de asimetría y el coeficiente de curtosis.

### MEDIDAS DE ASIMETRÍA

Se afirma que una distribución es simétrica cuando los casos se distribuyen simétricamente a ambos lados de la media. En caso contrario, es una distribución asimétrica. El indicador de esta asimetría se denomina coeficiente de asimetría y resulta de la sumatoria de todas las diferencias entre cada uno de los valores y la media.

La simetría perfecta da como resultado un coeficiente de '0' y, en este caso, existe el mismo número de valores a la derecha que a la izquierda de la media.

- **ASIMETRÍA POSITIVA** (a la derecha). La “cola” de la derecha de la media se alarga más que la izquierda, es decir, existen más valores separados de esa media a la derecha. Figura 5.11A.

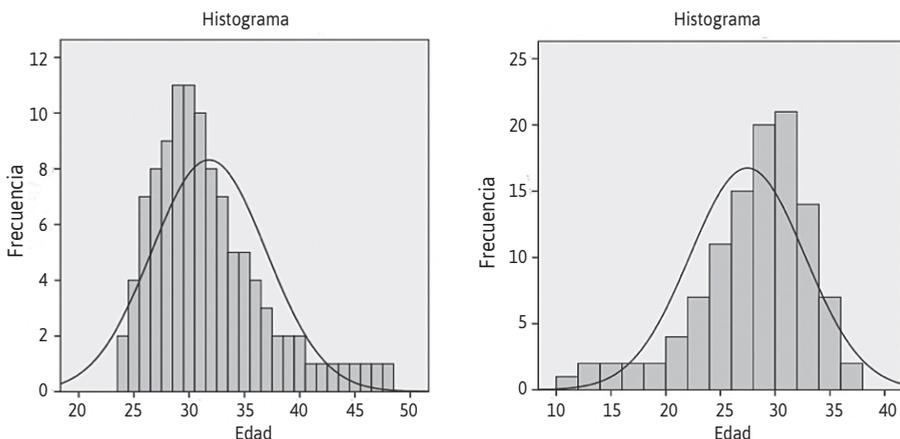


FIGURA 5.12. A. Asimetría positiva de la variable edad. B. Asimetría negativa de la variable edad.

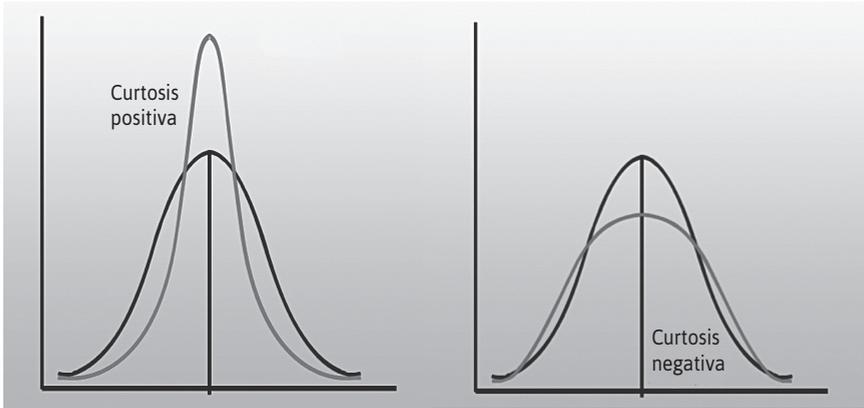


FIGURA 5.13. Curtosis. A. Curtosis positiva. B. Curtosis negativa.

- **ASIMETRÍA NEGATIVA** (a la izquierda). La “cola” de la izquierda de la media es más larga que la derecha, es decir, existen más valores separados de la media a la izquierda. Figura 5.11B.

En la Figura 5.12 se presentan las asimetrías positiva y negativa de la variable edad en dos muestras hipotéticas.

**MEDIDAS DE APUNTAMIENTO O CURTOSIS**

Con esta medida se valora el grado de concentración que presentan los valores en la región central de la distribución. El indicador de curtosis es el coeficiente de curtosis.

Según el valor del coeficiente de curtosis, se tiene lo siguiente:

- **VALOR POSITIVO.** Existe una gran concentración de valores (leptocúrtica). Figura 5.13A.
- **VALOR CERO (‘0’).** Existe una concentración normal (mesocúrtica).
- **VALOR NEGATIVO.** Existe una baja concentración (platicúrtica). Figura 5.13B.

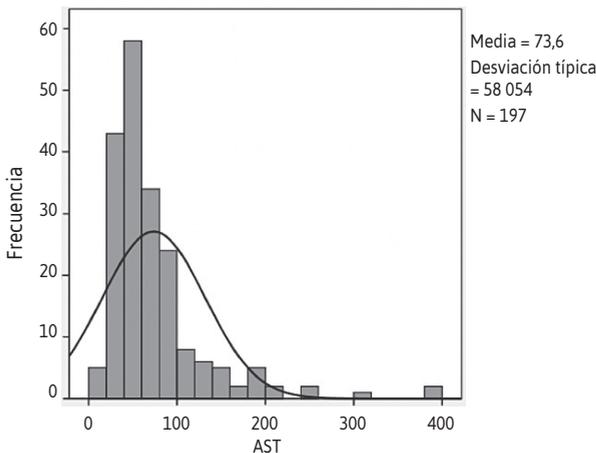


FIGURA 5.14. Curtosis positiva y asimetría positiva en la variable aspartato aminotransferasa (AST).

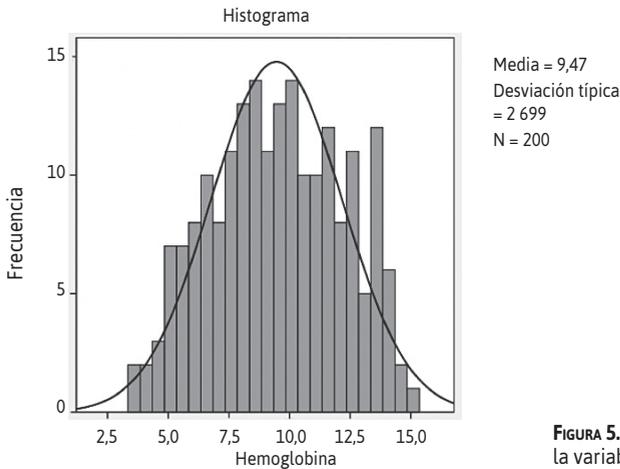


FIGURA 5.15. Curtosis negativa en la variable hemoglobina

Las Figuras 5.14 y 5.15 presentan las gráficas de las variables aspartato aminotransferasa (AST, antes llamada TGO) y hemoglobina, respectivamente, de la base de datos ficticia de cirrosis (Anexo 9).

### Valores extremos

Un valor extremo (*outlier*) es una observación cuyo valor no parece corresponderse con el resto de los valores en el grupo de datos. Estos valores atípicos suelen llamarse aberrantes o inconsistentes (Figura 5.16).

Los valores extremos se pueden clasificar en cuatro categorías.

- Los que surgen de un error de procedimientos.
- Los que surgen como consecuencia de un acontecimiento extraordinario. En este caso no representa ningún segmento válido de la población y puede ser eliminado del análisis.

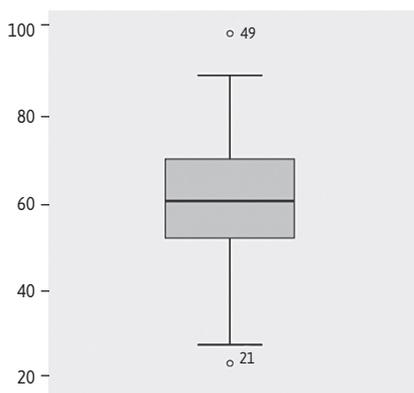


FIGURA 5.16. Gráfica caja y bigotes y datos extremos.

- Observaciones cuyos valores caen dentro del rango de las variables en estudio, pero que son únicas en la combinación de los valores de esas variables. Estas observaciones no se deben eliminar en el análisis. Se recomienda analizar qué influencia ejercen en las estimaciones, pruebas y modelos estadísticos considerados.
- Observaciones extraordinarias y que no tiene explicación aparente. En estos casos se realiza un análisis que considere estos datos y otro sin considerarlos. De este modo, se valora la influencia que tienen. Si estas observaciones tienen influencia, se reportan en las conclusiones y es importante investigar el porqué de dichas observaciones

El SPSS distingue dos tipos de *outliers*:

- **DÉBIL.** Dista del cuartil más cercano más de 1,5 veces el recorrido intercuartílico. Se representa por un '0'.
- **EXTREMO.** Dista del cuartil es mayor y se representa por un '\*'

### Evaluación de los supuestos estadísticos

Cuando se tiene la intención de utilizar pruebas paramétricas en variables cuantitativas, se debe evaluar que los datos de la variable en estudio sigan una distribución normal. Si no se cumple con este supuesto, no se podrán aplicar estas pruebas estadísticas y, en su lugar, se tendrá que recurrir a las pruebas no paramétricas.

Para la evaluación de la normalidad de la variable cuantitativa, se puede utilizar los siguientes procedimientos:

- **OBSERVACIÓN DE LAS MEDIDAS DE RESUMEN**
  - Media, mediana y moda.
  - Asimetría y curtosis.
- **MÉTODOS GRÁFICOS**
  - Histograma.
  - Diagramas de cuantiles (*Q-Q plots*).
- **PRUEBAS ESTADÍSTICAS**
  - Test de Kolmogorov-Smirnov.
  - Contraste de Shapiro-Wilks.

### OBSERVACIÓN DE LAS MEDIDAS DE RESUMEN

- **MEDIA, MEDIANA Y MODA.** En el caso de observar las medidas de resumen, se debe evaluar si la media, mediana y moda tienen el mismo valor o se aproximan. Esto indica que la variable tiene o se aproxima a una configuración normal. Si estas medidas son muy diferentes, induce a pensar que la variable no se aproxima a una distribución normal.
- **ASIMETRÍA Y CURTOSIS.** Se observan la asimetría y la curtosis, y se examinan sus valores. Si son cercanos a '0', se sospecha de normalidad. Si los valores se alejan del '0', es poco probable que los datos de la variable sigan una distribución normal.

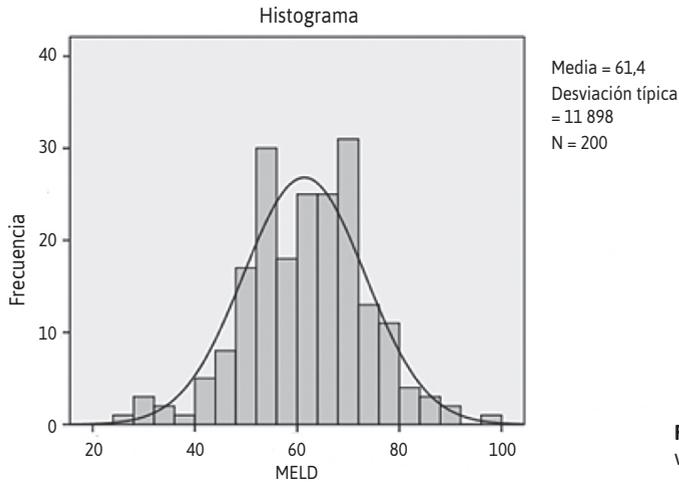


FIGURA 5.17. Histograma de la variable edad.

### MÉTODOS GRÁFICOS

Los métodos gráficos son prácticos para evaluar la normalidad, pero tienen el sesgo de la subjetividad del observador.

- **HISTOGRAMA.** Compara los valores de los datos observados con una distribución normal. En la Figura 5.17, se presentan dos imágenes, el histograma de la variable edad y una línea curva que representa la curva normal. Se observa que algunos valores escapan de la curva normal, sin embargo, la gran parte de ellos caen sobre la curva normal, la sospecha es de que la variable edad tiene una distribución normal.
- **DIAGRAMAS DE CUANTILES.** Otra gráfica importante en la evaluación de la normalidad es la gráfica de diagramas de cuantiles. Esta gráfica compara en un sistema de coordenadas cartesianas, los

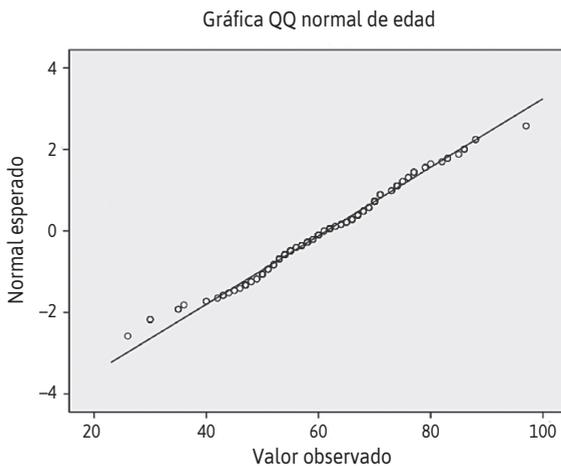


FIGURA 5.18. Diagrama de cuantiles de la variable edad.

**TABLA 5.5** PRUEBAS DE NORMALIDAD.

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
• Edad	0,055	200	0,200	0,991	200*	0,222

a. Corrección de la significación de Lilliefors  
 \* Este es un límite inferior de la significación verdadera

cuantiles muestrales (eje X) con los cuantiles esperados (eje Y) bajo la hipótesis normalidad. En caso de que la variable tenga una distribución normal, este diagrama tenderá a ser una recta que pasa por el eje origen. Mientras más se desvía de la recta, más se debe sospechar que no se trata de una distribución normal. En la Figura 5.18, se observa la gráfica *Q-Q plot* de variable edad, los valores de los casos se distribuyen aproximándose a la línea recta.

Sin embargo, las pruebas estadísticas de normalidad son las que en definitiva dan la valoración estadística de normalidad objetiva.

Para tener la certeza de que la variable cuantitativa tiene una distribución normal, se utilizan la prueba o test de Kolmogorov-Smirnov y el contraste de Shapiro-Wilk.

- **PRUEBA DE KOLMOGOROV-SMIRNOV.** Se toma en cuenta cuando la muestra es grande, por encima de 200.
- **CONTRASTE DE SHAPIRO-WILK.** Cuando la muestra es pequeña.

La significancia estadística se valora por el valor de  $p < 0,05$ . Con estas pruebas lo que se busca es evaluar si la distribución de los datos de la muestra es igual o diferente a una distribución normal. Se acepta la hipótesis de que los datos siguen una distribución normal si el valor de  $p > 0,05$ ; en caso de que el valor de  $p < 0,05$ , se rechaza esta hipótesis, entonces la variable de estudio no tiene una distribución normal. En SPSS, se presenta la valoración de las dos pruebas, el investigador debe tomar en cuenta la prueba adecuada de acuerdo con el número de casos de la muestra.

**POR EJEMPLO**, se considera la variable edad, que tiene un total de 200 casos. SPSS presenta el siguiente resultado:

La Tabla 5.5 presenta la prueba de Kolmogorov-Smirnov y el contraste de Shapiro-Wilk. En el contraste de Shapiro-Wilk (200 casos de la muestra) se observa que el valor de  $p = 0,222$ . Este valor es mayor que 0,05, por lo que no se rechaza la hipótesis nula de igualdad con la distribución normal. Se concluye que en la muestra estudiada la variable edad sigue una distribución normal.

# Análisis de variables cualitativas

## 6

Como se vio anteriormente, las variables cualitativas se clasifican en variables nominales y ordinales. Se realizará el análisis de cada una de ellas por separado.<sup>2</sup>

### Variabes nominales

Las variables nominales tienen solo como medida de resumen a las frecuencias. Estas frecuencias se pueden expresar en frecuencias absolutas y frecuencias relativas (proporciones, porcentajes); además, se pueden presentar como frecuencias simples o frecuencias acumuladas.

Cuando las variables nominales son politómicas otra medida de resumen importante es la moda.

### FRECUENCIAS

Se denomina frecuencia al número de veces que se repite cada una de las categorías de la variable cualitativa.

En la Tabla 6.1, se presentan los resultados de la variable ‘sexo’, de la base de datos ficticia.

En la columna ‘frecuencia’ se puede observar que la categoría ‘hombre’ se repite 115 veces y la categoría ‘mujer’, 85 veces. En este caso se están presentando las frecuencias absolutas.

Además, se observa que a cada una de estas categorías se acompaña en la columna ‘porcentaje’ su frecuencia relativa, es decir, el cociente entre la frecuencia absoluta y el tamaño de la muestra (total de sujetos tanto hombres como mujeres que conforman la muestra de estudio).

**TABLA 6.1.** MEDIDAS DE RESUMEN DE LA VARIABLE SEXO

	Frecuencia	Porcentaje	Frecuencia acumulada	Porcentaje acumulado
• Sexo Hombre	115	57,5	115	57,5
Mujer	85	42,5	200	100,0
Total	200	100,0		

Fuente: Base de datos ficticia de cirrosis (Anexo 9)

**TABLA 6.2.** MEDIDAS DE RESUMEN DE LA VARIABLE ETIOLOGÍA DE LA CIRROSIS

Etiología	Frecuencia	Porcentaje	Frecuencia acumulada	Porcentaje acumulado
• Alcohólica	80	40,0	80	40,0
• Virus de la hepatitis B	15	7,5	95	47,5
• Virus de la hepatitis C	14	7,0	109	54,5
• Alcohol-viral	4	2,0	113	56,5
• Autoinmune	11	5,5	124	62,0
• Colangitis biliar primaria	5	2,5	129	64,5
• Otro	4	2,0	133	66,5
• No determinada	67	33,5	200	100,0
Total	200	100,0		

Fuente: Base de datos ficticia de cirrosis (Anexo 9)

Para el cálculo de la frecuencia relativa (Fr) o porcentaje, se aplica la siguiente fórmula:

$$Fr = \frac{\text{Categoría de interés}}{n} \times 100$$

En este caso, las frecuencias relativas se expresan en porcentaje.

En el Tabla 6.1, se tiene un 57,5 % de hombres y 42,5 % de mujeres en la muestra de estudio.

Las frecuencias absolutas se pueden ir acumulando. Para esto, se agrega el valor absoluto de la categoría a la siguiente, progresivamente de acuerdo al número de categorías que tenga la variable. En el caso anterior, solo se puede acumular una sola vez, las frecuencias absolutas de los hombres se acumulan a las frecuencias absolutas de las mujeres y se obtiene la información de la columna ‘frecuencia absoluta acumulada’. En el caso de una variable politómica como ‘etiología de la cirrosis’, el proceso de acumulación tiene varias fases (Tabla 6.2).

Un proceso parecido se tiene en el caso de la ‘frecuencia relativa acumulada’ (porcentaje acumulados), en la cual se van acumulando las frecuencias relativas (porcentajes).

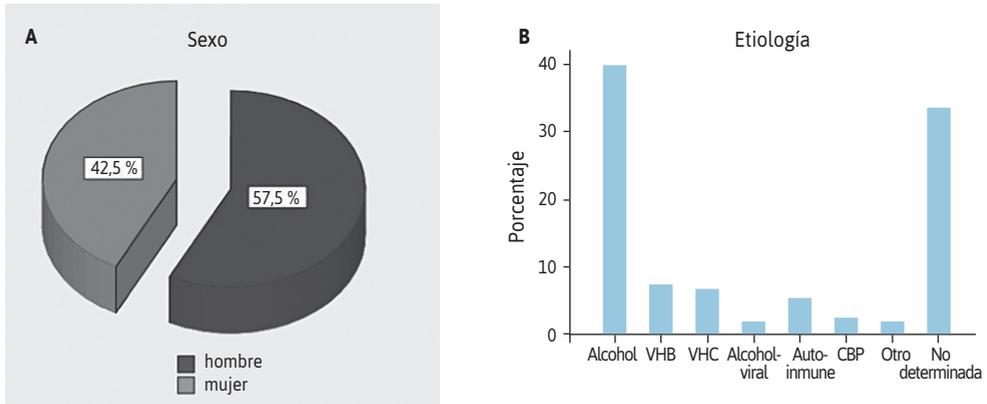
### MODA

En una variable nominal politómica, tiene sentido averiguar cuál de todas tiene la mayor frecuencia, es decir la moda. En la Tabla 6.2, se ve que la etiología ‘alcohólica’ de la cirrosis hepática es la más frecuente.

### GRÁFICAS

Las gráficas que se pueden utilizar en estas variables son las de sectores y de barras. Figura 6.1.

Las frecuencias acumuladas se representan en una gráfica llamada ojiva, en esta gráfica el eje de las abscisas corresponde a los límites de clase y el de las ordenadas a los porcentajes acumulados. Figura 6.2.



**FIGURA 6.1.** Gráficas. A) Gráfica de sectores de la variable 'sexo'. B) Gráfica de barras de la variable 'etiología' de la cirrosis hepática. VHB: virus de la hepatitis B; VHC: virus de la hepatitis C; CBP: colangitis biliar primaria.

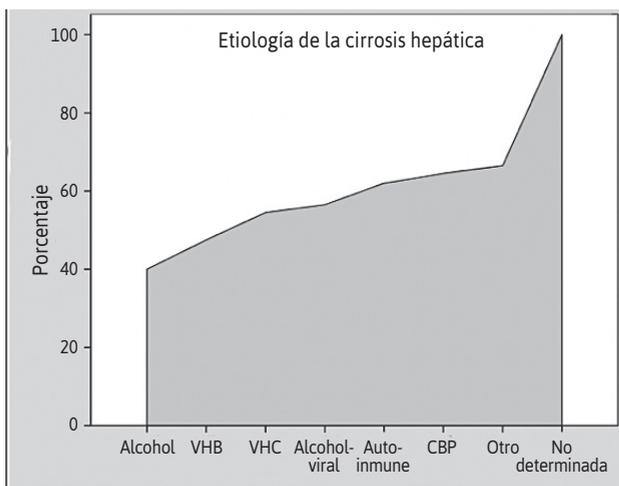
### RAZONES E ÍNDICES

A partir de las frecuencias obtenidas se pueden calcular razones e índices. En la razón se establece una relación por cociente entre las unidades de análisis que pertenecen a un grupo o categoría (a) y las unidades de análisis que pertenecen a otra categoría (b) de la misma variable.

**POR EJEMPLO,** en un determinado estudio, se halló que la cirrosis estaba presente en 230 hombres (H) y 112 mujeres (M). La relación hombre/mujer es 230/112.

$$\begin{aligned} \text{Razón H} &= \frac{\text{casos con cirrosis en hombres}}{\text{casos con cirrosis en mujeres}} \\ &= 230/112 = 2,05 \end{aligned}$$

La razón hombre/mujer con cirrosis es de 2,05; o 2,05:1 y expresa que hay más de 2 hombres con cirrosis por cada mujer con cirrosis en esa población, en ese período en que se realizó el estudio.



**FIGURA 6.2.** Gráfica de la ojiva de la variable 'etiología' de la cirrosis hepática.

En el caso del índice, lo que se busca es expresarlo, pero en 100 unidades.

$$\text{Índice} = H/M \times 100$$

**POR EJEMPLO**, en el caso anterior:

$$\begin{aligned}\text{Índice H/M} &= \text{casos con cirrosis en hombres} / \text{casos con cirrosis en mujeres} \\ &= 230/112 \times 100 = 205\end{aligned}$$

Es decir, que por cada 100 mujeres con cirrosis hay 205 hombres con cirrosis en esa población y en ese período.

## MEDIDAS EPIDEMIOLÓGICAS

Con las frecuencias obtenidas también se pueden calcular medidas epidemiológicas que valoren la magnitud, el riesgo y la velocidad con que las enfermedades se presentan en una determinada población.<sup>6,17</sup>

### ● MAGNITUD DE LA ENFERMEDAD

Se puede medir la magnitud de las enfermedades, de los factores de riesgo o de cualquier evento de interés. Para esto, se recurre a la prevalencia.

- **PREVALENCIA.** Número de casos dividido por la población en estudio o población en riesgo (PER: pasible de tener la condición). Es una de las medidas más básicas en epidemiología y da una imagen “instantánea” de la magnitud de los problemas de salud.

Matemáticamente, la prevalencia es una proporción expresada x 100, x 1 000, x 10 000 o x 100 000 personas, según criterios estandarizados en salud pública.

La prevalencia también expresa la probabilidad de que una persona seleccionada al azar de la PER tenga la enfermedad o problema de salud en estudio en un determinado tiempo y lugar.

$$\text{Prevalencia} = \text{casos} / \text{población en riesgo}$$

**POR EJEMPLO**, un estudio ficticio, realizado en el año 2010, en un distrito de Lima para determinar la prevalencia de cirrosis hepática en la población adulta mayor de 30 años. Para este, se tomó una muestra al azar de 5 150 personas, en las que se encontró un total de 42 casos diagnosticados de cirrosis hepática.

$$\begin{aligned}\text{Prevalencia de cirrosis hepática (PCH)} &= \text{casos con cirrosis} / \text{población mayor de 30 años} \\ &= 42/5150 \\ &= 815,5 \times 100\ 000 \text{ personas}\end{aligned}$$

En la muestra se tiene una prevalencia de cirrosis hepática de 815,5 x 100 000 personas mayores de 30 años en el año 2010.

Esta medida de resumen se podría extrapolar para toda la población del distrito de donde se obtuvo la muestra. Para esto, se tiene que agregar su intervalo de confianza (IC).

Para el cálculo del intervalo de confianza de la prevalencia, se utiliza la fórmula del IC de una proporción (Ver Capítulo 6).

- **RIESGO DE ENFERMEDAD.** La medida epidemiológica que informa sobre el riesgo de enfermar es la incidencia acumulada.

- **INCIDENCIA ACUMULADA.** Expresa en probabilidades el riesgo de enfermar (ser un nuevo caso) en un determinado tiempo y lugar.

Matemáticamente, la incidencia acumulada es el número de casos nuevos dividido por la población en estudio o población en riesgo (“pasible de ser caso nuevo”), en un período determinado. También es una proporción que se expresa x 100, x 1 000, x 10 000 o x 100 000 personas.

Incidencia acumulada = casos nuevos en el período especificado / población en riesgo

Al igual que la prevalencia, para que esta medida de resumen pueda extrapolarse a la población del distrito de donde se obtuvo la muestra, se debe calcular su intervalo de confianza (Ver Capítulo 6).

**POR EJEMPLO,** el estudio ficticio presentado se complementó con un seguimiento durante un año, de enero a diciembre de 2010. A los 5 108 que no presentaron cirrosis hepática en el momento del estudio se los volvió a evaluar y se encontró que 15 de ellos tenían el diagnóstico de cirrosis hepática. Se asume que esos 15 casos son casos nuevos en un año en la población en estudio o riesgo.

IA de cirrosis hepática = casos nuevos con cirrosis / población mayor de 30 años  
 = 15/5 108  
 = 293,7 x 100 000 personas

En el estudio se determinó que la IA de cirrosis hepática es de 293,7 x 100 000 personas mayores de 30 años en el año 2010.

- **VELOCIDAD DE APARICIÓN DE ENFERMEDAD.** La tasa de incidencia informa sobre la velocidad a la cual se desarrollan los casos nuevos en la población.

- **TASA DE INCIDENCIA.** Llamada también densidad de incidencia. Es el número de casos nuevos dividido por la población en estudio o la población en riesgo x tiempo. Es una razón que se expresa x 100, x 1 000, x 10 000 o x 100 000 años personas.

Tasa de incidencia = casos nuevos en el período especificado / población tiempo

En los estudios de seguimiento en el tiempo, los participantes constituyen una cohorte cuya característica común es el hecho de estar libre del evento o resultado de estudio y estar en riesgo de sufrir ese evento o resultado y a la cual se la debe vigilar para detectar todos los primeros eventos o resultados, así como también el registro de su pérdida.

Se debe tener en cuenta que otras tasas no se expresan en años persona. La tasa de mortalidad materna (TMM) se expresa x 100 000 nacidos vivos.

TMM = número de mujeres que mueren durante el embarazo y el parto / nacidos vivos

Algunos valores se presentan como tasas, pero realmente son prevalencias o incidencias acumuladas.

**TABLA 6.3.** CONTRIBUCIÓN DE AÑOS PERSONAS DE LOS CASOS Y NO CASOS

	N.º	Años persona
• Personas que enfermaron de cirrosis	15	7,91
• Personas que no enfermaron de cirrosis	5 093	5 093,00
Total	5 108	5 100,90

Fuente: Datos ficticios.

**POR EJEMPLO**, en el mismo estudio ficticio de seguimiento se tuvo la precaución de registrar la fecha de diagnóstico de la cirrosis y de registrar, en caso fuera necesario, la fecha de abandono o pérdida de los participantes.

En el estudio ficticio presentado, se realizó el seguimiento de 5 108 personas mayores de 30 años. La Tabla 6.3 registra los años persona que contribuyeron.

$$\begin{aligned} \text{Incidencia acumulada de cirrosis hepática} &= \text{casos nuevos con cirrosis} / \text{población tiempo} \\ &= 15/5\ 100,9 \\ &= 294,1 \times 100\ 000 \text{ años persona} \end{aligned}$$

La tasa de incidencia es de 294,1 x 100 000 años persona, es decir, la velocidad con que aparece un caso de cirrosis hepática en esta población.

Finalmente, estas medidas se pueden presentar en forma bruta o específica.

- **BRUTA** (cruda, general o global). Cuando en el denominador de la tasa aparece la población total.
- **ESPECÍFICA**. Cuando la población está restringida a una parte, por ejemplo, un grupo de edad o sexo, etc.

Es importante distinguir cuando una tasa es bruta o específica en el momento del análisis, porque las medidas brutas de diferentes territorios o de un mismo territorio en diferentes períodos de tiempo no se pueden comparar si no se le somete previamente a un procedimiento denominado estandarización.

### Variables ordinales

Las variables ordinales, como las variables cualitativas, tienen la característica de estar constituidas por categorías, pero a diferencia de las cualitativas, entre estas categorías existe una diferencia de valor, por lo que tienen un orden de menor a mayor o viceversa.

Las medidas de resumen en estas variables son las frecuencias, la moda, la mediana, el valor mínimo, el valor máximo, los cuartiles, los percentiles y el rango intercuartílico. Claro está, si el número de categorías de la variable ordinal lo permite; en caso de tener solo tres valores no se podrá tener todas las medidas mencionadas.

TABLA 6.4. FRECUENCIAS DE LA VARIABLE INGRESOS ECONÓMICOS

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
• Ingresos muy bajos	42	31,1	31,1	31,1
• Ingresos bajos	37	27,4	27,4	58,5
• Ingresos medios	25	18,5	18,5	77,0
• Ingresos altos	19	14,1	14,1	91,1
• Ingresos muy altos	12	8,9	8,9	100,0
Total	135	100,0	100,0	

Fuente: Datos ficticios.

## FRECUENCIAS

Las frecuencias pueden ser absolutas o relativas.

Con fines de explicación, se tomarán los resultados ficticios de la variable ‘ingresos económicos’ que tienen las categorías ordinales de ingresos muy bajos, ingresos bajos, ingresos medios, ingresos altos e ingresos muy altos. En la Tabla 6.4 se presentan los resultados.

En la columna ‘frecuencia’ se observa que el grupo de ingresos muy bajos se repite 42 veces; el grupo de ingresos bajos, 37 veces; el grupo de ingresos medios, 25 veces; el grupo de ingresos altos, 19 veces; el grupo de ingresos muy altos, 12 veces. Además, se observa los valores de los cinco grupos en las columnas ‘porcentaje’, para las frecuencias relativas, la columna ‘frecuencias acumuladas’ y la columna ‘porcentajes acumulados’.

En la Tabla 6.5, se observan otras medidas de resumen.

TABLA 6.5. MEDIDAS RESUMEN DE VARIABLE INGRESOS ECONÓMICOS

Medida de Resumen	Valor
• Media	2,42
• Mediana	2,00
• Moda	1
• Rango	4
• Mínimo	1
• Máximo	5
• Percentil 25	1,00
• Percentil 50	2,00
• Percentil 75	3,00

Fuente: Datos ficticios.

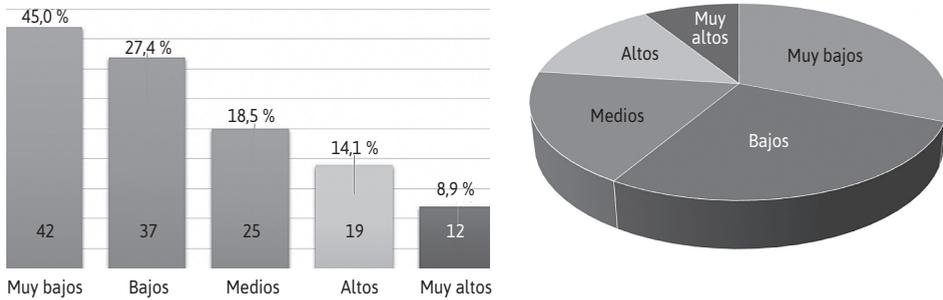


FIGURA 6.4. Gráficas. A. Histograma de la variable ingresos económicos. B. Gráfica de ingresos económicos.

### MEDIA

La media o promedio tiene un valor de 2,42, que corresponde a un nivel intermedio entre el valor 2 (ingresos bajos) y 3 (ingresos medios).

### MEDIANA

La mediana tiene un valor de 2 (que corresponde al grupo de ingresos bajos), es decir, que el 50 % de la población tiene un valor por debajo y por encima de ingresos bajos.

### MODA

El grupo de ingresos muy bajos es el que más se repite (valor 1).

### MÍNIMO Y MÁXIMO

Son los casos 1 y 5, respectivamente, que corresponden al grupo de ingresos muy bajos y el grupo de ingresos muy altos.

### CUARTILES

El cuartil 1 es 1, significa que el 25 % de la población tiene ingresos muy bajos. El cuartil 2 llega al valor 2 y representa también a la mediana. El cuartil 3 tiene un valor de 3, que significa que hasta el 75 % de la población tiene ingresos medios.

### GRÁFICAS

Las gráficas que se pueden utilizar con estas variables son las de barras y las de sectores, que se observan en las Figuras 6.3A y 6.3B, respectivamente.

El intervalo de confianza es un rango de valores calculados a partir de una muestra y en el cual se estima con una probabilidad determinada que se encuentra el verdadero valor del parámetro.

En la Figura 7.1, se tiene hipotéticamente la determinación de la tasa de incidencia (TI) de la población de un determinado problema de salud (parámetro) y la representación del estadístico (TI) y su intervalo de confianza de la muestra del estudio de esa población.

En el caso de una investigación, en la que se tiene datos de una muestra, las medidas de resumen como el promedio (variables cuantitativas) o la proporción (variables cualitativas) que se obtienen dan información de esta muestra. El investigador puede construir a partir de este dato, un intervalo de valores para estimar con un determinado grado de probabilidad de acierto, el valor que tiene el parámetro en la población.

La probabilidad de acierto o éxito en la estimación del parámetro se representa como  $1 - \alpha$  y se denomina nivel de confianza.

En esta diferencia,  $\alpha$  es el error aleatorio o nivel de significación, y se constituye en una medida de las posibilidades de fallar en la estimación mediante el intervalo construido. El investigador puede elegir el grado de error al considerar un determinado nivel de confianza. Así, por ejemplo, si se admite un 90 % de confianza, el error será el 10 %; al 95 % de confianza, el error será del 5 % y al 99 % de confianza, el error será del 1 %.

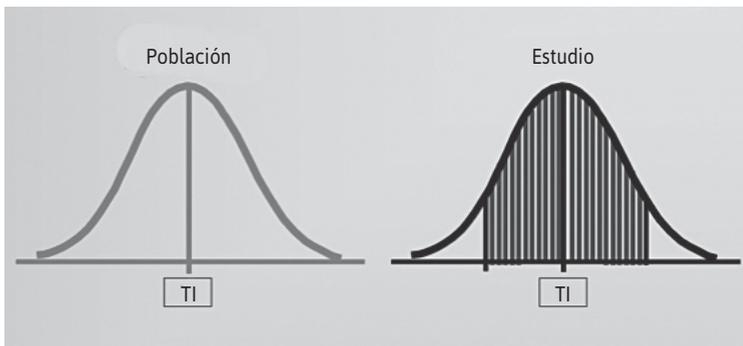


FIGURA 7.1. Gráfica de tasa de incidencia en la población y en el estudio con su respectivo intervalo de confianza.

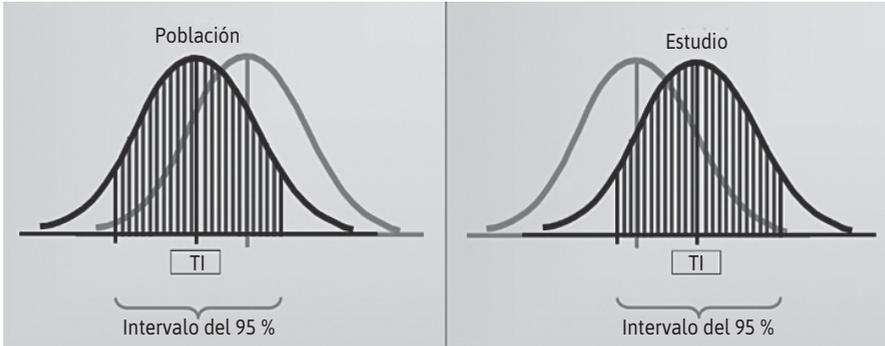


FIGURA 7.2. Gráficas que esquematizan las posibilidades de acertar el valor del parámetro poblacional a partir de muestras estudiadas de esa población.

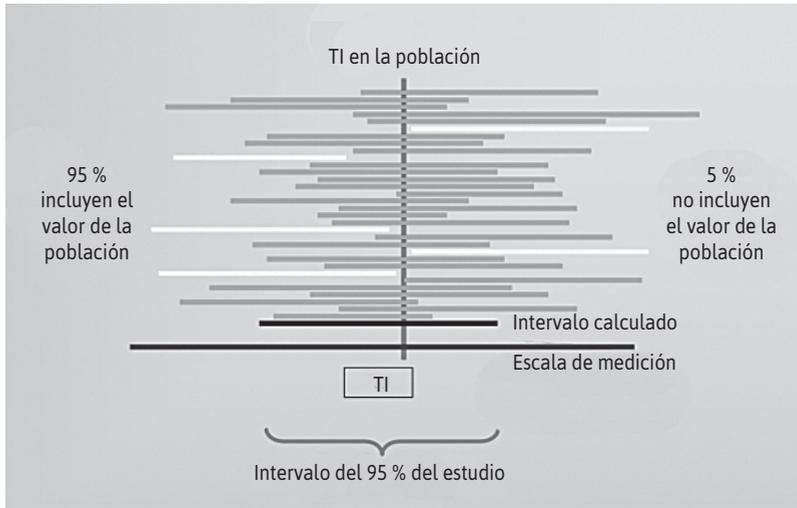
En la Figura 7.2, se grafica las posibilidades de acertar, en los intervalos de confianza, el valor del parámetro de la población. En la imagen de la izquierda, el parámetro se ubica en los valores más altos del intervalo de confianza, mientras que, en la imagen de la derecha, el parámetro se ubica en los valores más bajos del intervalo de confianza.

En la Figura 7.3, se observa la posibilidad de equivocarnos en la valoración del parámetro poblacional a partir de los hallazgos en el estudio. En este caso, el nivel de confianza es al 95 % y la probabilidad de error, del 5 %. Es decir, esa equivocación está en ese 5 %.

En la Figura 7.4, se esquematiza el 95 % de confianza de incluir el valor del parámetro en el intervalo de confianza del estudio y también el 5 % de error en el cálculo del parámetro. Se puede decir que de 100 estudios muestrales de esa población, 95 de ellos incluirán en su intervalo de confianza el valor del parámetro poblacional, mientras que 5 de ellos no incluirán el valor del parámetro.



FIGURA 7.3. Gráficas que esquematizan la posibilidad de error del intervalo de confianza en el cálculo del parámetro poblacional a partir de muestras estudiadas de esa población.



**FIGURA 7.4.** Gráficas que esquematan la posibilidad de error del intervalo de confianza en el cálculo del parámetro poblacional a partir de muestras estudiadas de esa población.

En un intervalo de confianza son importantes el nivel de confianza y la amplitud del intervalo. En el caso de un intervalo muy amplio, habrá mayores probabilidades de acierto (mayor nivel de confianza), pero menor precisión, mientras que un intervalo más corto ofrece una estimación más precisa, pero aumenta las probabilidades de error.

Es importante también el tamaño de la muestra o cantidad de casos. Con un mayor tamaño de muestra, se mejora la precisión.

#### UTILIDAD DE LOS INTERVALOS

- Medida inferencial que permite la extrapolación de la medida de resumen de la muestra a la población.
- Permite estimar el valor del parámetro (medida de la población) a partir del estadígrafo de la muestra con una probabilidad conocida.
- Es un estimador de nivel de significancia estadística, debido al nivel de confianza de incluir en el intervalo al parámetro de la población.
- Permite valorar el nivel de precisión del estudio, un intervalo muy amplio tiene menor precisión.

En caso del análisis bivariado, valora la significancia estadística de la fuerza de asociación entre dos variables a partir de las medidas de efecto (diferencias o razones).

Cuando se realiza el análisis univariado, se puede calcular y valorar el intervalo de confianza para una media (variable cuantitativa) y para una proporción, prevalencia, incidencia acumulada

o tasa de incidencia (variable cualitativa). Con el intervalo obtenido, se puede inferir que el parámetro poblacional se encuentra comprendido en este intervalo.

Para que los intervalos de confianza sean fiables es necesario que se cumplan las dos condiciones siguientes:

- Los datos deben proceder de una muestra aleatoria de la población de interés.
- En el caso de una media de la muestra, esta debe seguir una distribución aproximadamente normal.

## FÓRMULAS

### • INTERVALO DE CONFIANZA PARA UNA MEDIA

$$\text{IC media} = \text{media de la muestra} \pm Z \text{ EE de la media}$$

Donde:

Z es el nivel de significancia expresado en valores de desviación en la curva normal.

Para el 90 %, el 95 % y el 99 % le corresponde un valor de 1,65, 1,96 y 2,34, respectivamente.

EE de la media es el error de la media. Para su cálculo se utiliza la fórmula siguiente:

$$\text{EE de la media} = \text{desviación estándar} / \sqrt{n}$$

### • INTERVALO DE CONFIANZA PARA UNA PROPORCIÓN

Aplicado también para una prevalencia o para una incidencia acumulada.

$$\text{IC proporción} = \text{proporción} \pm Z \text{ EE de la proporción}$$

Donde:

Z es el nivel de significancia expresado en valores de desviación en la curva normal.

EE de la proporción es el error de la proporción. Para su cálculo, se utiliza cualquiera de las fórmulas siguientes:

$$\text{EE de la proporción} = \sqrt{\frac{p(1-p)}{n}}$$

$$\text{EE de la proporción} = \sqrt{\frac{p^2(1-p)}{a}}$$

Donde: a es el numerador de la proporción

### • INTERVALO DE CONFIANZA PARA UNA TASA

$$\text{IC tasa} = \text{tasa} \pm Z \text{ EE de la tasa}$$

Donde:

Z es el nivel de significancia expresado en valores de desviación en la curva normal.

EE de la tasa es el error de la tasa. Para su cálculo, se utiliza la fórmula siguiente:

$$\text{EE de la tasa} = \frac{\text{tasa}}{\sqrt{a}}$$

Donde: a es el numerador de la proporción

**POR EJEMPLO**, con fines meramente aplicativos, se verá un estudio hipotético en donde se estudia la prevalencia de cirrosis en una población y el promedio de aminotransferasas (transaminasas) que presentan los supuestos pacientes con diagnóstico de cirrosis.

El estudio se realizó con una muestra de 2 500 personas, en la que se detectó un total de 35 casos de cirrosis. Los resultados de alanina-aminotransferasa (ALT, antes TGP) en los pacientes con cirrosis tuvieron un promedio de 92 U, con una desviación estándar de 40 U. El objetivo del estudio es determinar el intervalo de confianza para la prevalencia de cirrosis en esa población y el intervalo de confianza para el promedio de aminotransferasas en los pacientes con cirrosis, y, finalmente, la interpretación de estos hallazgos.

#### CÁLCULO DE LA PREVALENCIA Y DE SU IC

$$\text{Proporción} = 35/2\,500 = 0,014$$

Expresada como prevalencia: 1 400 x 100 000 habitantes

IC proporción = proporción  $\pm$  Z EE de la proporción

$$\text{EE de la proporción} = \sqrt{\frac{p^2(1-p)}{a}}$$

$$\begin{aligned} \text{IC proporción} &= 0,014 \pm (1,96 \times 0,00234981) \\ &= 939 - 1861 \times 100\,000 \text{ habitantes} \end{aligned}$$

#### INTERPRETACIÓN DE LA PREVALENCIA Y DE SU IC

- En la población de estudio existe una prevalencia de cirrosis hepática de 1 400 x 100 000 habitantes.
- El valor verdadero de la prevalencia en la población está comprendido en el intervalo comprendido entre 939 x 100 000 habitantes y 1 861 x 100 000 habitantes, con un nivel de confianza del 95 %.

Es decir, si se realizan 100 estudios en la misma población, 95 de ellos (95 %) tendrían una prevalencia comprendida en este intervalo y solo en 5 de ellos (5 %) la prevalencia no estaría incluida en este intervalo (5 % de error).

#### CÁLCULO DEL IC PARA EL PROMEDIO DE ALANINA-AMINOTRANSFERASA (ALT)

IC media = media de la muestra  $\pm$  Z EE de la media

EE de la media = desviación estándar /  $\sqrt{n}$

$$\begin{aligned} \text{IC media} &= 92 \pm (1,96 \times 7,3029674) \\ &= 78 \text{ U} - 106 \text{ U} \end{aligned}$$

#### INTERPRETACIÓN DE LA MEDIA Y SU IC

- El promedio de ALT en pacientes con cirrosis hepática es de 92 U.
- El valor verdadero del promedio de la ALT en la población de pacientes con cirrosis hepática está comprendido en el intervalo comprendido entre 78 U y 106 U con un nivel de confianza del 95 %. Es decir, si se realizan 100 estudios en la población de pacientes con cirrosis, 95 de ellos (95%) tendrían una media comprendida en este intervalo y solo en 5 de ellos (5 %) la media no estaría incluida en este intervalo (5 % de error).

Otra aplicación del IC es saber si existen diferencias entre el parámetro con el estadígrafo que se ha obtenido en la muestra estudiada. En este caso se sabe de antemano cual es el parámetro de la población. Para esto, solo se observa si en el intervalo de confianza de la muestra está incluido el valor poblacional.

- Si lo incluye, se dirá que no existen diferencias y, por lo tanto, la medida obtenida en la muestra, aunque no sea exacta al de la población, no es estadísticamente significativa.
- Si no lo incluye, se afirmará que existen diferencias significativas entre el estadígrafo y el parámetro.

En el análisis bivariado, cuando se quiere valorar el efecto de una variable independiente (factor de riesgo o de protección), se puede calcular para una diferencia de proporciones (incluidas las prevalencias, incidencias acumuladas o tasas de incidencias) y también para el OR y el RR. Solo se presentan las fórmulas para conocimiento, por cuanto en este manual solo se está tratando el análisis univariado.

- **INTERVALO DE CONFIANZA PARA UNA DIFERENCIA DE PROPORCIONES (ICDP) O DE TASAS**

ICDP = diferencia de proporciones  $\pm$  Z EE de diferencia de proporciones

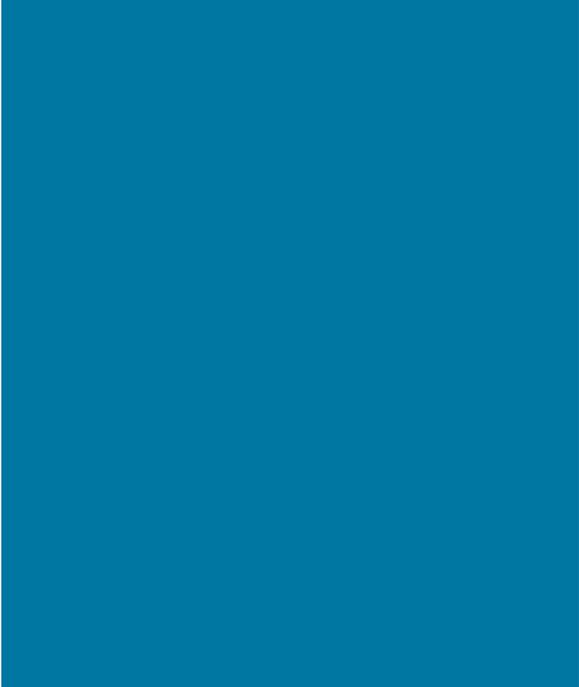
- **INTERVALO DE CONFIANZA PARA UNA RAZÓN DE PROPORCIONES O DE TASAS**

ICRP =  $e^{\ln RP \pm Z EE (\ln RP)}$

- **INTERVALO DE CONFIANZA PARA UNA OR Y RR**

ICOR =  $e^{\ln OR \pm Z EE (\ln OR)}$

ICRR =  $e^{\ln RR \pm Z EE (\ln RR)}$



## ANEXOS





## Diseño de base de datos

### Creación de una base de datos en hoja de cálculo

1. Se utilizará el programa Excel y la base de datos ficticia de cirrosis que está en el Anexo 9 (Tabla A9.1). La lista de las variables se presenta en el Tabla A1.1.
2. En la primera fila de la hoja de cálculo, se coloca el nombre de las variables de acuerdo con

**TABLA A1.1.** LISTA DE VARIABLES

N.º	Nombre	Tipo	Etiqueta	Valores	Escala
1	iden	numérico	Código del paciente		Escalar
2	sex	numérico	Sexo	1 = Hombre 2 = Mujer	Nominal
3	ed	numérico	Edad		Escalar
4	etio	numérico	Etiología	1 = Alcohólica 2 = VHB 3 = VHC 4 = Alcohol – Viral 5 = Autoinmune 6 = CBP 7 = Otro 8 = No determinada	Nominal
5	cchp	numérico	Causa de Hospitalización	1 = Ascitis 2 = Encefalopatía 3 = Falla renal 4 = Hemorragia Digestiva Alta 5 = Infección bacteriana 6 = Ictericia	Nominal
6	pcp	numérico	Puntaje CHILD-PUGH		Escalar
7	meld	numérico	MELD		Escalar
8	asc	numérico	Severidad de ascitis	0 = Ausente 1 = Leve 2 = Moderada 3 = Severa	Ordinal
9	hb	numérico	Hemoglobina		Escalar
10	tgo	numérico	TGO		Escalar
11	tgp	numérico	TGP		Escalar
12	urea	numérico	Urea		Escalar
13	comp	numérico	Complicaciones	0 = Sin complicaciones 1 = Presenta complicaciones	Nominal
14	esthos	numérico	Estancia hospitalaria		Escalar
15	alt	numérico	Condición de alta	1 = Mejorado 2 = Fallecido	Nominal

Fuente: Base de datos ficticia de cirrosis (Anexo 9)

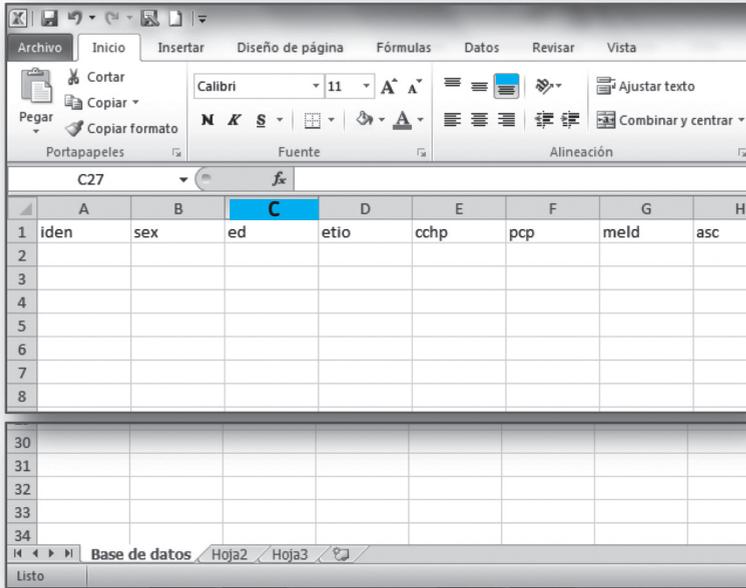


FIGURA A1.1. Base de datos en Excel.

la Lista de variables que se ha diseñado para el estudio. También se podría darle un nombre a la hoja (Base de datos) Figura A1.1.

3. A cada una de las variables se le incluye una etiqueta donde se adjuntan los valores de la variable. Para esto, en la celda donde figura el nombre de la variable, se hace clic con el botón derecho del ratón para que aparezca el menú contextual (Figuras A1.2 y A1.3) y se elige el comando 'Insertar comentario'.

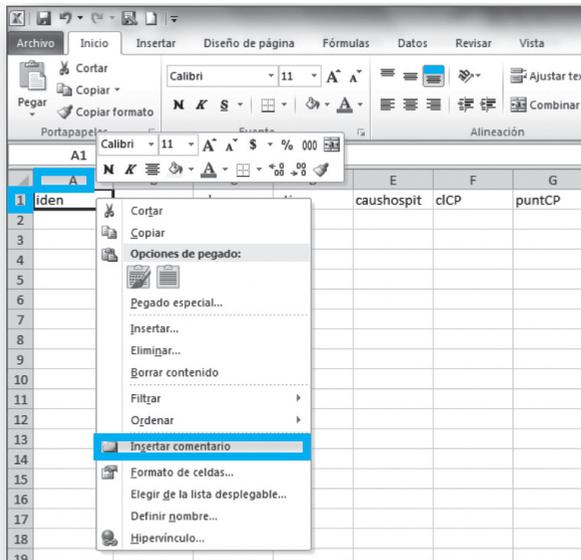


FIGURA A1.2. Insertar comentario.

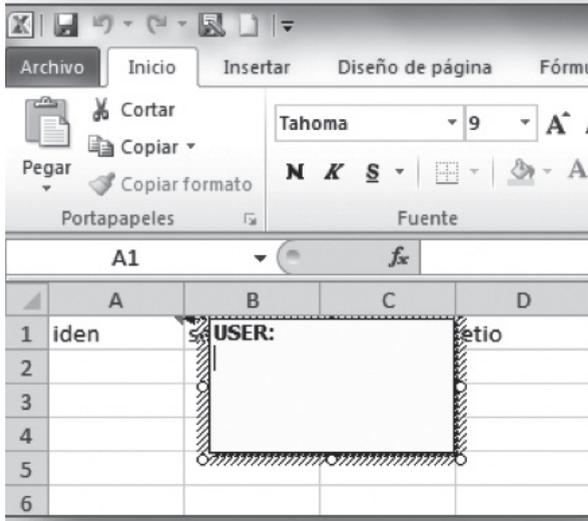


FIGURA A1.3. Etiqueta para incluir información sobre las variables.

4. En la etiqueta para el comentario, colocar la etiqueta de las variables y sus valores, si corresponde (Figura A1.4).

El programa insertará un triángulo rojo en el extremo superior y derecho de la celda de la variable.

Se puede variar el tamaño de la etiqueta arrastrando los bordes de la etiqueta. Al pasar el ratón por la variable que tiene incluido el comentario/etiqueta, se activa la Etiqueta.

5. Se puede colocar en toda la columna de cada una de las variables el comando 'Validación de datos', que evitará introducir valores no permitidos. Para esto, primero, se selecciona la columna haciendo clic sobre la letra de columna; en este caso, se ha hecho el clic sobre la Columna A para la variable 'iden'. Se hace clic sobre el 'Menú Datos'; luego, se hace clic en el comando 'Validación de datos' (Figura A1.5).

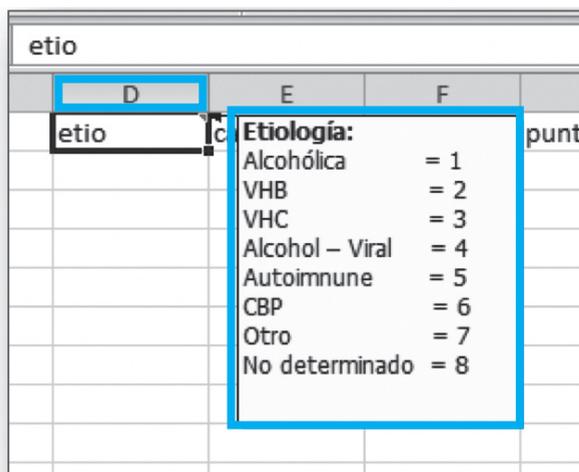


FIGURA A1.4. Etiqueta donde se incluye la información sobre la variable

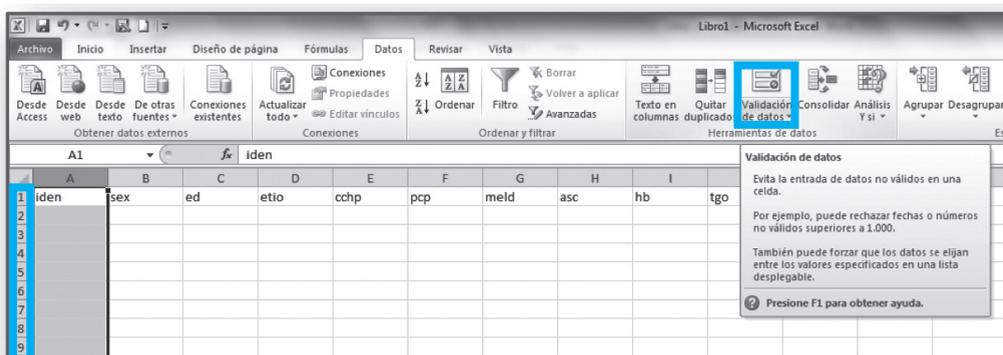


FIGURA A1.5. Validación de los datos de las variables.

Para la variable 'iden', se han considerado algunos 'criterios de validación'. En 'Permitir' el criterio es que sea 'Número entero'; en 'Datos' el criterio 'entre', para luego explicitar que será 'Mínimo' 1 y 'Máximo' 500 (Figura A1-6). Si se digita por error 520, el programa no lo acepta (Figura A1.7).

Es más evidente, por ejemplo, en el caso de Sexo. Se ha incluido los criterios que sea número entero y esté comprendido entre 1 y 2. Si se ingresa el texto Mujer, no lo acepta, se debe introducir el número 2 (valor asignado para el sexo femenino).

La 'Validación de datos' permite evitar la introducción de datos erróneos. Se procede a completar la validación de todas las variables que se crea conveniente.

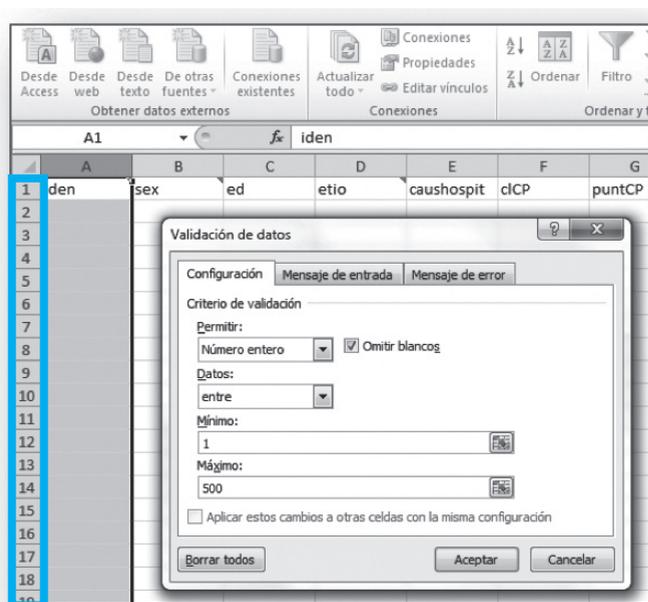


FIGURA A1.6. Validación de datos.

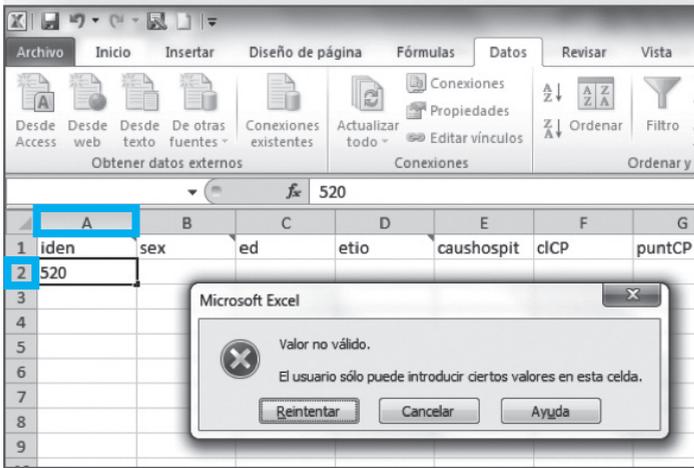


FIGURA A1.7. Mensaje de error por ingreso incorrecto de dato.

6. Se procede a guardar la base de datos, para lo cual se debe asignar un nombre y direccionar la ubicación en la carpeta o lugar que se tiene previsto ubicar el archivo. Es importante direccionar la ubicación del archivo que se va a guardar para evitar su pérdida. Figura A1.8. En el ejemplo, se está direccionando a que el archivo se guarde en 'Equipo' > 'Nuevo vol (E) 04 JMTP' > 'Libro de estadística JMTP'. En 'Nombre del archivo' se le está dando el nombre de Base de Cirrosis, el 'Tipo de archivo', automáticamente Excel le asigna 'Libro de Excel'. Finalmente, se hace clic en el botón 'Guardar'. Posteriormente, se puede abrir y guardar cuantas veces sea necesario hasta que se complete la base.
7. La base de datos está lista para el ingreso de los datos. Es necesario guardar la información con cierta frecuencia mientras se ingresan los datos, es una buena costumbre para evitar pérdida de información.

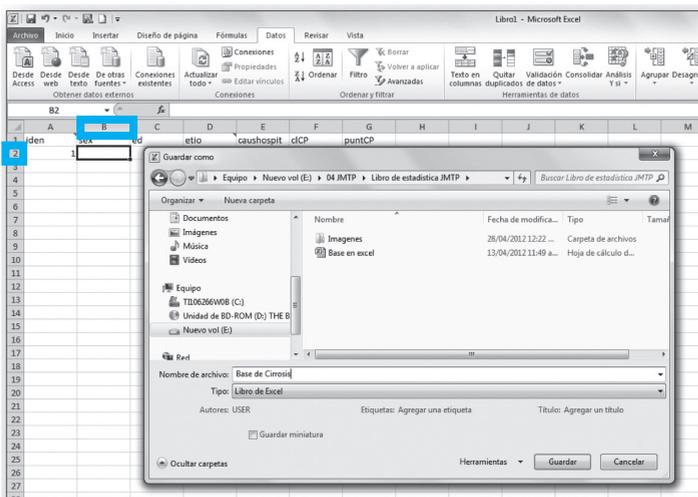


FIGURA A1.8. Guardar archivo de la Base de datos.

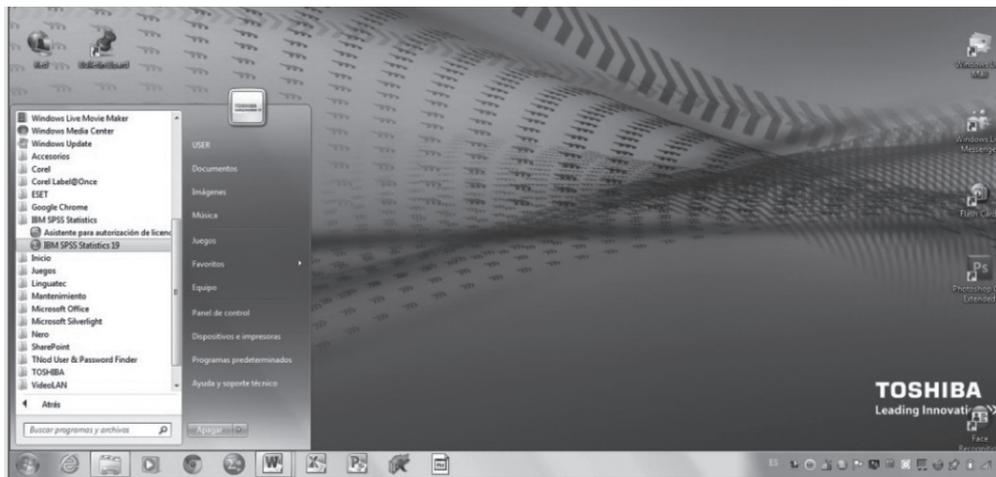


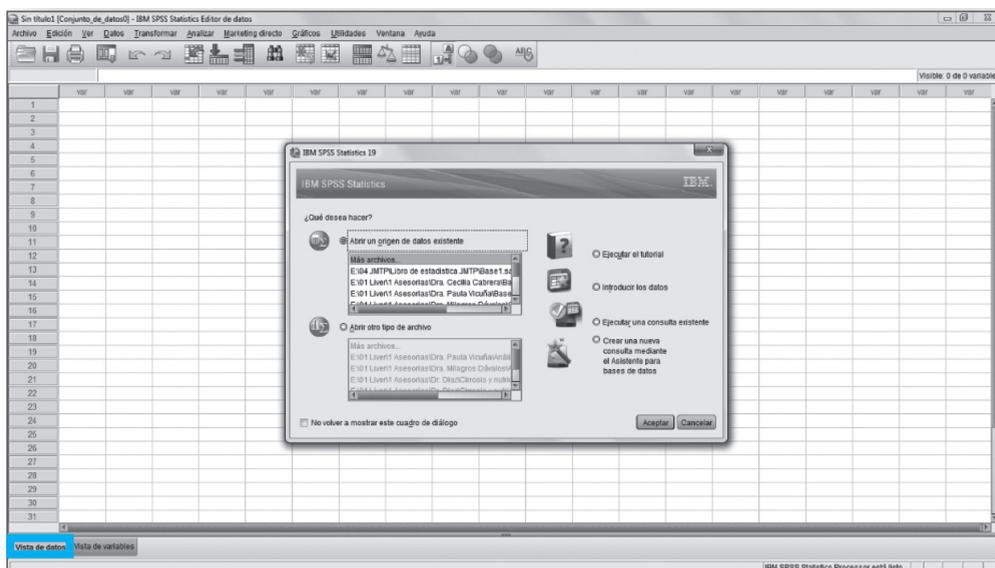
FIGURA A1.9. Ingreso a SPSS 23.

## Creación de una base de datos en SPSS

1. Abrir el programa SPSS: hacer clic en el ícono del programa.

El ícono del programa puede estar en la Barra de tareas o en el botón inicio en la carpeta IBM SPSS Statistics (Figura A1.9).

2. Se presenta la interfaz de SPSS 23 con un cuadro de diálogo (Figura A1.10), donde se presentan opciones que se deben elegir, de acuerdo a los requerimientos del usuario.



76 FIGURA A1.10. Interfaz de SPSS 19 y cuadro de diálogo inicial.

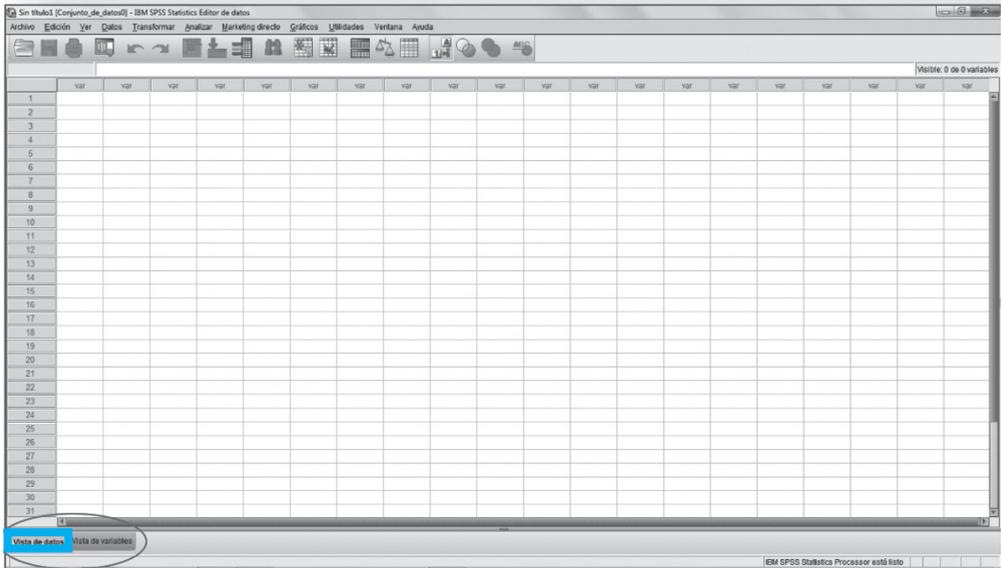


FIGURA A1.11. Editor de datos de SPSS 23: 'Vista de datos' activa.

3. Como se quiere crear una base en SPSS, se hace clic en la opción 'Introducir datos' o también sobre el botón 'cancelar'. El programa ofrece el interfaz o 'Editor de datos' (Figura A1.11).
4. En el extremo inferior izquierdo del 'Editor de datos', se observan dos etiquetas: 'Vista de datos', en la cual se pueden introducir los datos de los casos y 'Vista de variables', en la cual se crean las variables de la base de datos (Figura A1.11).

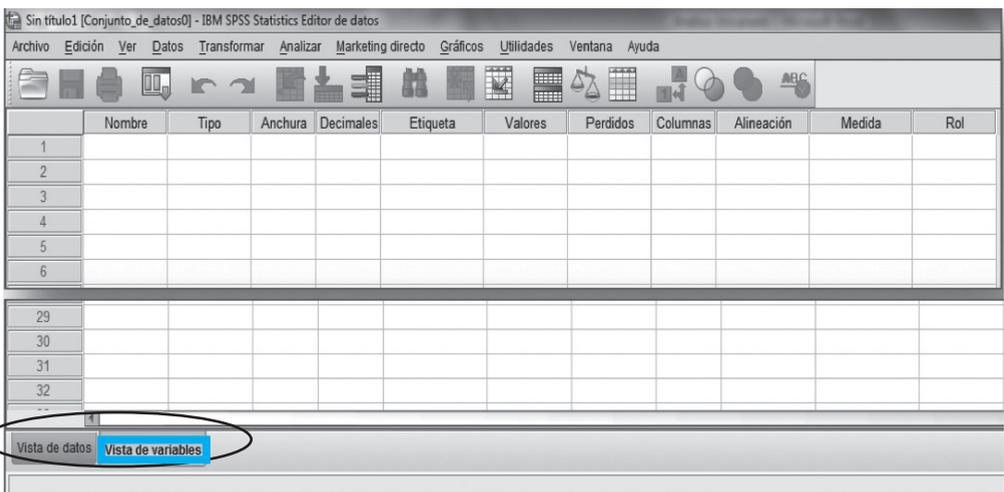


FIGURA A1.12. 'Vista de variables' activa.

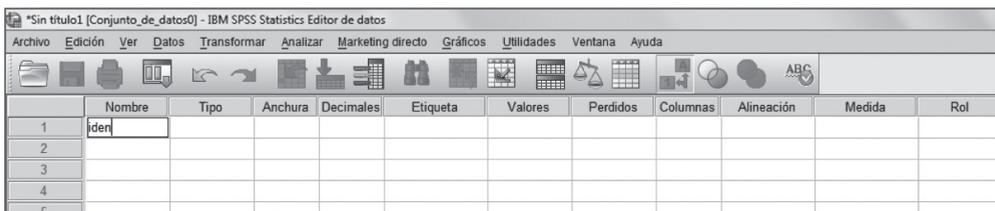


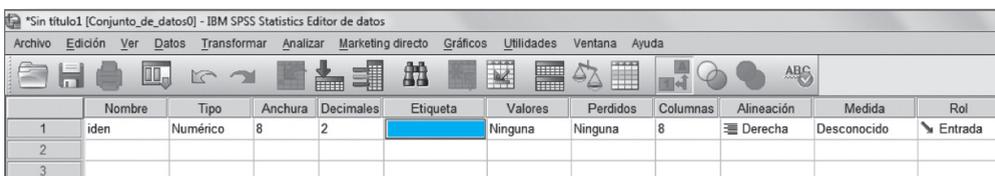
FIGURA A1.13. Introducción del nombre de la primera variable: 'iden'.

- Se hace clic en 'Vista de variables' (Figura A1.12) y se procede a la introducción de las características de las variables que se van a crear, para lo cual se usará el listado de variables:
- La primera variable para introducir es "iden", se refiere al número o código de identificación del caso. Se ingresan todas las características estipuladas de esta primera variable:
  - Se coloca el cursor en la primera celda de la fila 1, debajo de la columna 'Nombre' y se hace doble clic con el botón izquierdo del ratón y se procede a escribir el texto 'iden' (Figura A1.13).
  - Automáticamente el programa completa las otras características por defecto, excepto en Etiqueta (Figura A1.14).
  - Se procede a ajustar todas las características:
    - Tipo, al hacer clic en la celda 'Tipo' aparece un botón con puntos suspensivos (Figura A1.15).

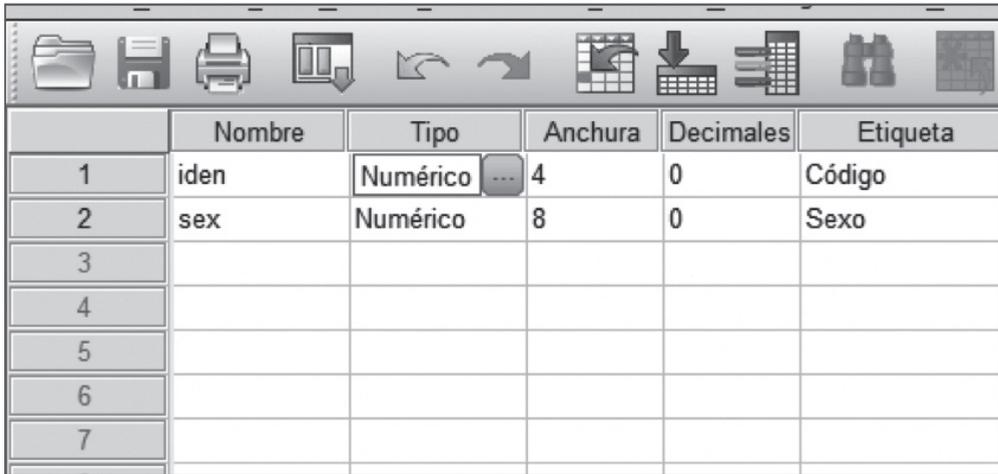
Se hace clic en el botón y aparece el cuadro de diálogo 'Tipo de variable', se elige la opción 'Numérico' y, luego, en 'Anchura', que se refiere al número de códigos que se acepta en la característica 'iden'. En este caso, si se van a incorporar unos 3 000 casos en la base, se tendría que especificar una anchura de 4 o 5, para que abarque todas las unidades (Figura A1.16).

En 'Cifras decimales', en este caso, debe ser números enteros, por lo que en decimales se indica '0'. Luego, se hace clic en el botón 'Aceptar'.

- En la celda 'Etiqueta', se escribe el texto 'Código'. Esta etiqueta permite que, en la 'Vista de datos', si se quiere saber el nombre completo de la variable, bastará con colocar el ratón sobre el nombre de la variable para que se pueda apreciar el nombre completo de la variable. El programa también considera esta etiqueta cuando presenta los resultados en el análisis univariado o multivariado que se realicen posteriormente.



78 FIGURA A1.14. Generación automática de las características de la primera variable.



	Nombre	Tipo	Anchura	Decimales	Etiqueta
1	iden	Numérico	4	0	Código
2	sex	Numérico	8	0	Sexo
3					
4					
5					
6					
7					

FIGURA A1.15. Celda 'Tipo' y visualización de botón con puntos suspensivos.

- En 'Valores', se deja la opción 'Ninguna', porque esta variable no tiene niveles ni categorías.
- En 'Perdidos', se deja la opción 'Ninguno'. En caso necesario, se le puede decir al programa que considere casos perdidos cuando se asignan los valores de '9', '99' o '999'.
- En 'Columna', se debe especificar el ancho que se desea que tenga la columna en la 'Vista de datos'.
- En 'Alineación', se puede elegir izquierda, derecha o centrado.

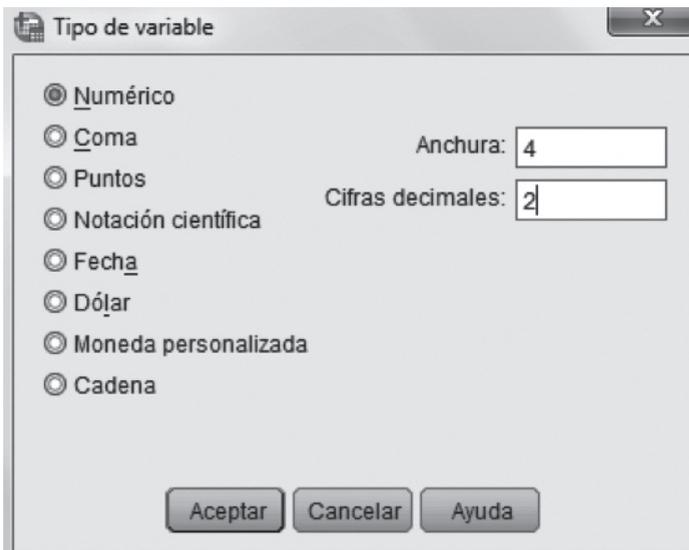
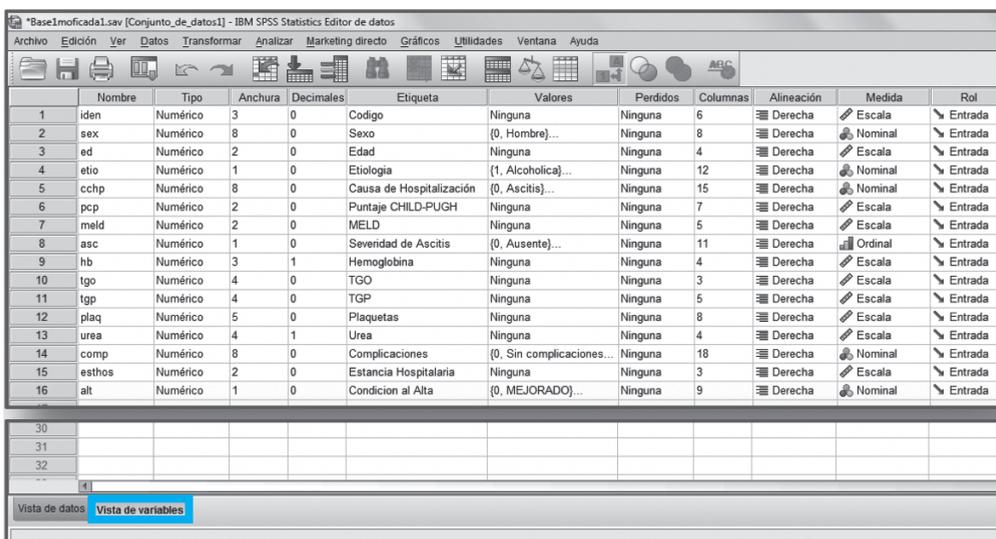


FIGURA A1.16. Cuadro de diálogo 'Tipo de variable'.



FIGURA A1.17. Lista plegable de 'Rol'.

- En 'Medida', se elige la opción 'Escalar', 'Ordinal' o 'Nominal', según el tipo de variable.
  - En 'Rol', se tiene las opciones de 'Entrada', 'Objetivo', 'Ambos', 'Ninguna', 'Partición' y 'Segmentación'. En este caso, se eligió 'Entrada'. Figura A1.17.
- Todo este proceso se repite para cada una de las variables y de acuerdo con la lista de variables que previamente se ha elaborado.
- Al final la 'Vista de Variables' queda como se visualiza en la Figura A1.18.



80 FIGURA A1.18. 'Vista de variables'.

	iden	sex	ed	etio	cchp	pcp	meld	asc
1	1	1	74	3	4	8	13	1
2	2	0	50	1	3	7	9	1
3	3	0	51	1	3	9	10	3
4	4	0	55	8	3	10	15	3
5								
6								
7								

FIGURA A1.19 'Vista de datos': ingreso de los datos hasta el cuarto paciente.

- Después se procede a ingresar los datos en 'Vista de datos' (Figura A1.19).
- El primer caso se trata de una paciente mujer de 74 años, con 'Etiología' VHC, cuya causa de hospitalización es hemorragia digestiva alta, que al ingreso presenta un puntaje de Child-Pugh de 8, etc.
- El programa tiene un comando 'Etiquetas de valor' que intercambia la forma numérica o en texto de las variables (Figura A1.20).

Al pasar el ratón sobre el nombre de la variable aparece el nombre completo de la variable en una etiqueta, en este caso se pasó el mouse por la variable 'cchp' y aparece una etiqueta 'Causa de Hospitalización', como se ve en la Figura A1.20.

La variable 'sexo' ya no presenta los números 1 y 2, sino el texto Femenino y Masculino. Lo mismo sucede con 'Etiología', 'Causa de Hospitalización' y las demás variables.

	iden	sex	ed	etio	cchp	pcp	meld	asc	h
1	1	Mujer	74	VHC	Infección Bacteri...	7	9	Leve	1
2	2	Hombre	50	Alcoholica	Hemorragia Digestiva ...	7	9	Leve	
3	3	Hombre	51	Alcoholica	Hemorragia Digestiva ...	9	10	Severa	
4	4	Hombre	55	No determinada	Hemorragia Digestiva ...	10	15	Severa	
5	5	Hombre	46	Alcoholica	Ascitis	9	6	Severa	1
6	6	Hombre	64	Alcoholica	Hemorragia Digestiva ...	11	21	Moderada	
7	7	Hombre	60	Alcoholica	Ascitis	11	14	Severa	
8	8	Hombre	61	No determinada	Encefalopatía	7	12	Ausente	

FIGURA A1.20. 'Vista de datos': en lugar de número el programa presenta en texto los valores asignados en cada variable.

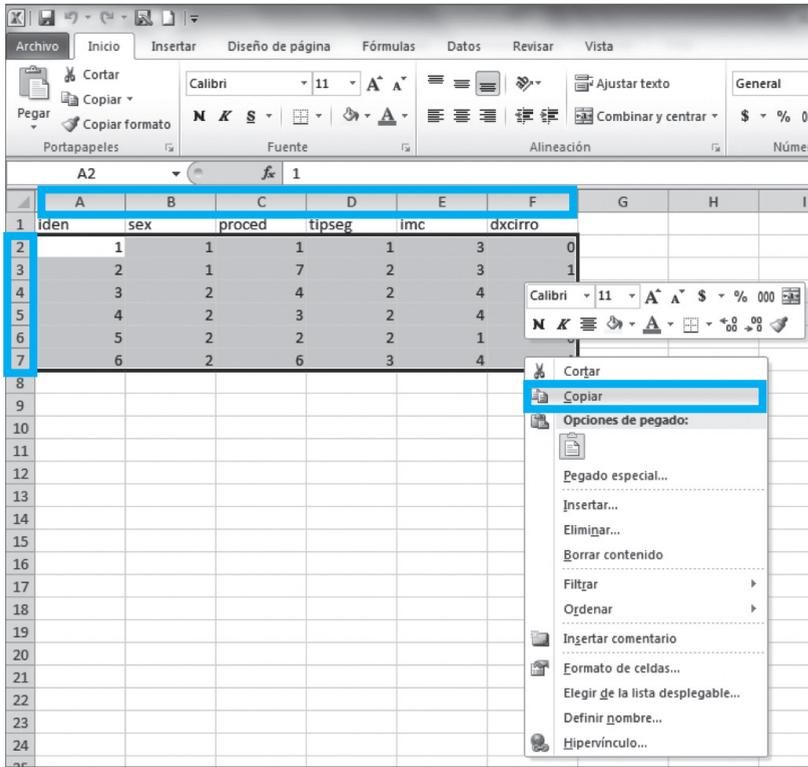
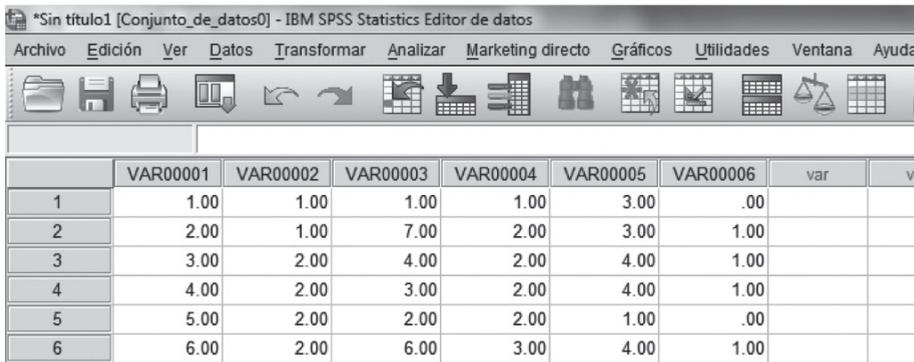


FIGURA A1.21. Base de datos en Excel.

### Importación de una base de datos diseñada en otro programa

En caso de que se tenga hecha la base de datos en una hoja de cálculo u otro programa, se selecciona todos los casos y se invoca la orden 'Copiar' (Figura A1.21). Luego, en SPSS, se coloca el cursor en la primera celda de la primera fila y se procede a 'Pegar'. Automáticamente se generan variables sin nombres y se copian los valores de todos los casos (Figura A1.22).



82 FIGURA A1.22. Aspecto del visor 'Vista de datos' en SPSS después de haber pegado la base de datos de Excel.

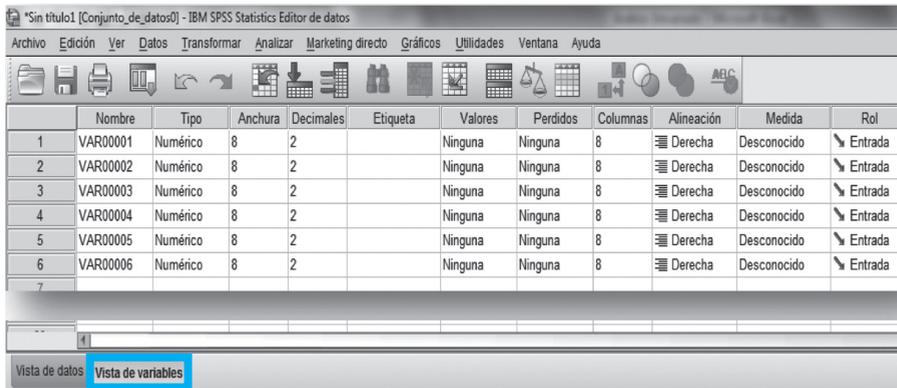


FIGURA A1.23. Aspecto del visor 'Vista de variables' en SPSS.

En la 'Vista de variables' se procede a especificar las características de cada una de las variables (Figura A1.23).

También se puede abrir directamente una base de datos hecha en una hoja de cálculo con la utilización del comando 'Archivo' >>> 'Abrir' >>> 'Datos' (Figura A1.24).

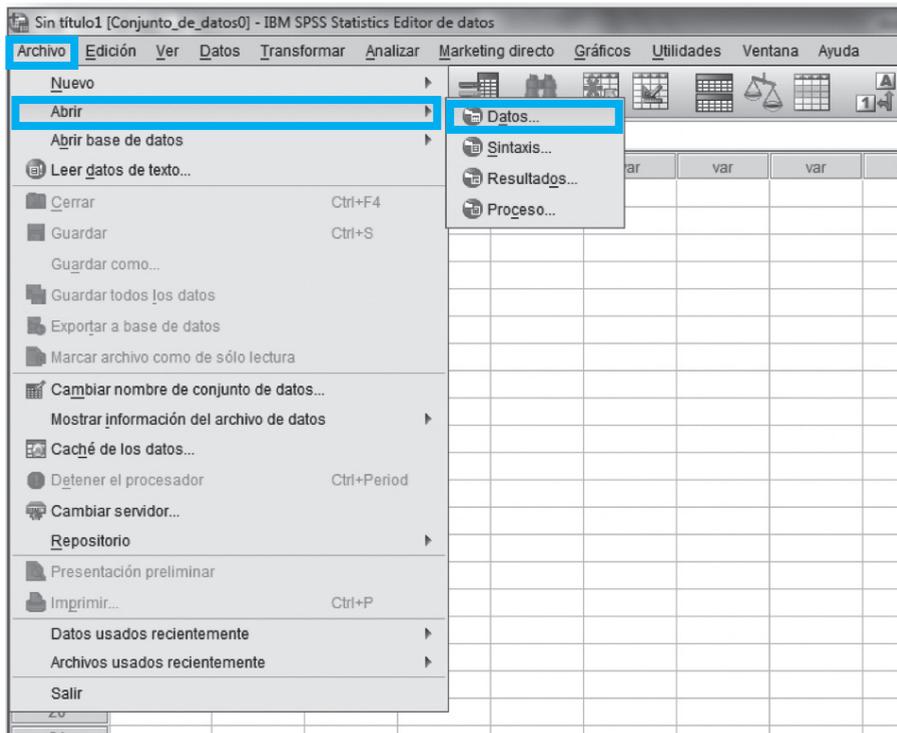


FIGURA A1.24. Lista del menú 'Archivo' de SPSS.

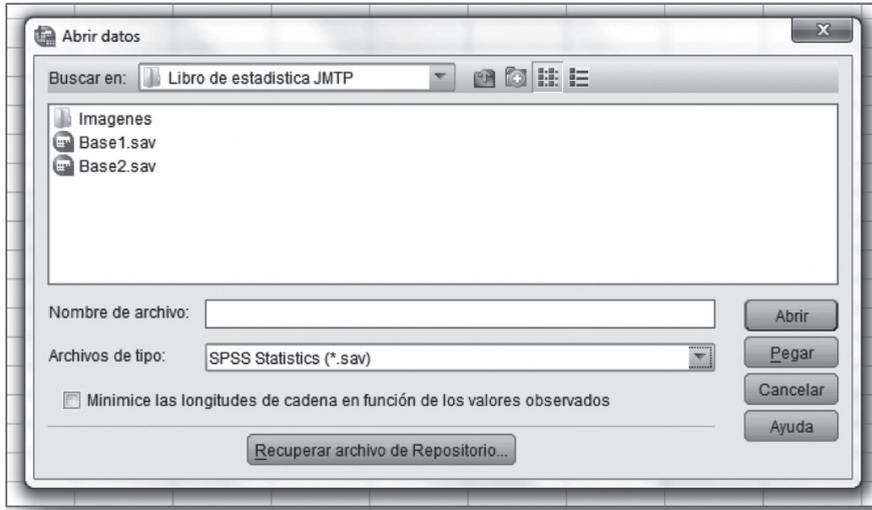
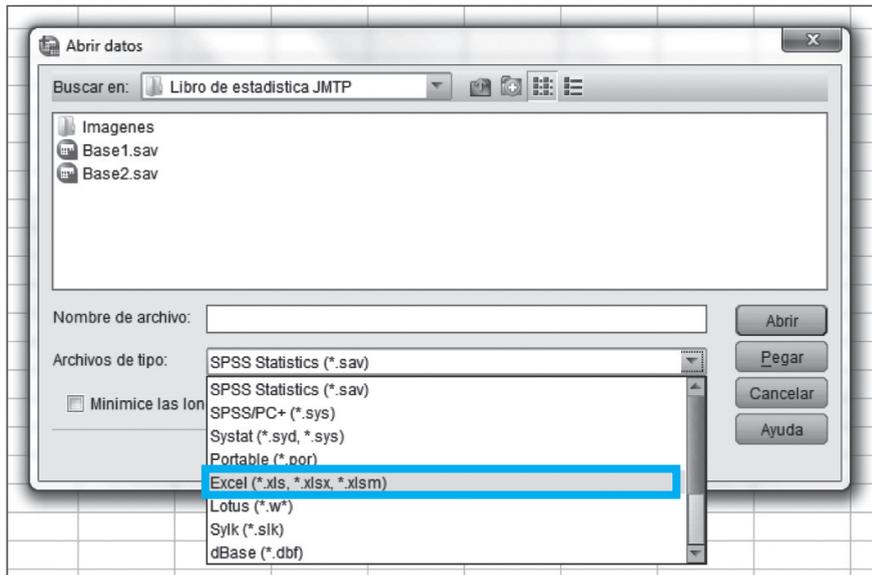


FIGURA A1.25. Cuadro de diálogo 'Abrir datos' del menú 'Archivo' de SPSS.

El programa presenta un cuadro de diálogo (Figura A1.25) en el que se direcciona en 'Buscar' en la ubicación en que se encuentra el archivo que se desea abrir. Luego, en la pestaña de 'Archivos de tipo', se procede a elegir la extensión que tiene el archivo, en el ejemplo la Base fue elaborada en Excel y tiene una extensión '\*.xlsx' (Figura A1.26).

Luego, se selecciona el archivo Excel que ya se visualiza en la lista de archivos para que el programa automáticamente lo incluya en 'Nombre de archivo' (Figura A1.27).



84 FIGURA A1.26. Seleccionar 'tipo de archivo' a importar en 'Abrir datos'.

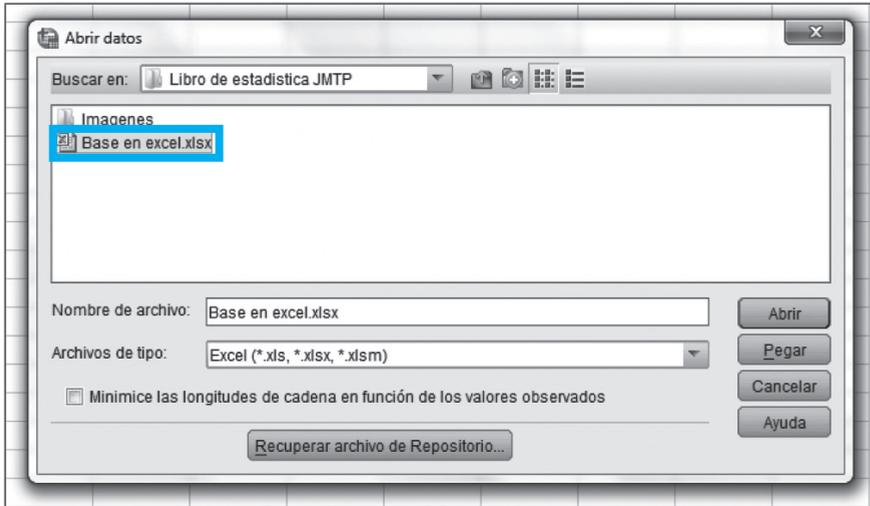


FIGURA A1.27. Cuadro de diálogo 'Abrir datos' del menú 'Archivo' de SPSS.

A continuación, se hace clic en el botón 'Abrir', el programa presenta un cuadro de diálogo para completar la información y en el caso solo se debe hacer clic en el botón 'Aceptar'.

Esto lleva al Cuadro de diálogo 'Apertura' de origen de datos Excel, en el cual se puede elegir la hoja de trabajo que se quiere importar.

Finalmente, se presiona el botón 'Aceptar' (Figura A1.28).

El programa genera la base de datos e incluye los nombres de las variables en la 'Vista de Datos'. Se debe seleccionar la 'Vista de Variables' para realizar los cambios pertinentes de las características de las variables (Figura A1.29).

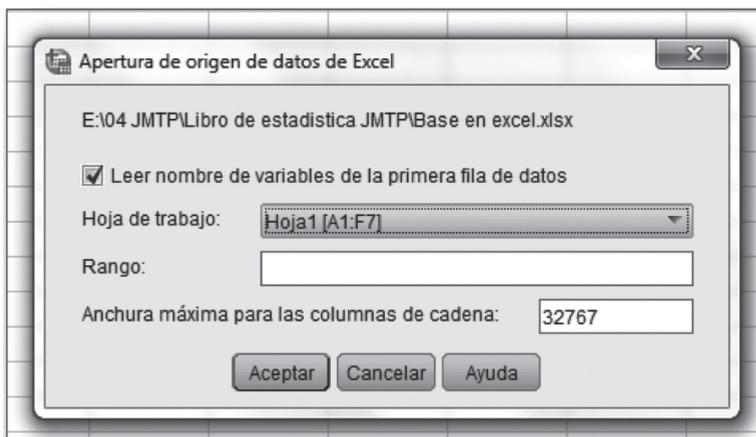


FIGURA A1.28. Cuadro de diálogo 'Apertura de origen de datos de Excel'.

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	iden	Numérico	12	0		Ninguna	Ninguna	12	Derecha	Nominal	Entrada
2	sex	Numérico	12	0		Ninguna	Ninguna	12	Derecha	Nominal	Entrada
3	proced	Numérico	12	0		Ninguna	Ninguna	12	Derecha	Nominal	Entrada
4	tipseg	Numérico	12	0		Ninguna	Ninguna	12	Derecha	Nominal	Entrada
5	imc	Numérico	12	0		Ninguna	Ninguna	12	Derecha	Nominal	Entrada
6	dxcirro	Numérico	12	0		Ninguna	Ninguna	12	Derecha	Nominal	Entrada
7											

FIGURA A1.29. 'Vista de variables' de SPSS.

### Importación de una base de datos mediante la utilización del Asistente

1. Para esto, en el Menú 'Archivo' de SPSS, se selecciona el comando 'Abrir base de datos' > 'Nueva consulta' (Figura A1.30).
2. El programa presenta el cuadro de diálogo del Asistente para base de datos (Figura A1.31).  
Se procede a realizar todos los pasos del asistente para la importación de base de datos diseñados en otros programas.

	imc	dxcirro	var
1	3	0	
2	3	1	
4	2	4	1
3	2	4	1
2	2	1	0
6	3	4	1

FIGURA A1.30. Lista de menú 'Archivo' de SPSS, para abrir directamente una base de datos hecha en otro software.

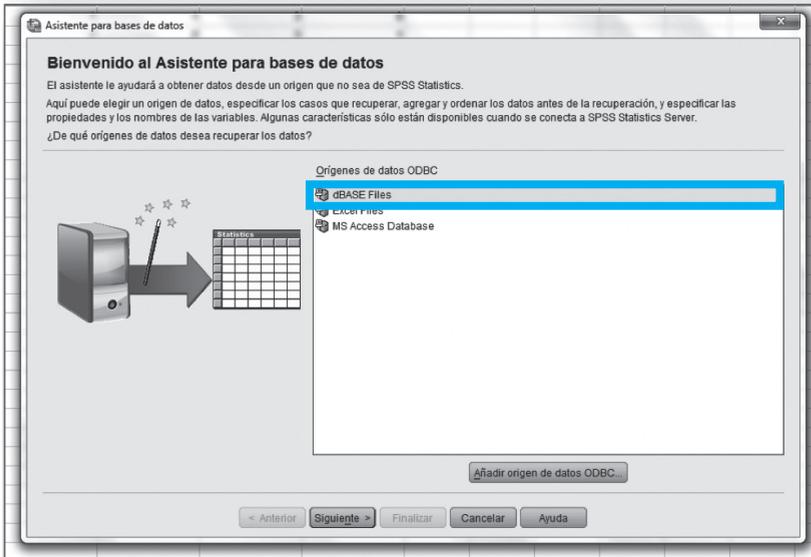


FIGURA A1.31. 'Asistente para base de datos' del SPSS, para abrir directamente una base de datos hecha en otro software.

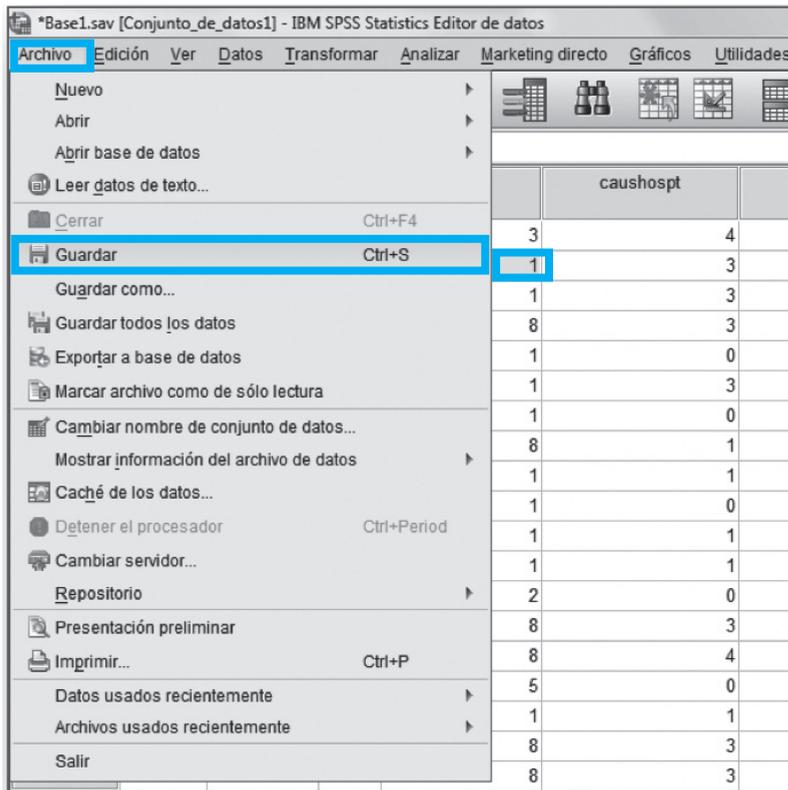


FIGURA A1.32. Menú 'Archivo' y comando 'Guardar' de SPSS.

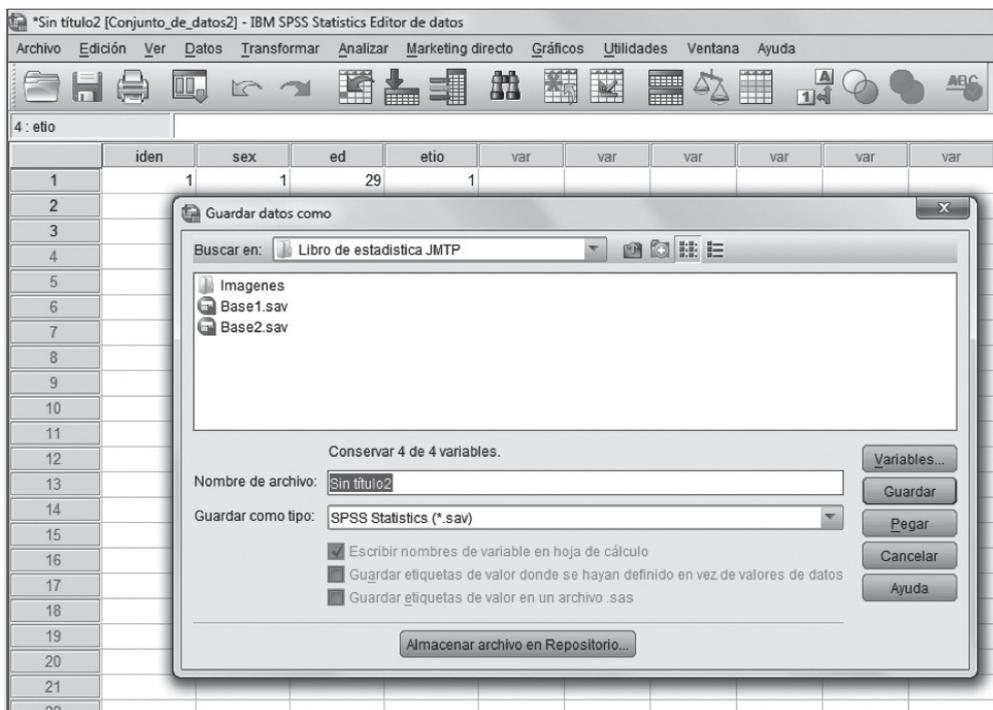


FIGURA A1.33. Cuadro de diálogo 'Guardar datos como' de SPSS.

## Guardar la base de datos en SPSS

En el proceso de editar una base de datos, frecuentemente se estará abriendo y guardando una base de datos hasta que se complete el ingreso de todos los datos.

1. En SPSS, los comandos para 'Guardar' un archivo son semejantes a los de Excel que se han presentado (Figuras A1.32 y A1.33).
2. Se direcciona hacia el lugar donde se quiere que se guarde el archivo y, luego, en 'Nombre de archivo', se da el nombre de correspondencia. Finalmente, se hace clic en el botón 'Guardar'.

## Análisis de variable cuantitativa

1. Se debe tener presente que SPSS trata a las variables cuantitativas discretas o continuas como variables con medida escalar. Para el ejemplo, se utilizará la variable 'Edad' de la base de datos que está en el archivo base1.sav y se indica a SPSS los siguientes comandos:

*Analizar >>> Estadísticos Descriptivos >>> Frecuencias* (Figura A2-1).

2. En la ventana de diálogo 'Frecuencias', se selecciona la variable 'Edad'; luego, se hace clic en el botón 'Frecuencia Estadísticos' y en 'Valores percentiles' se activan 'Cuartiles', 'Puntos de corte' (y se añade '10' grupos iguales) y 'Percentiles (y se van añadiendo 5, 25

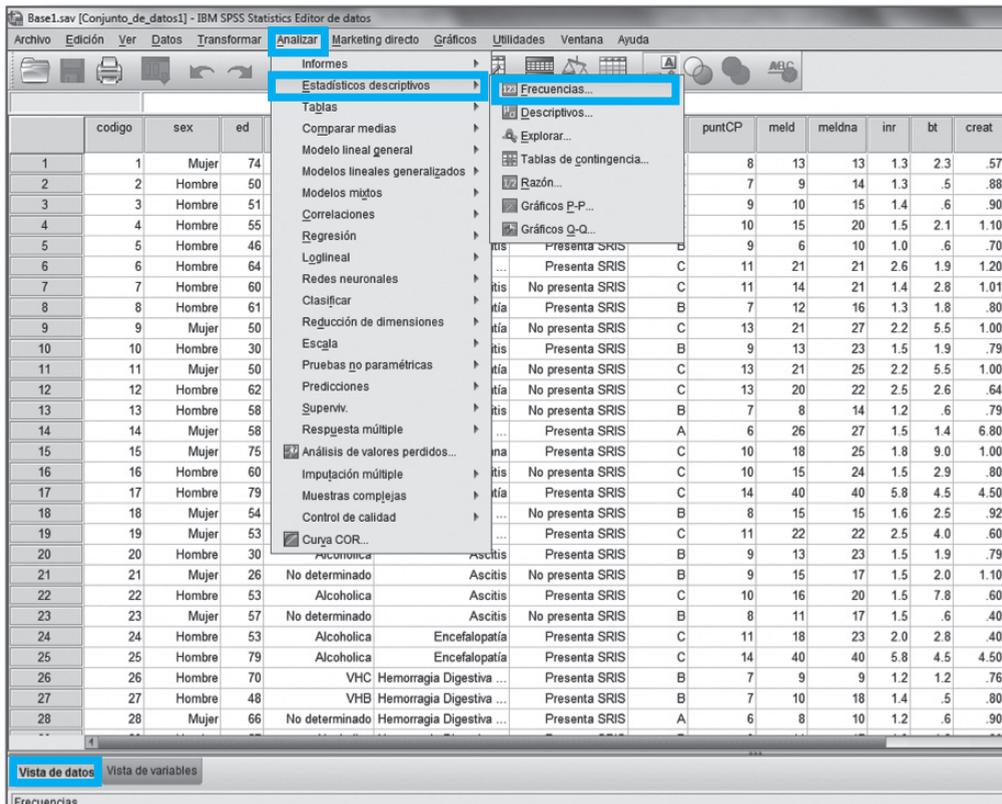


FIGURA A2.1. Menú 'Analizar' y comando 'Estadísticos descriptivos', 'Frecuencias' de SPSS.

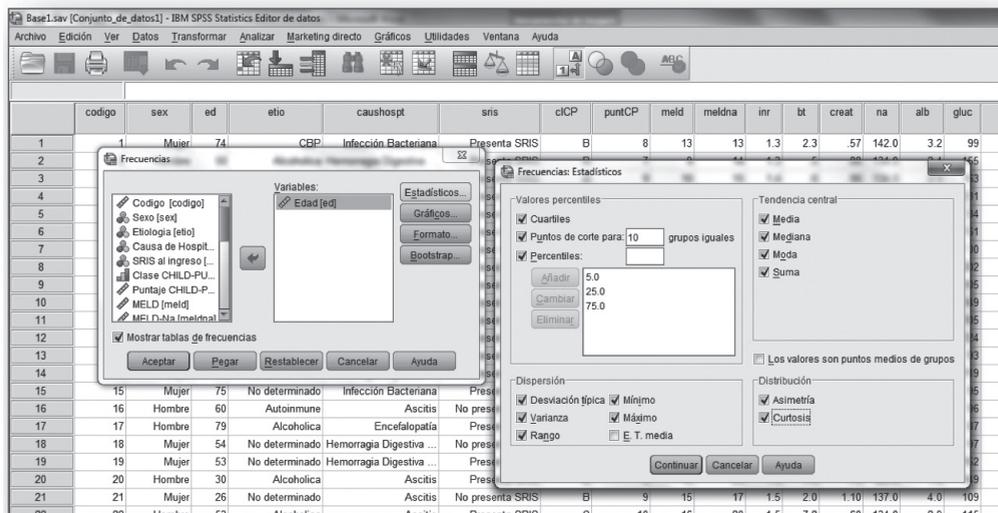


FIGURA A2.2. Cuadro de diálogo 'Frecuencias', seleccionar 'Variables'. Cuadro de diálogo 'Frecuencias: Estadísticos' seleccionar los pertinentes.

y 75). En 'Tendencia central', se activan 'Media', 'Mediana', 'Moda' y 'Suma'; en 'Dispersión', 'Desviación típica', 'Varianza', 'Rango', 'Mínimo' y 'Máximo'; en 'Distribución', 'Asimetría' y 'Curtosis'. Finalmente, se hace clic en el botón 'Continuar' (Figura A2-2 y Figura A2-3).

- Posteriormente, se hace clic en botón 'Gráficos'. En cuadro de diálogo 'Frecuencias: Gráficos', se activa 'Histogramas' y 'Mostrar curva normal en el histograma' (Figura A2-2 y Figura A2-3).

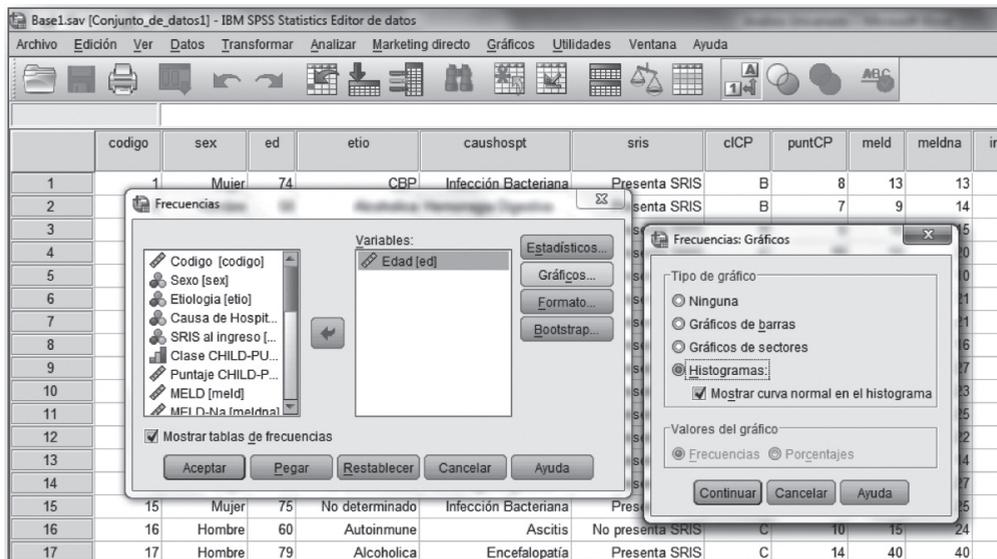


FIGURA A2.3. Cuadro de diálogo 'Frecuencias'. Cuadro de diálogo 'Frecuencias: Gráficos', seleccionar 'Histogramas' y 'Mostrar curva normal en el histograma'.

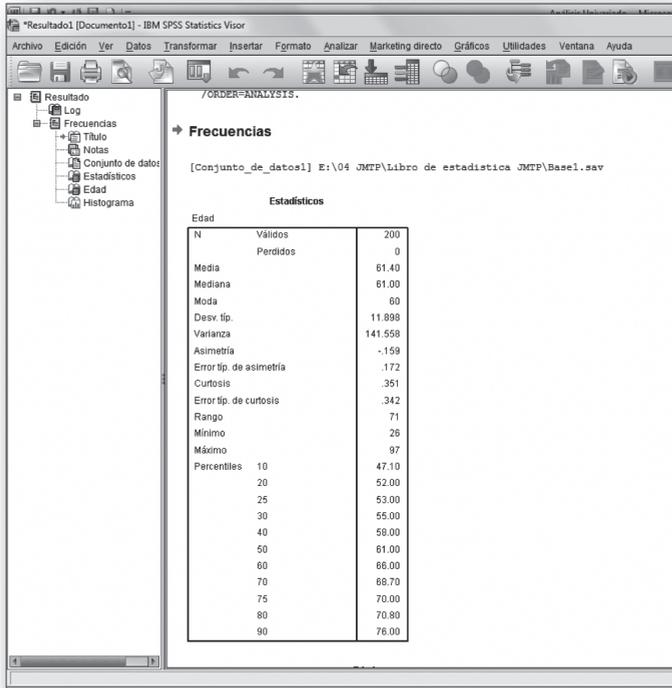


FIGURA A2.4. Editor de resultados: Estadísticos solicitados a SPSS.

Finalmente, se hace clic en botón 'Continuar'. El programa presenta la ventana Visor los resultados (Figuras A2.4 y A2.5).

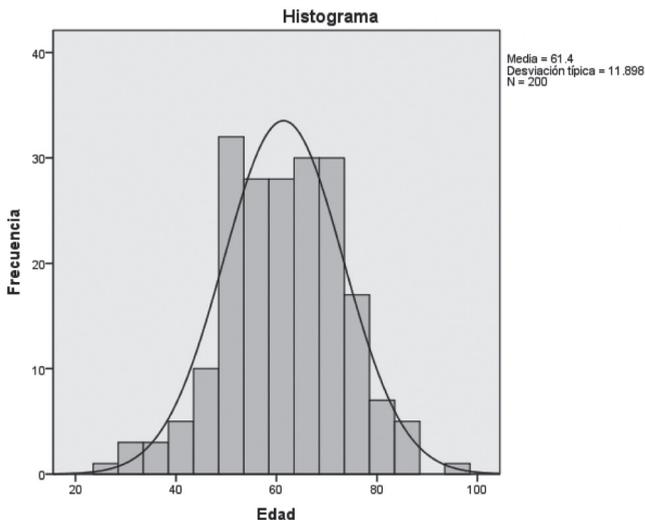
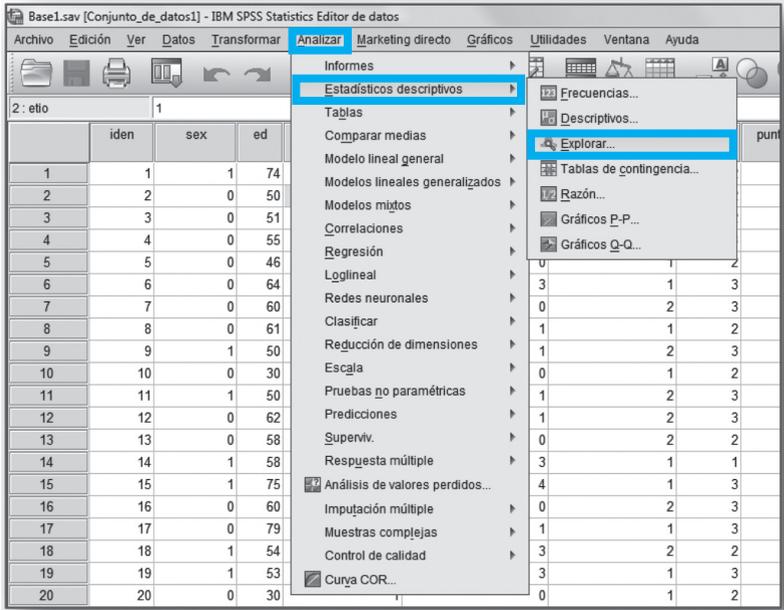


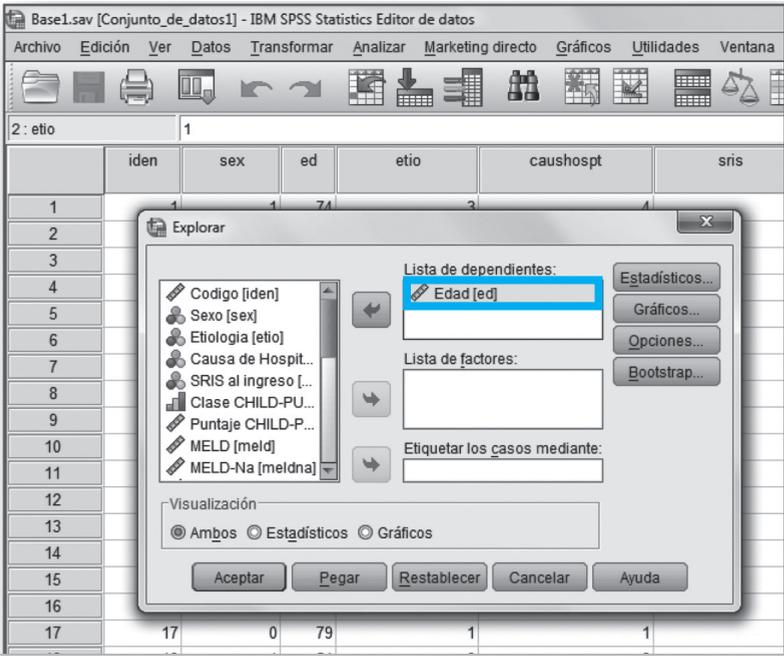
FIGURA A2.5. Editor de resultados: gráfica del histograma de frecuencias con curva normal.



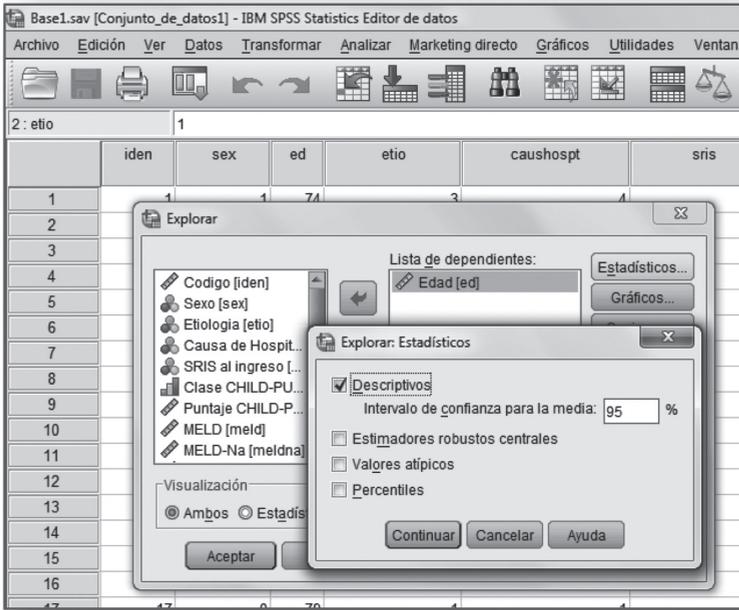
**FIGURA A2.6.** Menú 'Analyze', 'Estadísticos descriptivos' comando 'Explorar'.

4. Faltan los 'Intervalos de confianza' de la media y otros gráficos como 'Tallos y hojas' y 'Caja y bigotes'. Para esto, en el menú 'Analyze' (Figura A2.6).

*Analyze >>> Estadísticos Descriptivos >>> Explorar*

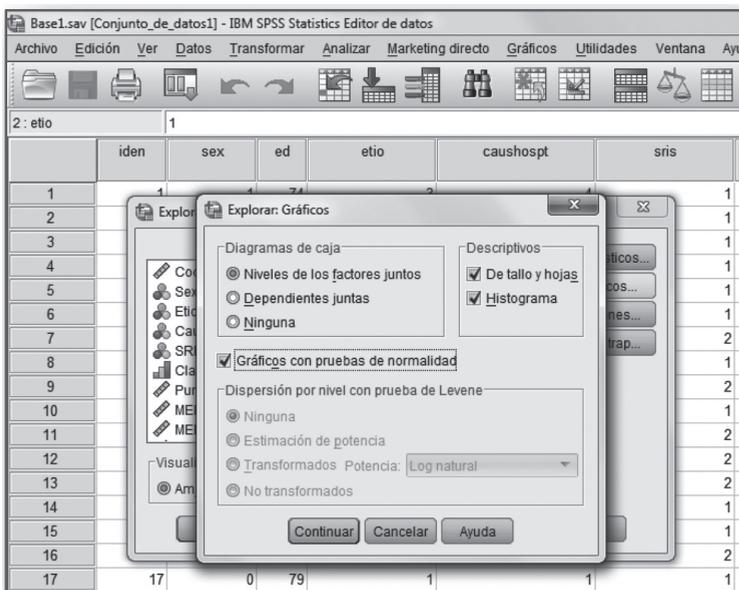


**FIGURA A2.7.** Cuadro de diálogo 'Explorar'. Pasar la variable 'edad' a 'Lista de dependientes'.

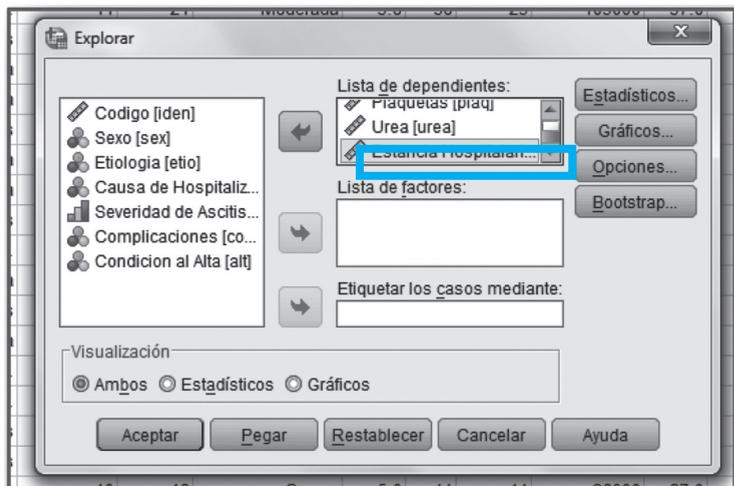


**FIGURA A2.8.** En 'Explorar', 'Explorar estadísticos'. Activar 'Descriptivos' y en 'Intervalo de confianza para la media' introducir '95'.

5. El programa presenta el cuadro de diálogo 'Explorar' (Figuras A2.7 y A2.8), en el cual se selecciona la variable 'edad' y se agrega a 'Lista de dependientes'. Se hace clic en el botón 'Estadísticos' y se conserva lo que por defecto presenta SPSS. Se hace clic en el botón 'Continuar'.



**FIGURA A2.9.** Explorar 'Explorar Gráficos'.



**FIGURA A2.10:** En 'Explorar', en 'Lista de dependientes', se observa que se han pasado todas las variables cuantitativas, para realizar de una sola vez el análisis de todas ellas.

6. Nuevamente, en el cuadro de diálogo 'Explorar', se hace clic en el botón 'Gráficos' y se seleccionan: en 'Diagrama de caja', se deja activa 'Niveles de los factores juntos'; en 'Descriptivos', se seleccionan 'De tallo y hojas' e 'Histograma'. Además, se activa 'Gráficos con pruebas de normalidad' (Figura A2.9). Se hace clic en 'Continuar'; luego, en 'Aceptar' y el programa presenta los 'Gráficos con pruebas de normalidad', que se proceden a copiar en el procesador de texto.
7. Para no tener que repetir los pasos, variable por variable, se pueden pasar todas las variables cuantitativas en el cuadro de diálogo 'Frecuencias' y 'Explorar' que se han utilizado y pedir todos los estadísticos que se necesitan para el análisis (Figura A2-10).
8. El programa presenta los siguientes resultados y gráficas (Figuras A2-11 hasta la Figura A2.16).

			Estadístico	Error t <sub>p</sub> .
Edad	Media		61.40	.841
	Intervalo de confianza para la media al 95%	Límite inferior	59.74	
		Límite superior	63.06	
	Media recortada al 5%		61.54	
	Mediana		61.00	
	Varianza		141.558	
	Desv. t <sub>p</sub> .		11.898	
	Mínimo		26	
	Máximo		97	
	Rango		71	
	Amplitud intercuartil		17	
	Asimetría		-.159	.172
	Curtosis		.351	.342

**FIGURA A2.11.** Resultados de variable 'edad' en 'Editor de Resultados'.

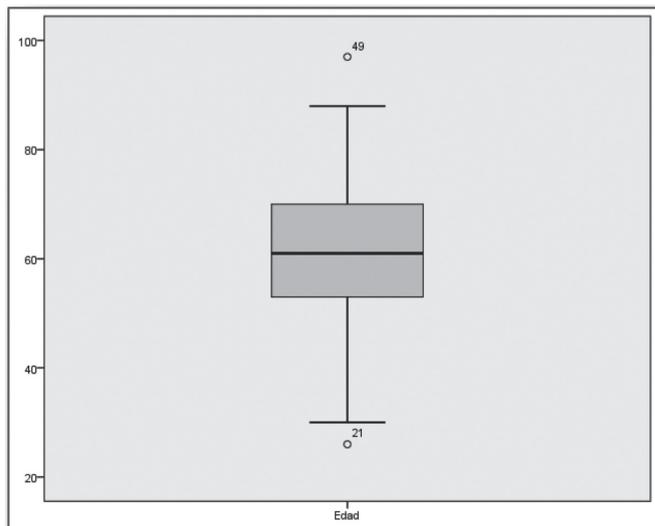


FIGURA A2.12. Gráfica de caja y bigotes en el 'Editor de Resultados'.

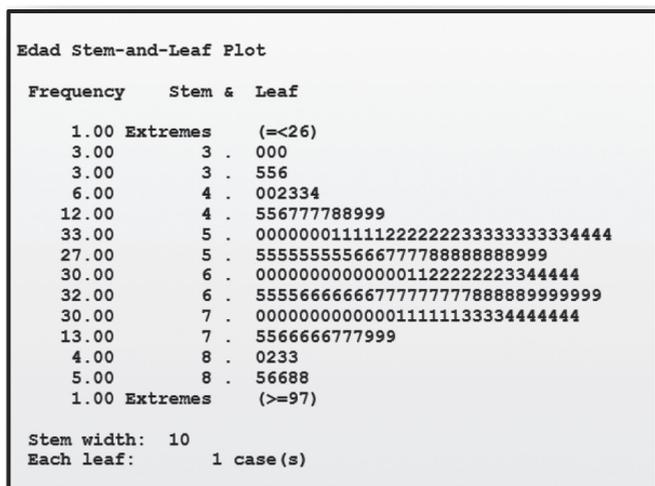


FIGURA A2.13. Gráfica de tallo y hojas en el 'Editor de Resultados'.

Pruebas de normalidad						
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Edad	.055	200	.200*	.991	200	.222

a. Corrección de la significación de Lilliefors

\*. Este es un límite inferior de la significación verdadera.

FIGURA A2.14. Prueba de normalidad 'Editor de Resultados'.

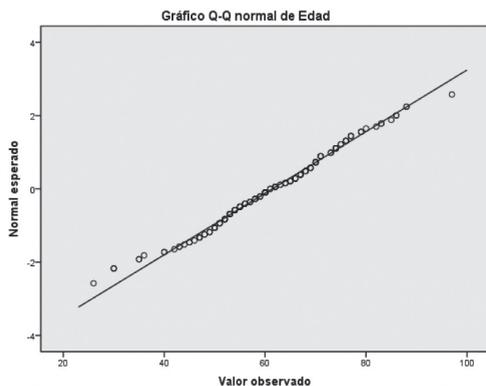


FIGURA A2.15. Gráfico Q-Q normal de la variable 'edad' en 'Editor de Resultados'.

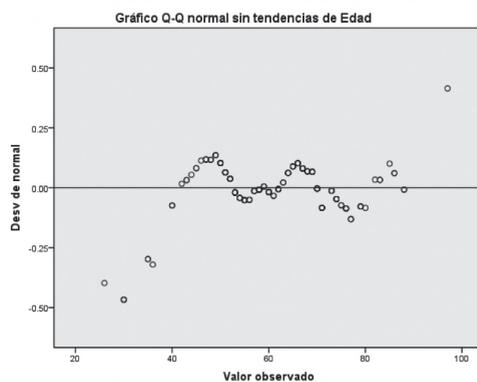


FIGURA A2.16. Gráfico Q-Q normal sin tendencias de la variable 'edad' en 'Editor de Resultados'.

- De toda esta información, se identifica y selecciona la necesaria para elaborar las Tablas o Gráficos que se utilizarán en el informe de la variable en estudio.

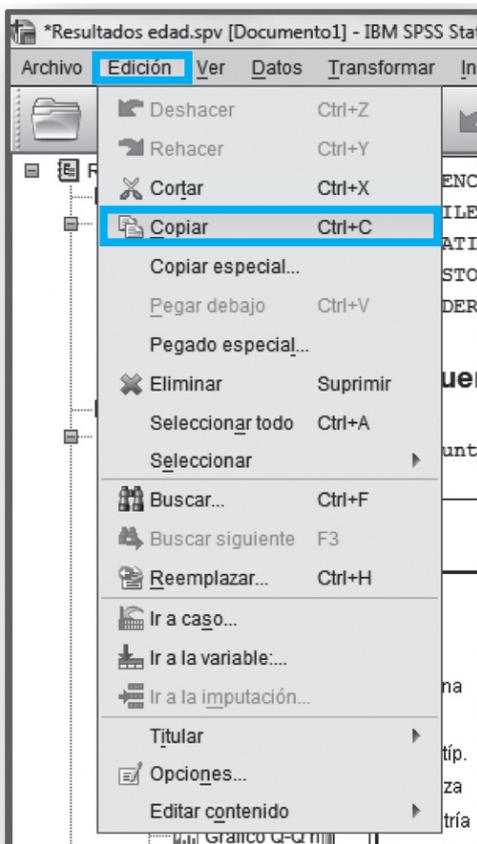


FIGURA A2.17. Menú 'Edición de Editor resultados'.

10. Insertar las tablas y gráficos al procesador de texto. Para hacerlo se tienen que dar los siguientes comandos:

- Seleccionar la Tabla o Gráfico que se desea copiar, directamente se puede teclear simultáneamente las teclas 'Ctrl' + 'C'.
- El programa genera una copia que está en la memoria de trabajo de la computadora. Se va al procesador de texto y se ubica el cursor en el lugar que se desea que se incluya la copia; se tecléa 'Ctrl' + 'V', y aparece en el procesador de texto la copia. También se puede hacer utilizando los comandos 'Edición' (Figura A2-17).



## Análisis de variable ordinal

1. Para este análisis, se utilizará la variable 'gradca' ('grado de extensión del cáncer') de la base de datos que está en el archivo base2.sav y se indica a SPSS los siguientes comandos (Figura A3.1):  
'Analizar' >>> 'Estadísticos Descriptivos' >>> 'Frecuencias'
2. En la ventana de diálogo 'Frecuencias' (Figura A3.1 y Figura A3.2), se selecciona la variable 'gradca'. Luego, se hace clic en el botón 'Estadísticos' y se activa solo en 'Dispersión: Mínimo y Máximo', y en 'Tendencia central: Mediana y Moda'. Se hace clic en botón 'Continuar'.
3. Se hace clic en el botón 'Frecuencia: Gráficos' y en 'Tipo de gráfico', se activa 'Gráficos de barras' y en 'Valores del gráfico, 'Porcentajes' Figura A3.3. Se hace clic en botón 'Continuar'. Finalmente, en el cuadro de diálogo 'Frecuencias: Gráficos', se hace clic en el botón 'Aceptar'.
4. Finalmente, se hace clic en el botón 'Continuar'. El programa presenta la ventana 'Editor resultados' (Figura A3.4).
5. En la Gráfica de Grado de extensión del cáncer (Figura A3.4C), se observa que las barras están separadas. Sin embargo, por tratarse de una variable ordinal, estas barras deberían

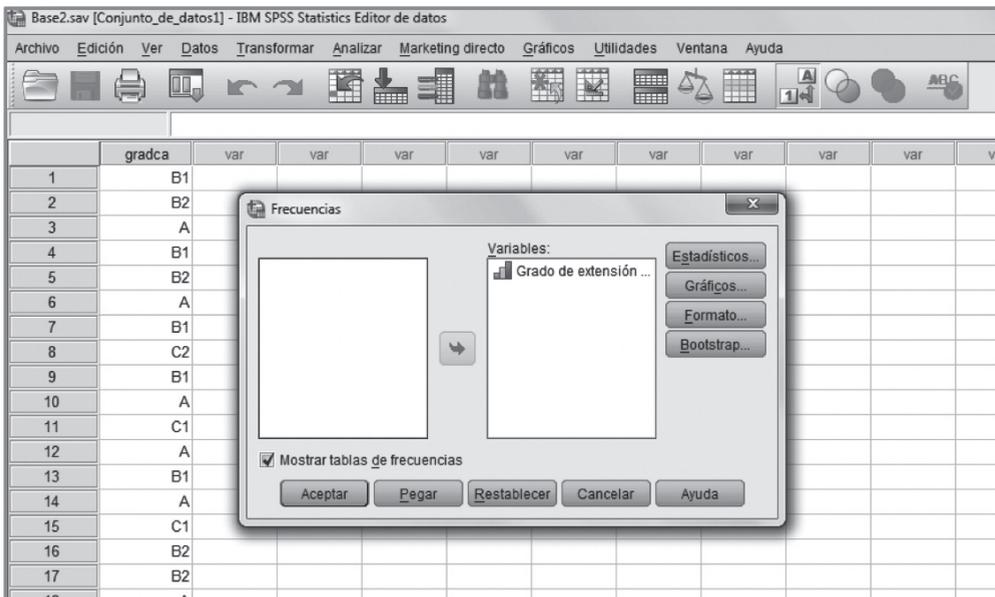


FIGURA A3.1. Cuadro de diálogo 'Frecuencias'.

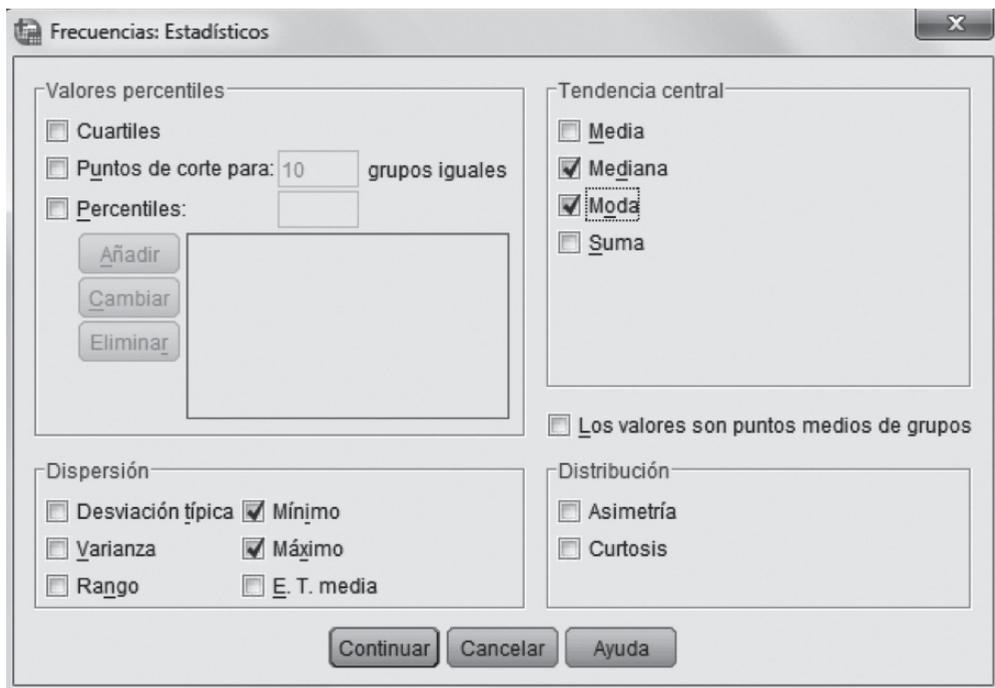


FIGURA A3.2: Cuadro de diálogo 'Frecuencias: Estadísticos'.

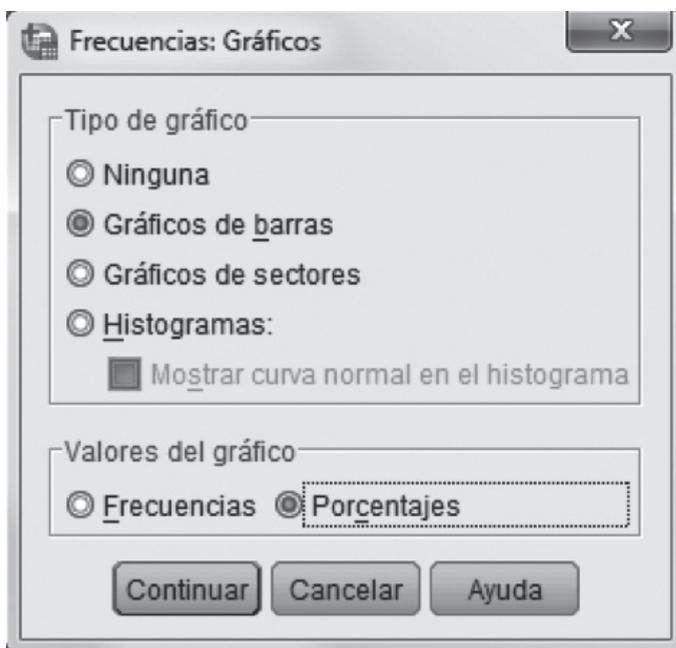


FIGURA A3.3. Cuadro de diálogo 'Frecuencias: Estadísticos'.

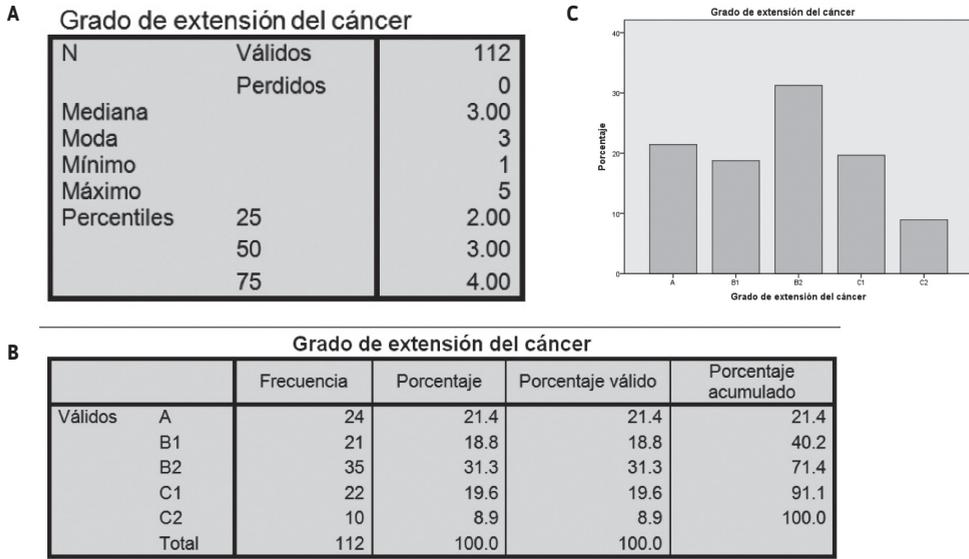


FIGURA A3.4. A) Cuadro de 'Estadísticos Descriptivos'. B) Cuadro de 'Estadísticos Distribución de frecuencias'. C) Distribución de Frecuencias.

tener continuidad, por lo que se modificará la presentación del gráfico. Para esto, en 'Visor de resultados' del SPSS, se hará doble clic sobre la gráfica, para invocar al 'Editor de Gráficos' (Figura A3.5).

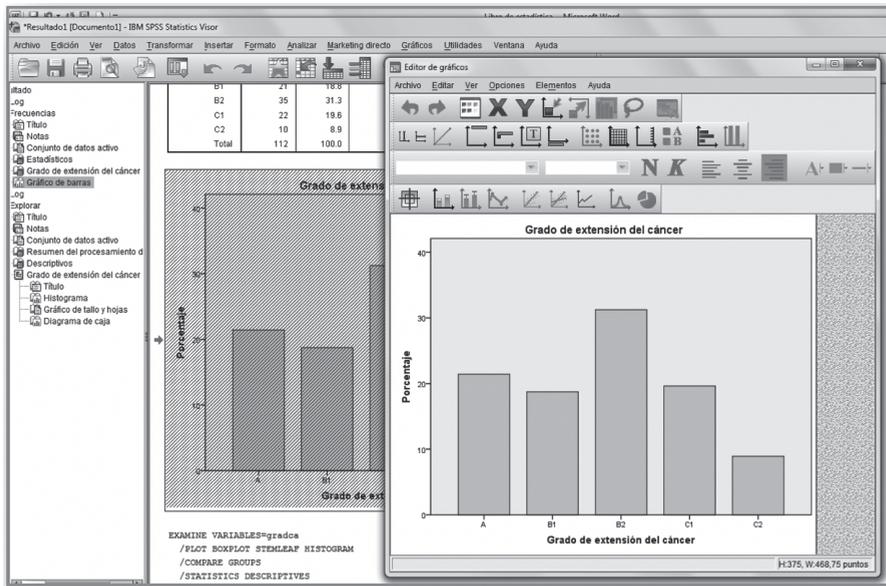


FIGURA A3.5. 'Editor de gráficos': modificar algunas características del gráfico.

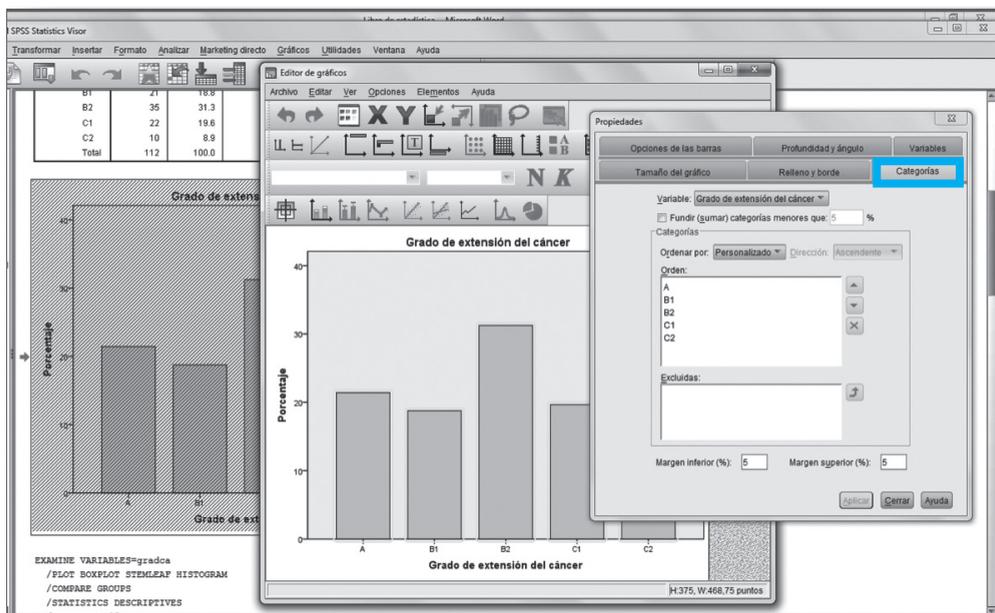


FIGURA A3.6. 'Editor de gráficos': 'Propiedades'.

6. Se señalan las barras y se hace doble clic, para invocar el cuadro de diálogo 'Propiedades' (Figura A3.6).
7. En el cuadro de diálogo 'Propiedades', se selecciona 'Opciones de las barras' (Figura A3.7). En el ítem 'Anchura', con el ratón, arrastrar el botón de 'Barras' hasta 100, luego hacer clic en el botón Aplicar, luego clic en el botón de Cerrar la ventana del Editor de Gráficos.



FIGURA A3.7. Propiedades: Modificar anchura de las barras (100%).

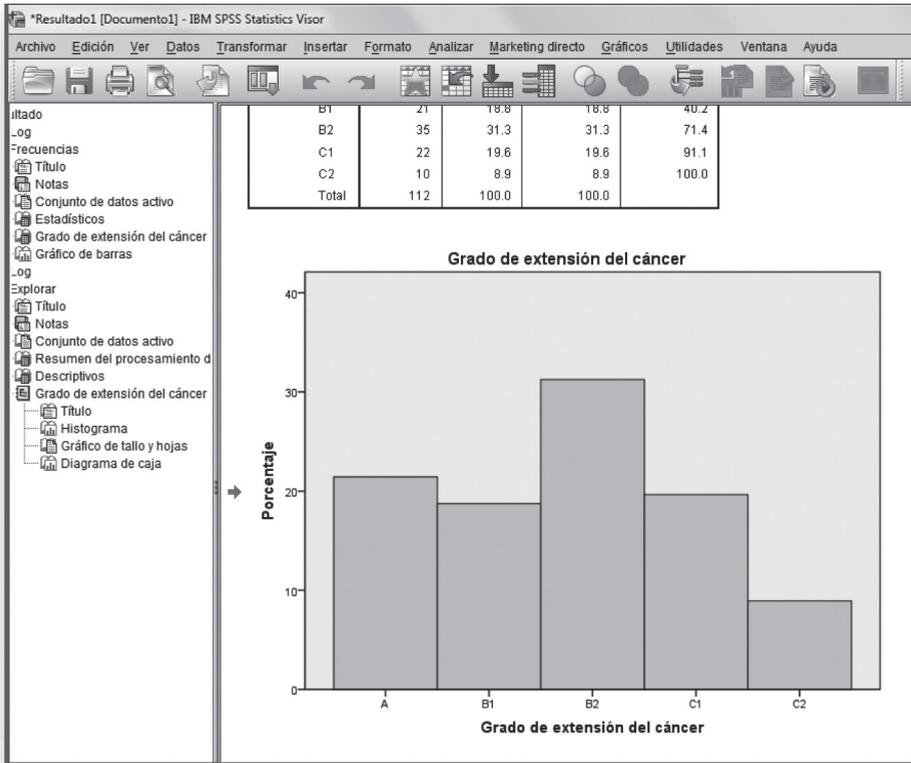


FIGURA A3.8. Vista final del 'Editor de resultados'.

8. SPSS presenta la gráfica en el 'Editor resultados' (Figura A3.8).
9. Para que SPSS diseñe la gráfica Caja y bigotes, se debe ir nuevamente a Menú 'Analizar', y se selecciona el comando 'Explorar' (Figura A3.9).

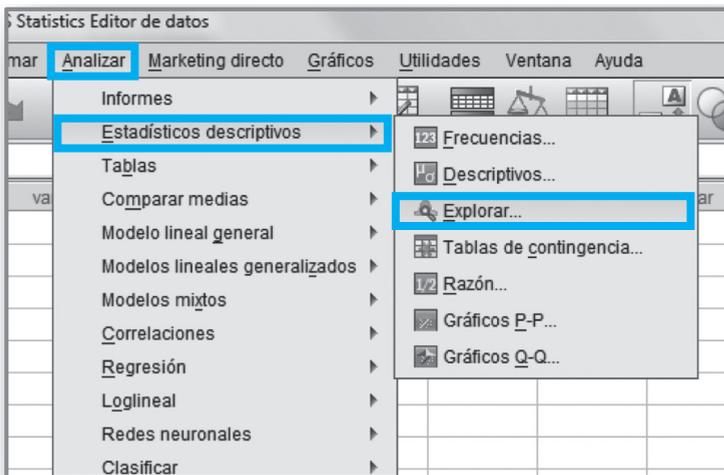


FIGURA A3.9. Lista de menú 'Analizar': 'Estadísticos descriptivos', 'Explorar'.

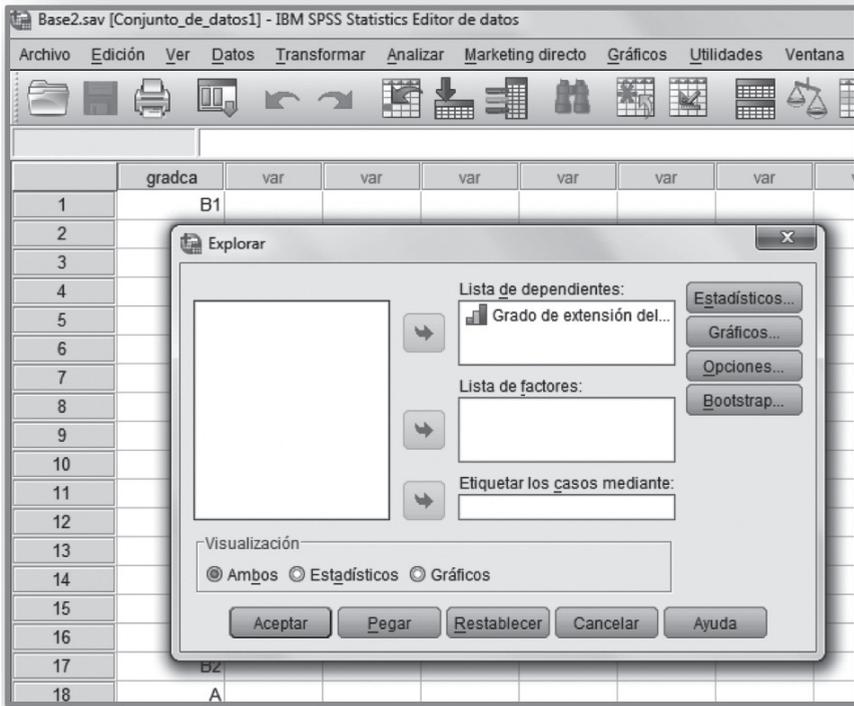


FIGURA A3.10. Cuadro de diálogo 'Explorar': 'Explorar'.

10. En el cuadro de diálogo 'Explorar' (Figura A3.10) se constata que en 'Visualización' esté activo 'Ambos'. Luego, se hace clic en el botón 'Aceptar': se selecciona la gráfica de Caja y bigotes.
11. SPSS presenta la gráfica de caja y bigotes (Figura A3.11).
12. Para la gráfica de la ojiva, se va al menú 'Gráficos' y se elige la opción 'Cuadros de Diálogos antiguos' y, luego, la opción 'Áreas' (se puede también optar por 'Líneas') Figura A3.12.

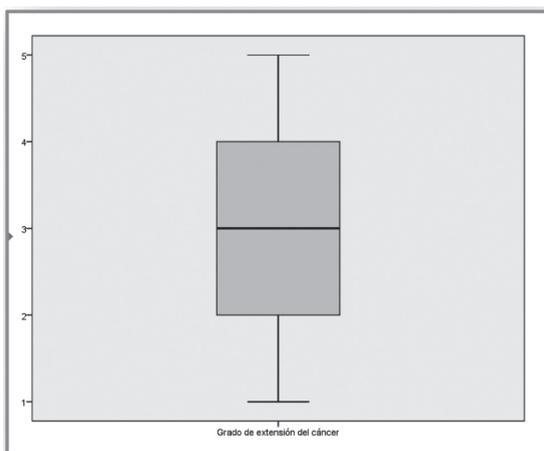


FIGURA A3.11. Gráfica caja y bigotes.

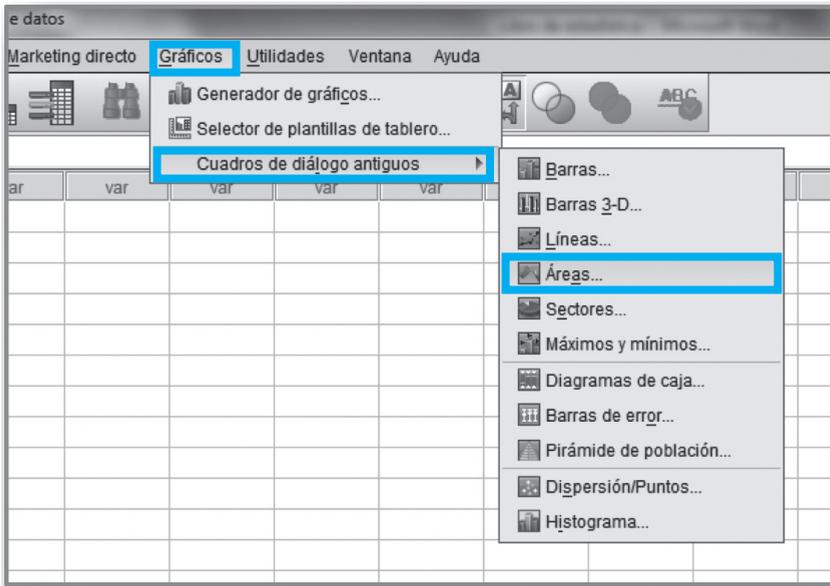


FIGURA A3.12. Lista de Menú 'Gráficos'.

13. En el cuadro de diálogo 'Áreas', se elige la opción 'Simple', se hace clic en el botón 'Definir' y aparece el cuadro de diálogo 'Gráficos de Áreas', se hace clic en 'Simple'; en 'Los datos del gráfico son', se hace clic en 'Resúmenes para grupos de casos'. En este cuadro, se pasa la variable 'Grado de cáncer' al rectángulo 'Eje de Categorías'. Luego, en el ítem 'El Área representa', se activa la opción '% acum', para que el programa presente la gráfica considerando los porcentajes acumulados (Figuras A3.13 y A3.14). Luego, clic en el botón 'Aceptar'.
14. En las Figuras A3.15 y Figura A3.16, se observa la ojiva obtenida con los comandos 'Áreas' y



FIGURA A3.13. Cuadro de diálogo 'Gráficos de áreas'.

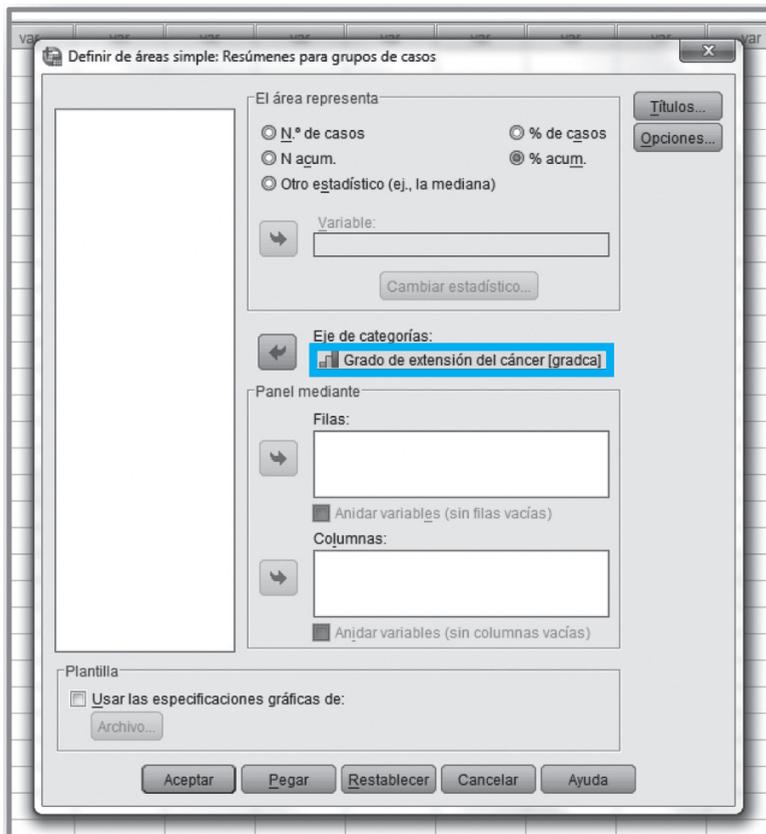
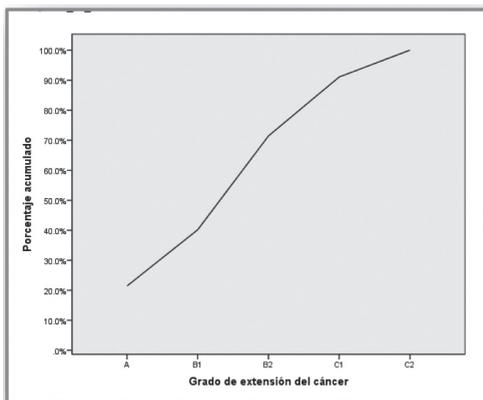


FIGURA A3.14. Cuadro de diálogo 'Definir de áreas simple: Resúmenes para grupos de datos'.

'Líneas', respectivamente.



106 FIGURA A3.15. Gráfica ojiva (áreas).

FIGURA A3.16. Gráfica ojiva (líneas).

## Análisis de variable nominal

1. Se utiliza el archivo base1.sav y se le indica a SPSS los siguientes comandos: 'Analizar' >>> 'Estadísticos Descriptivos' >>> 'Frecuencias' (Figura A4.1).
2. En el cuadro de diálogo 'Frecuencias', se selecciona la variable 'sex' y se la pasa al recuadro 'Variables', mediante un clic en el botón ubicado entre ambos recuadros. Hacer clic en el botón 'Estadísticos', en el cuadro de diálogo 'Frecuencias: Estadísticos', se activa solo la opción Moda (Figura A4.2).

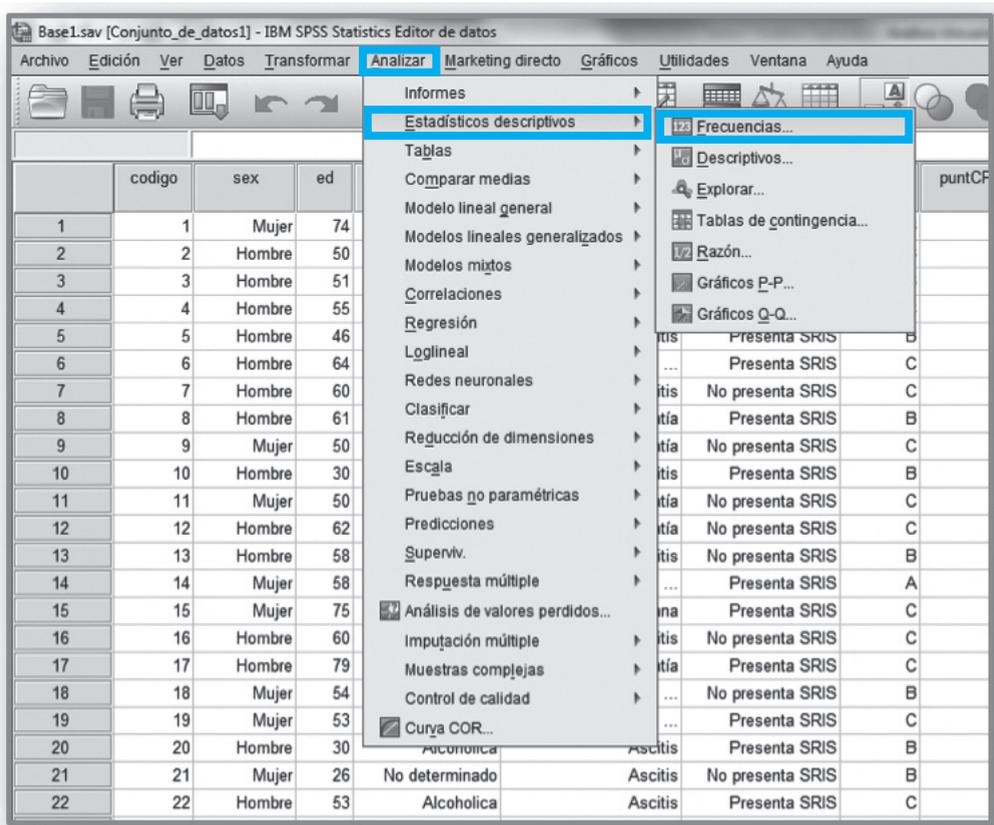


FIGURA A4.1. Lista de menú 'Analizar', 'Estadísticos descriptivos' y 'Frecuencias'.

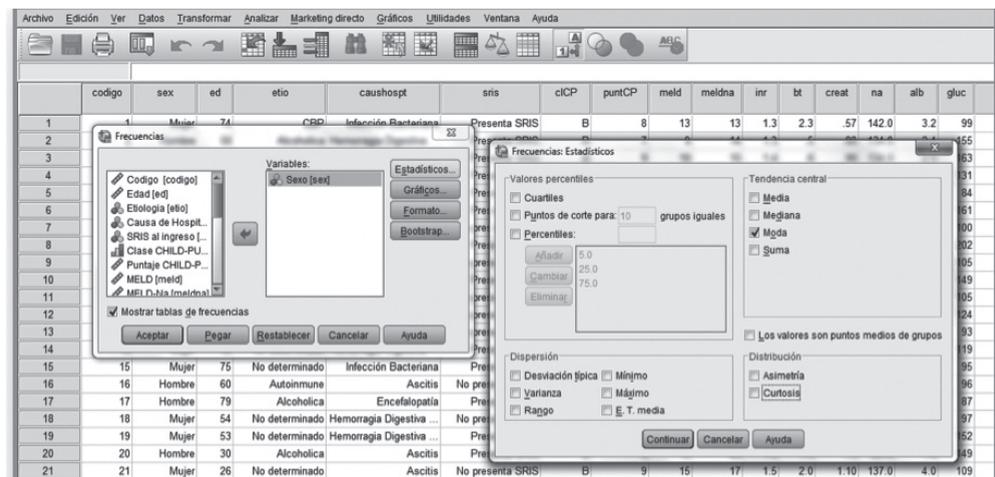


FIGURA A4.2. Medidas de resumen en variables nominales.

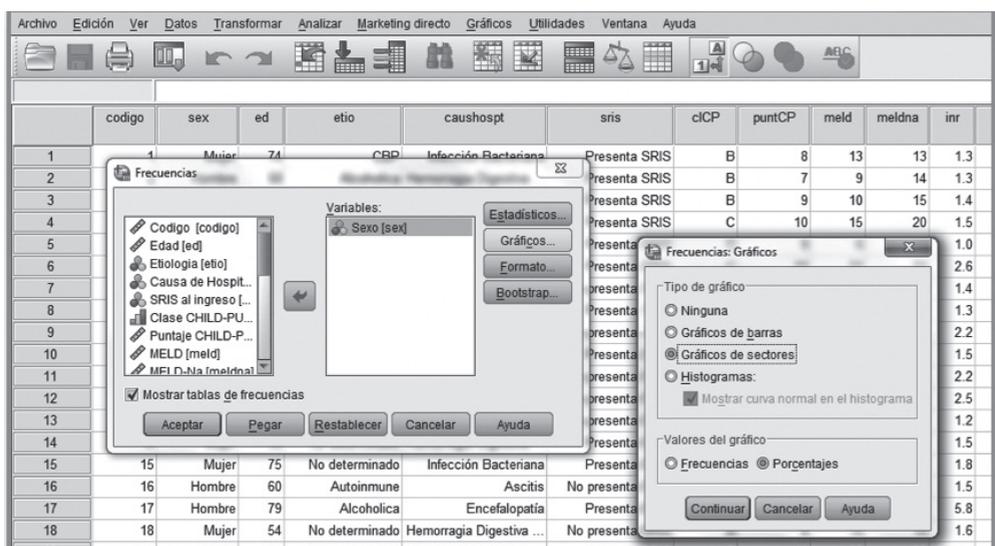


FIGURA A4.3. Frecuencias y gráficos para una variable nominal.

Sexo		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Hombre	115	57.5	57.5	57.5
	Mujer	85	42.5	42.5	100.0
	Total	200	100.0	100.0	

108 FIGURA A4.4. Cuadro de frecuencias absolutas, relativas y acumuladas.

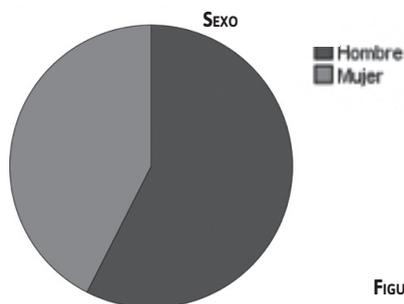


FIGURA A4.5. Gráficos de sectores de variable nominal sexo.

3. Se hace clic en el botón 'Continuar'. En el cuadro de diálogo 'Frecuencias', se hace clic en el botón 'Gráficos'; en el cuadro de diálogo 'Frecuencias: Gráficos', clic en la opción 'Gráficos de Sectores' y en el ítem 'Valores del Gráfico', se activa 'Porcentajes'. Se hace clic en el botón 'Continuar' y en el cuadro de diálogo 'Frecuencias' y se vuelve a hacer clic en el botón 'Aceptar' (Figura A4-3).
4. El programa presenta los resultados mostrados en las Figuras A4-4 y A4-5.
5. Se puede mejorar la presentación del gráfico:
  - a. En la ventana del 'Editor de Resultados', se hace doble clic sobre el gráfico y aparece el 'Editor de Gráficos'. En esta ventana, se hace doble clic sobre la gráfica para que aparezca el cuadro de diálogo 'Propiedades', en el cual se selecciona la etiqueta 'Profundidad y Ángulo': se puede elegir 'imagen 3D' y el nivel de ángulo que se desea. Para darle profundidad a los sectores, se debe dar por ejemplo 20 en el ítem 'Profundidad' (%) y en el ítem 'Distancia' se puede elegir 1 (100) con solo desplazar el indicador seleccionándolo y arrastrándolo con el ratón (Figura A4-6).

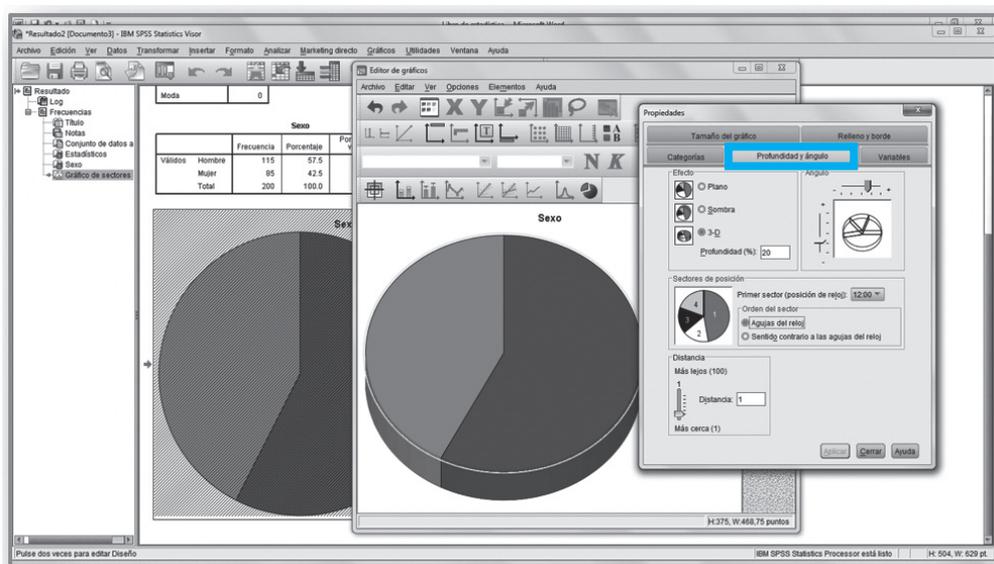


FIGURA A4.6. Realizar cambios en la Gráfica.

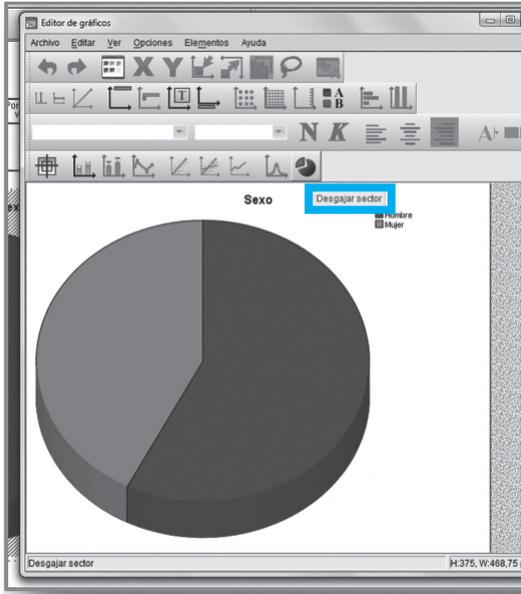


FIGURA A4.7. Uso de comando 'Desgajar sector' en 'Editor de gráficos'.

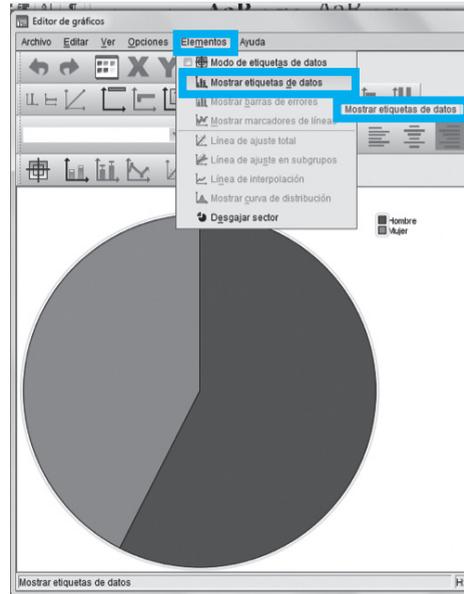


FIGURA A4.8. Uso de comando 'Desgajar Gráfico' en Editor de gráficos.

- b. Luego, se hace clic en el botón 'Aplicar' y, después, sobre el botón 'Cerrar'. Se cierra el 'Editor de Gráfico' y se pueden observar los cambios efectuados sobre el gráfico de 'sexo' (Figura A4.7).

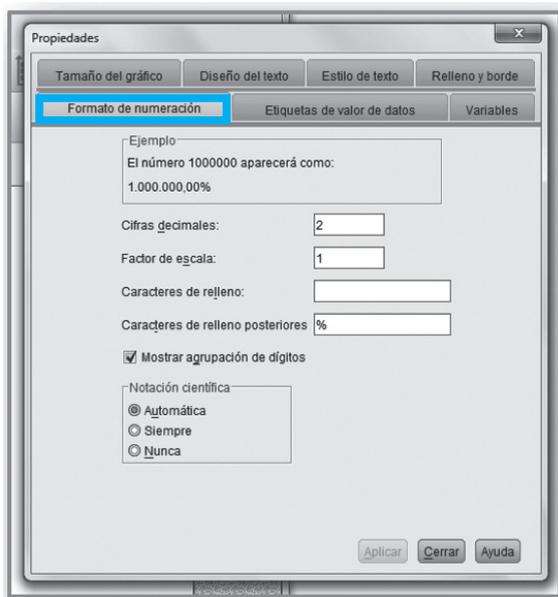
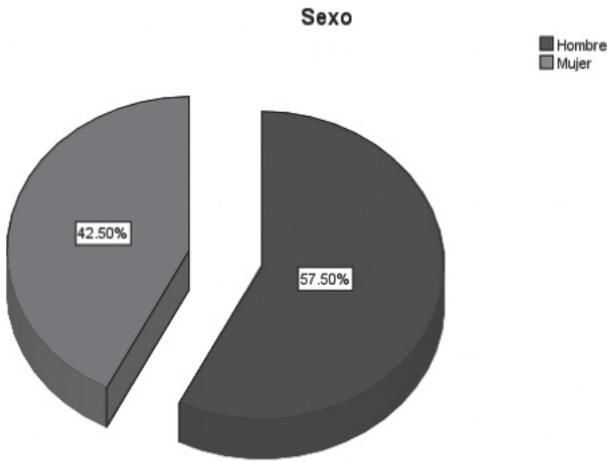


FIGURA A4.9. Cuadro de diálogo 'Propiedades'



**FIGURA A4.10.** Gráfica de variable 'sexo', con las modificaciones.

- c. Para que los sectores se aprecien separados, igual que un pastel recortado, en el 'Editor de Gráfico', se hace clic en el comando 'Desgajar Gráfico', que se encuentra en la barra de comando del 'Editor de Gráficos' (Figura A4.8).
- d. Para que se incluya en los sectores la 'Etiqueta de datos', en el mismo 'Editor de Gráficos', se seleccionan el menú 'Elementos' y se 'elige' la opción 'Mostrar Etiqueta de Datos'. De inmediato, se activa el cuadro de diálogo 'Propiedades'.

En este cuadro de diálogo se observan siete 'Etiquetas' para poder realizar cambios sobre el gráfico (Figura A4-9). Luego de realizar los cambios deseados, se hace clic en el botón 'Aplicar' y, luego, clic sobre el botón 'Cerrar'. Finalmente, se cierra el 'Editor de Gráficos' y se presenta en el 'Editor de Resultados' el gráfico con los cambios efectuados (Figura A4.10).



## Crear una nueva variable: comando Calcular variable

1. Mediante la utilización de la base de datos IMC del Anexo 9, se va a crear una nueva variable a la que se llamará 'índice de masa corporal' ('imc').  
En el menú, se despliega 'Transformar' y se elige 'Calcular variable' (Figura A5.1).
2. En el cuadro de diálogo 'Calcular variable' (Figura A5.2), en el recuadro 'Variable de destino' se debe dar el nombre de la nueva variable.

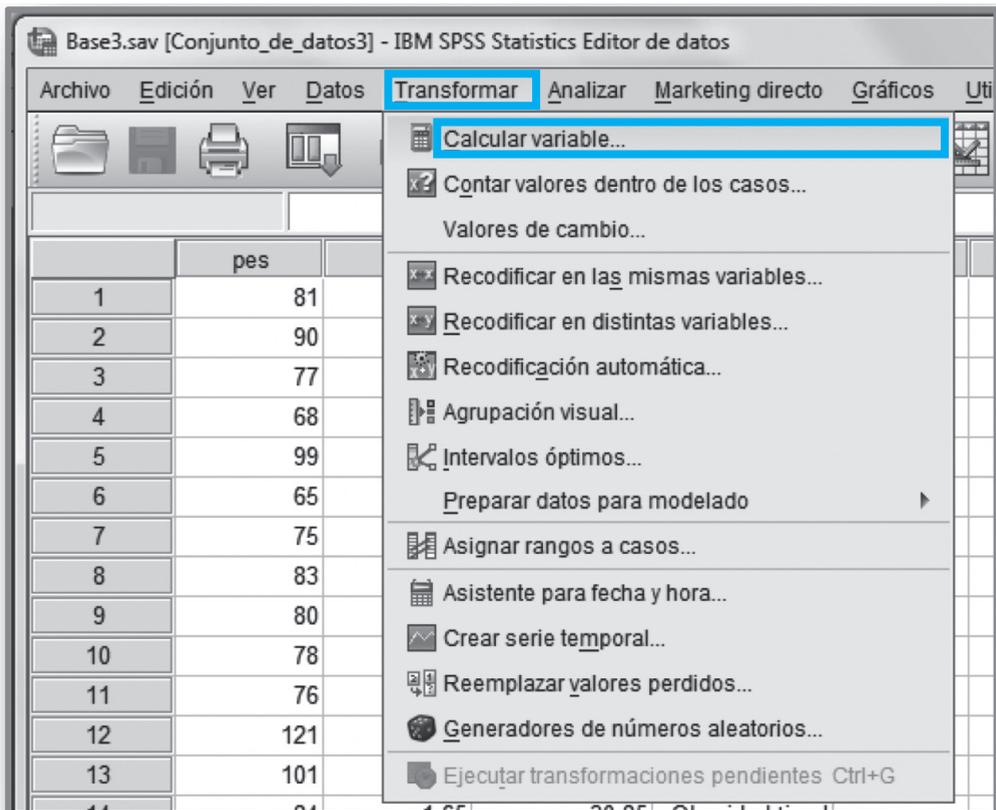


FIGURA A5.1. Menú 'Transformar', 'Calcular variable'.

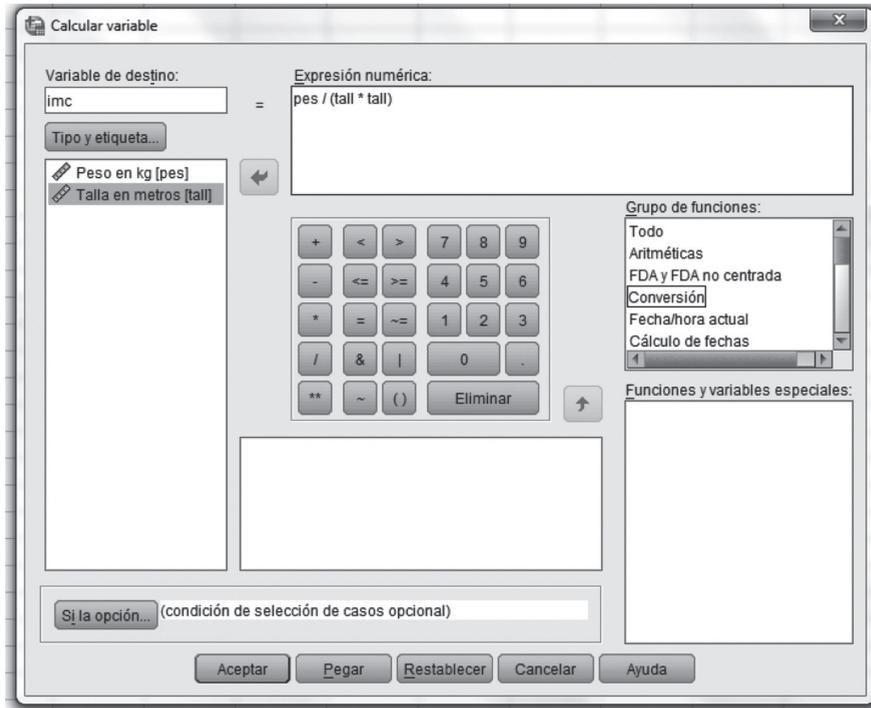


FIGURA A5.2. Cuadro de diálogo 'Calcular variable'.

Luego, se hace clic en el botón 'Tipo y etiqueta', para activar el cuadro de diálogo 'Calcular variable: Tipo y etiqueta' (Figura A5.3), en donde se da el texto para la 'Etiqueta' de la variable y se especifica el tipo 'Numérico' o 'Cadena' de la variable. Se cierra el cuadro de diálogo 'Tipo y etiqueta'.

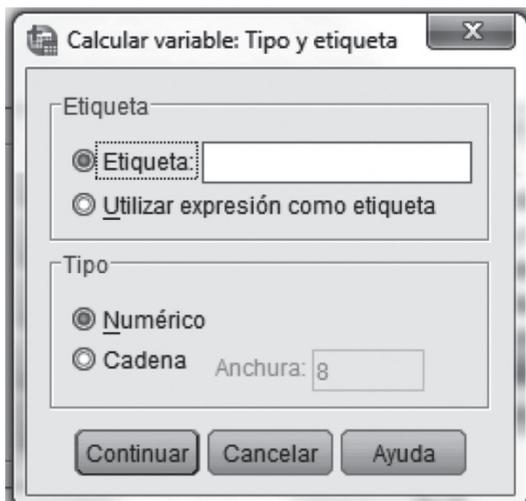


FIGURA A5.3. Cuadro 'Tipo y etiqueta'.

	pes	tall	imc	var	var
1	81	1.67	29.04		
2	90	1.72	30.42		
3	77	1.62	29.34		
4	68	1.71	23.26		
5	99	1.72	33.46		
6	65	1.67	23.31		
7	75	1.66	27.22		
8	83	1.77	26.49		
9	80	1.65	29.38		
10	78	1.70	26.99		

FIGURA A5.4. Vista de 'Datos' y los valores de la nueva variable creada.

En el recuadro 'Expresión numérica' del cuadro de diálogo 'Calcular variable' (Figura A5.2), se procede a expresar la fórmula del índice de masa corporal. Se selecciona la variable peso y se la pasa al recuadro presionando la tecla de fecha hacia la derecha. Luego, se teclea el operador 'de división' que se encuentra en el teclado de calculadora.

Seguidamente, se teclea el operador 'paréntesis' y dentro del mismo se realiza la introducción de la variable 'talla', el operador 'de multiplicación' y, nuevamente, la variable 'talla'. Terminado este proceso, se teclea el botón 'Aceptar'.

3. En 'Editor de datos' y en 'Vista de datos' se crea automáticamente la nueva variable y en 'Vista de variables' aparecen las características de la nueva variable 'imc' (Figuras A5.4 y A5.5).

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	pes	Numérico	8	0	Peso en kg	Ninguna	Ninguna	8	Derecha	Escala	Entrada
2	tall	Numérico	8	2	Talla en metros	Ninguna	Ninguna	8	Derecha	Escala	Entrada
3	imc	Numérico	8	2	Indice de Masa...	Ninguna	Ninguna	10	Derecha	Escala	Entrada
4											

FIGURA A5.5. Vista de variables y las características de la nueva variable creada.



## Recodificar una variable

### Variable índice de masa corporal recodificar en la misma variable

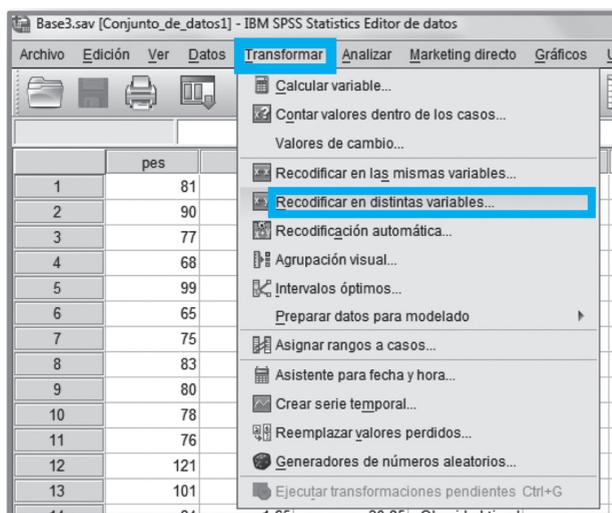
1. Siguiendo el caso de la variable 'imc' (índice de masa corporal), se realizará la recodificación de esta variable de acuerdo con la Tabla A6.1.
2. Desplegar el menú 'Transformar' y seleccionar el comando 'Recodificar en distintas variables'. En el cuadro de diálogo 'Recodificar en distintas variables', trasladar la variable 'imc' hacia el recuadro 'Variable numérica' → 'variable de resultado'. Luego, en el ítem 'Variable de resultado', dar el nombre de la variable y su etiqueta. A continuación, cliclear el botón 'Cambiar' (Figuras A6.1 y A6.2).
3. Posteriormente, hacer clic en el botón 'Valores antiguos y nuevos' y el SPSS presentará el cuadro de diálogo 'Recodificar en distintas variables: Valores antiguos y nuevos'. En este cuadro de diálogo, se introducen cada uno de los valores de 'imc' de acuerdo a la Tabla A6.1.
4. Para el primer valor, se activa 'Rango, INFERIOR hasta valor' y se introduce el valor. Luego, en el ítem 'Valor nuevo', se coloca '1' en el recuadro 'Valor', y, posteriormente, se hace clic en el botón 'Añadir', para aceptar los valores.

Repetir los mismos pasos para los siguientes valores. Se debe tener en cuenta que para los valores del 2 al 5, se activar 'Rango' y se introducen los rangos respectivos. Para el valor 6, activar 'Rango, valor hasta SUPERIOR'. Posteriormente, hacer clic en el botón 'Continuar'.

**TABLA A6.1.** TABLA DE VALORES DE IMC

Diagnóstico de IMC	Valores de IMC
• 1	Menor de 18,5
• 2	De 18,5 a 25
• 3	De 25 a 30
• 4	De 30 a 35
• 5	De 35 a 40
• 6	Mayor de 40

Fuente: Organización Mundial de la Salud.



**FIGURA A6.1.** Comando recodificar en distintas variables.

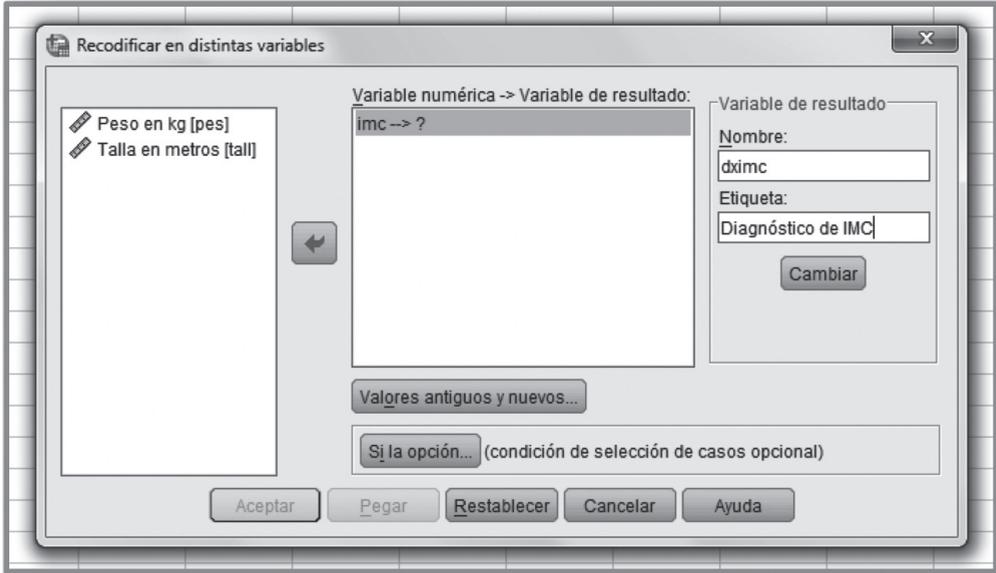
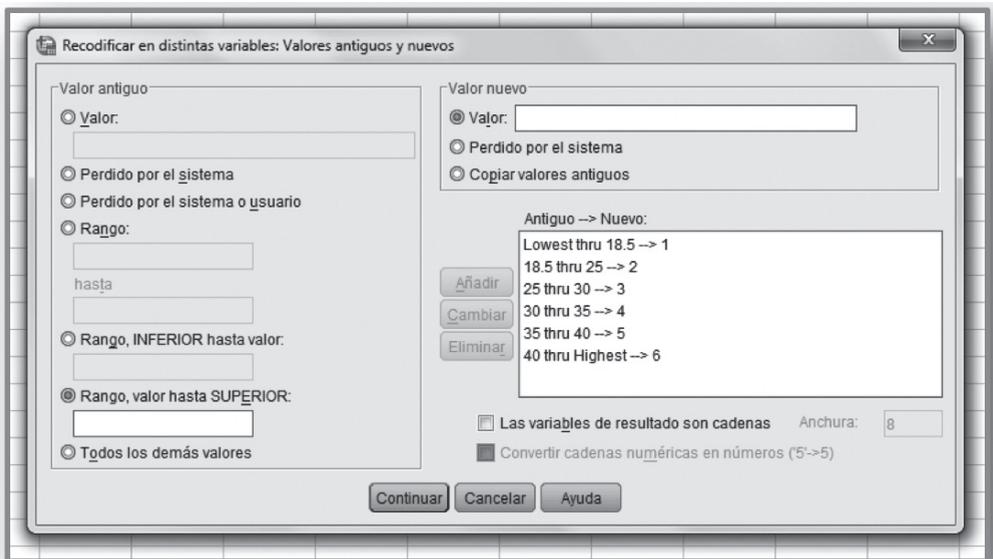


FIGURA A6.2. Cuadro de diálogo 'Recodificar en distintas variables).

SPSS genera la nueva variable en 'Vista datos' y en 'Vista variables' se realizan los cambios necesarios (Figura A6.4A).

5. En la misma 'Vista variables', se activa en la variable 'dximc' la característica 'Valores' (Figura A6.4B), para introducir la información de los diagnósticos de acuerdo con el valor del 'imc' (Tabla A6.2).



118 FIGURA A6.3. Cuadro de diálogo 'Valores antiguos y nuevos'.

A

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	pes	Numérico	8	0	Peso en kg	Ninguna	Ninguna	8	Derecha	Escala	Entrada
2	tall	Numérico	8	2	Talla en metros	Ninguna	Ninguna	8	Derecha	Escala	Entrada
3	imc	Numérico	8	2	Índice de Masa...	Ninguna	Ninguna	10	Derecha	Escala	Entrada
4	dximc	Numérico	8	0	Diagnóstico de ...	Ninguna	Ninguna	10	Derecha	Ordinal	Entrada

B

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores
1	pes	Numérico	8	0	Peso en kg	Ninguna
2	tall	Numérico	8	2	Talla en metros	Ninguna
3	imc	Numérico	8	2	Índice de Masa...	Ninguna
4	dximc	Numérico	8	0	Diagnóstico de ...	Ninguna

FIGURA A6.4. A) 'Vista de variables' con la nueva variable 'dximc'. B) 'Vista de variables': Activar valores.

- En 'Etiquetas de valor', se coloca 1 en el recuadro 'Valor' y en 'Etiqueta', se escribe 'Bajo peso'. Luego, hacer clic en el botón 'Añadir'. Repetir para el resto de los valores (Figura A6.5). Seguidamente, se hace clic en el botón 'Aceptar'.
- En 'Vista de datos' se observa la nueva variable 'Diagnóstico de índice de masa corporal' (Figura A6.6).

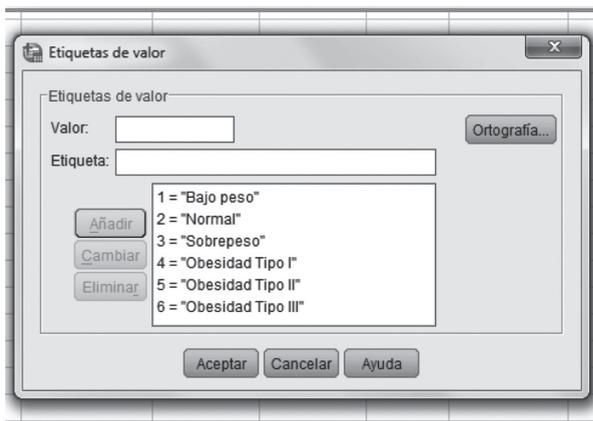


FIGURA A6.5. Cuadro de diálogo 'Etiquetas de valores'.

TABLA A6.2. TABLA DE VALORES DE IMC

Diagnóstico de IMC	Valores de IMC
• Bajo peso	Menor de 18,5
• Normal	De 18,5 a 25
• Sobrepeso	De 25 a 30
• Obesidad tipo 1	De 30 a 35
• Obesidad tipo 2	De 35 a 40
• Obesidad tipo 3	Mayor de 40

Fuente: Organización Mundial de la Salud

	pes	tall	imc	dximc
1	81	1.67	29.04	Sobrepeso
2	90	1.72	30.42	Obesidad Tipo I
3	77	1.62	29.34	Sobrepeso
4	68	1.71	23.26	Normal
5	99	1.72	33.46	Obesidad Tipo I
6	65	1.67	23.31	Normal
7	75	1.66	27.22	Sobrepeso

FIGURA A6.6. Resultado del ingreso de los valores de basas de datos.

### Variable puntaje Child-Pugh recodificar en distinta variable

1. Para este caso, se utiliza la variable 'pcp' (puntaje Child-Pugh), para crear la variable 'gcp' (grado de Child-Pugh). Tabla A6.3.
2. En el comando 'Transformar' (Figura A6.7), seleccionar 'Recodificar en distintas variables' y en el cuadro de diálogo 'Recodificar en distintas variables'. Se selecciona la variable 'pcp' (puntaje Child-Pugh) hacia el recuadro vacío, mediante la utilización del botón de flecha hacia la derecha (Figura A6.8). En el ítem 'Variable de resultado', se procede a ingresar el nombre de la nueva variable y la 'Etiqueta' de esta. Luego, se teclea el botón 'Cambiar'.

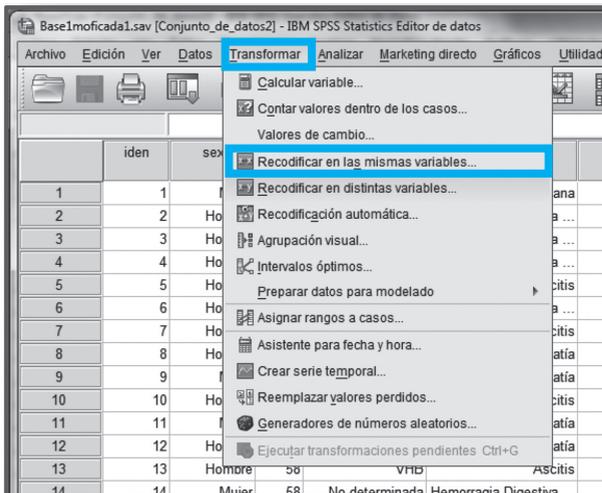


TABLA A6.3. GRADO CHILD-PUGH DE ACUERDO CON EL PUNTAJE

Grado Child-Pugh	Puntaje Child-Pugh
• 1	5-6
• 2	7-9
• 3	10-15

Fuente: <https://medicina.ufm.edu/eponimo/escala-de-child-pugh/#:-:text=Es%20un%20sistema%20de%20estadificaci%C3%B3n,de%20un%20trasplante%20de%20h%C3%ADgado>

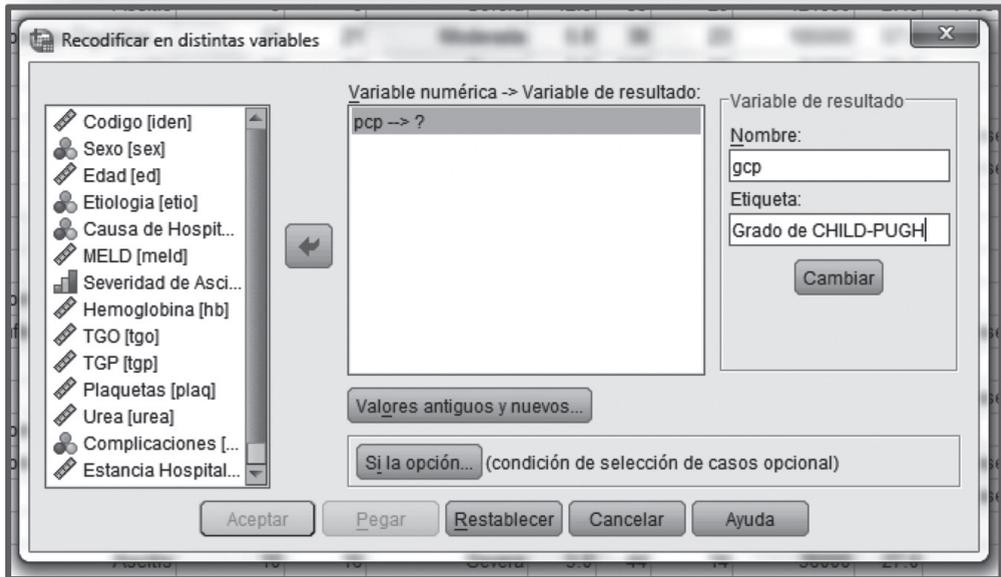


FIGURA A6.8. Cuadro de diálogo 'Recodificar en distintas variables'.

Posteriormente, se hace clic en el botón 'Valores antiguos y nuevos', para introducir los parámetros de transformación de la variable existente para que se genere la nueva variable.

3. En el cuadro de diálogo 'Recodificar en distintas variables: Valores antiguos y nuevos' (Figura A6.9), especificar en el ítem 'Valor antiguo', los valores que corresponden a cada grado y teclear el botón Añadir para aceptar los nuevos valores.

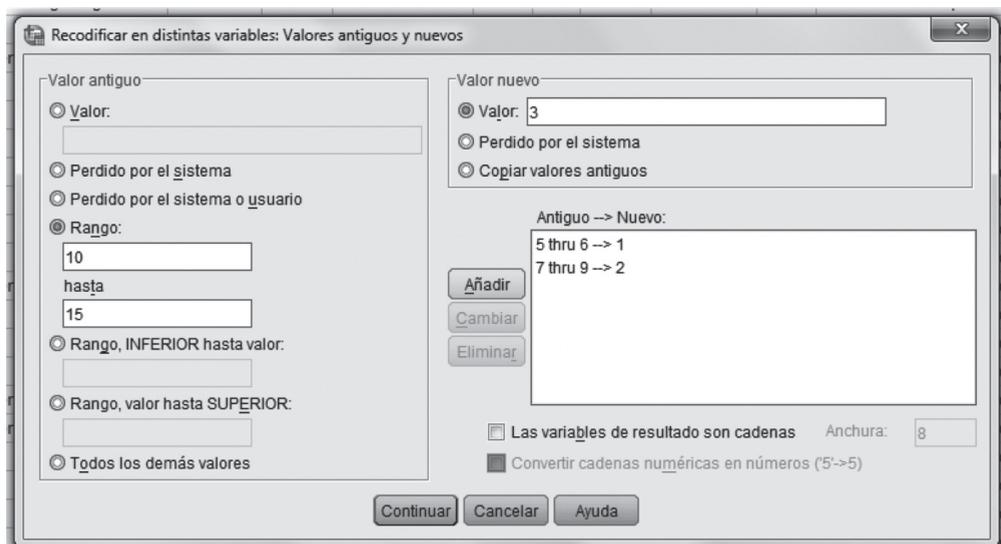


FIGURA A6.9. 'Recodificar en distintas variables: Valores antiguos y nuevos'.

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	iden	Númerico	3	0	Codigo	Ninguna	Ninguna	6	Derecha	Escala	Entrada
2	sex	Númerico	8	0	Sexo	{0, Hombre}...	Ninguna	8	Derecha	Nominal	Entrada
3	ed	Númerico	2	0	Edad	Ninguna	Ninguna	4	Derecha	Escala	Entrada
4	etio	Númerico	1	0	Etiologia	{1, Alcoholi...	Ninguna	12	Derecha	Nominal	Entrada
5	cchp	Númerico	8	0	Causa de Hosp...	{0, Ascitis}...	Ninguna	15	Derecha	Nominal	Entrada
6	pcp	Númerico	2	0	Puntaje CHIL...	Ninguna	Ninguna	7	Derecha	Escala	Entrada
7	meld	Númerico	2	0	MELD	Ninguna	Ninguna	5	Derecha	Escala	Entrada
8	asc	Númerico	1	0	Severidad de A...	{0, Ausente}...	Ninguna	11	Derecha	Ordinal	Entrada
9	hb	Númerico	3	1	Hemoglobina	Ninguna	Ninguna	4	Derecha	Escala	Entrada
10	tgo	Númerico	4	0	TGO	Ninguna	Ninguna	3	Derecha	Escala	Entrada
11	tgp	Númerico	4	0	TGP	Ninguna	Ninguna	5	Derecha	Escala	Entrada
12	plaq	Númerico	5	0	Plaquetas	Ninguna	Ninguna	8	Derecha	Escala	Entrada
13	urea	Númerico	4	1	Urea	Ninguna	Ninguna	4	Derecha	Escala	Entrada
14	comp	Númerico	8	0	Complicaciones	{0, Sin com...	Ninguna	18	Derecha	Nominal	Entrada
15	esthos	Númerico	2	0	Estancia Hospi...	Ninguna	Ninguna	3	Derecha	Escala	Entrada
16	alt	Númerico	1	0	Condicion al Alta	{0, MEJOR...	Ninguna	9	Derecha	Nominal	Entrada
17	gcp	Númerico	8	2	Grado de CHIL...	Ninguna	Ninguna	10	Derecha	Nominal	Entrada

FIGURA A6.10. Vista 'Variables'.

4. Después de introducir los nuevos valores, se hace clic en el botón 'Continuar'. Luego, en el cuadro de diálogo 'Recodificar en distintas variables', hacer clic en botón 'Aceptar'. SPSS genera la nueva variable. Se debe realizar los cambios necesarios de las características de la variable en 'Vista de variables' (Figura A6.10), en la característica 'Decimales' cambian de '2' a '0', en 'Medida' cambiar 'Nominal' a 'Ordinal'.

	iden	sex	ed	etio	cchp	pcp	sthos	alt	gcp
1	1	Mujer	74	VHC	Infección Bacteriana	8	6	MEJORADO	2
2	2	Hombre	50	Alcoholica	Hemorragia Digestiva ...	7	8	MEJORADO	2
3	3	Hombre	51	Alcoholica	Hemorragia Digestiva ...	9	19	MEJORADO	2
4	4	Hombre	55	No determinada	Hemorragia Digestiva ...	10	4	MEJORADO	3
5	5	Hombre	46	Alcoholica	Ascitis	9	7	MEJORADO	2
6	6	Hombre	64	Alcoholica	Hemorragia Digestiva ...	11	5	MEJORADO	3
7	7	Hombre	60	Alcoholica	Ascitis	11	6	MEJORADO	3
8	8	Hombre	61	No determinada	Encefalopatía	7	4	MEJORADO	2
9	9	Mujer	50	Alcoholica	Encefalopatía	13	2	FALLECIDO	3
10	10	Hombre	30	Alcoholica	Ascitis	9	26	MEJORADO	2
11	11	Mujer	50	Alcoholica	Encefalopatía	13	2	FALLECIDO	3
12	12	Hombre	62	Alcoholica	Encefalopatía	13	7	MEJORADO	3
13	13	Hombre	58	VHB	Ascitis	7	13	MEJORADO	2
14	14	Mujer	58	No determinada	Hemorragia Digestiva ...	6	11	MEJORADO	1
15	15	Mujer	75	No determinada	Infección Bacteriana	10	3	FALLECIDO	3
16	16	Hombre	60	Autoinmune	Ascitis	10	3	MEJORADO	3
17	17	Hombre	79	Alcoholica	Encefalopatía	14	3	FALLECIDO	3
18	18	Mujer	54	No determinada	Hemorragia Digestiva ...	8	7	MEJORADO	2

122 FIGURA A6.11. 'Vista de datos'. La variable 'gcp' (grado de Child-Pugh) creada y sus respectivos valores en cada caso.

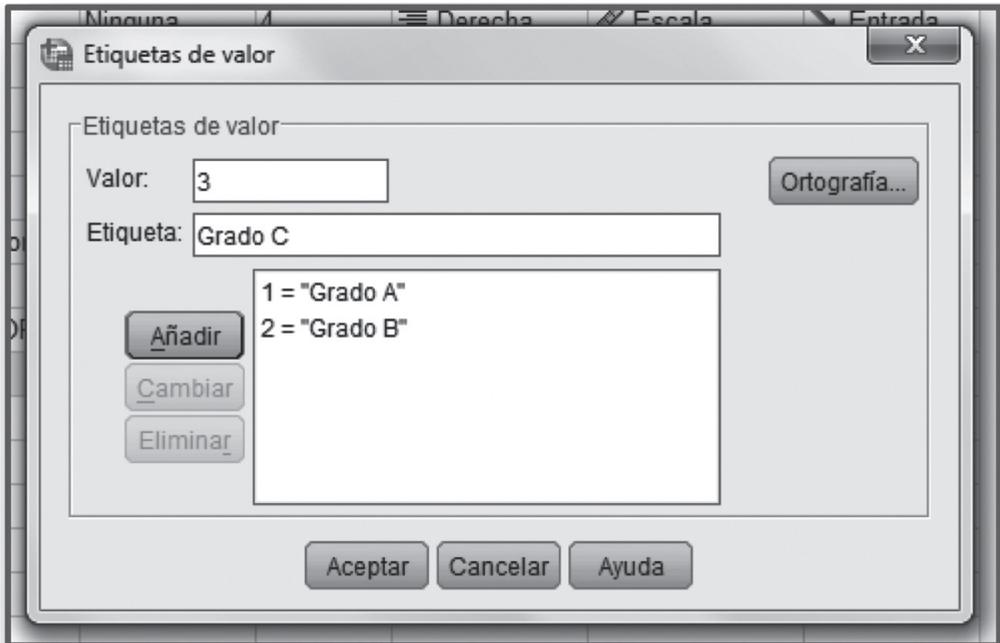


FIGURA A6.12. 'Etiquetas de Valor'.

5. Se observan los cambios en 'Vista de datos' (Figura A6.11).
6. En 'Vista variables', se activa en la variable 'gcp' la característica 'Valores' (Figura A6.12) para introducir la información de los diagnósticos de acuerdo con el valor del IMC. Después de introducir todos los valores y grados, se hace clic en el botón 'Aceptar'. En 'Vista datos', se puede observar la variable 'grado de Child-Pugh'.



## Selección de casos

A veces es necesario realizar el análisis de las características de una muestra o población considerando un atributo o variable en particular. Por ejemplo, se quiere evaluar las características de los pacientes con cirrosis hepática de acuerdo con su sexo. Para esto, se tiene que dividir la base en dos grupos de acuerdo con su sexo. SPSS ofrece la selección de casos como artificio para poder realizar el análisis estadístico por separado de acuerdo con un factor de clasificación.

1. Se despliega el menú 'Datos' (Figura A7-1) y se selecciona el comando 'Selección de casos'. Aparece el cuadro de diálogo 'Seleccionar casos', en el ítem Seleccionar existen cinco posibilidades de selección:

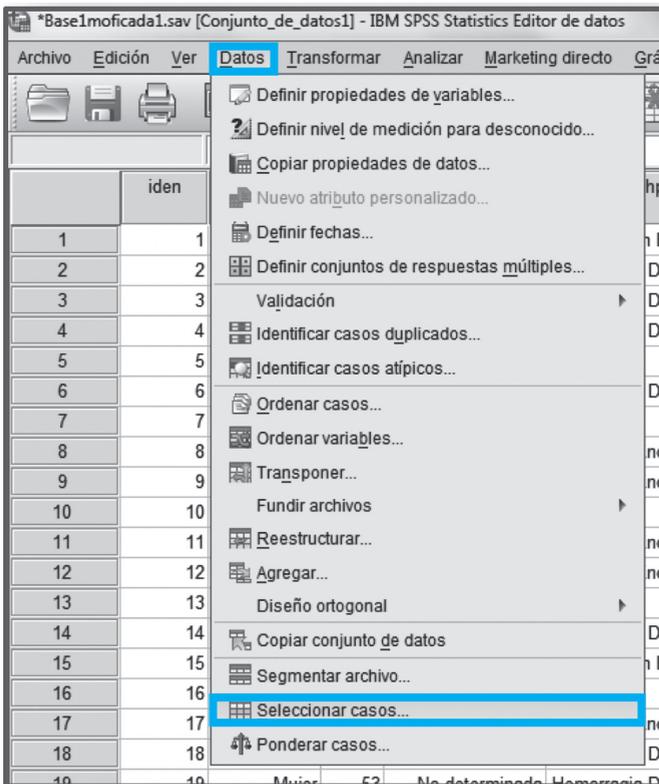


FIGURA A7.1. Comando 'Seleccionar casos'.

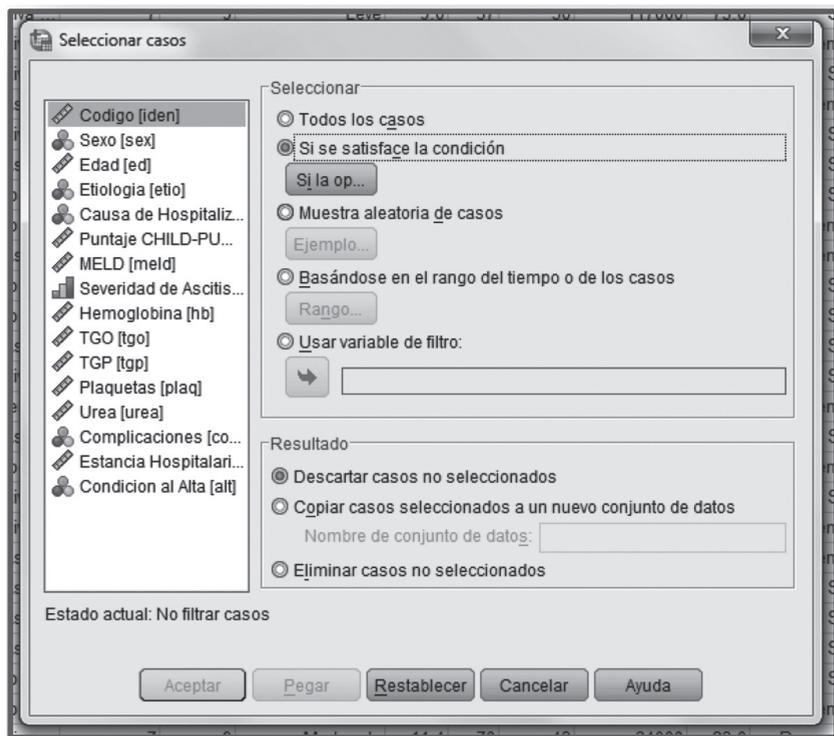
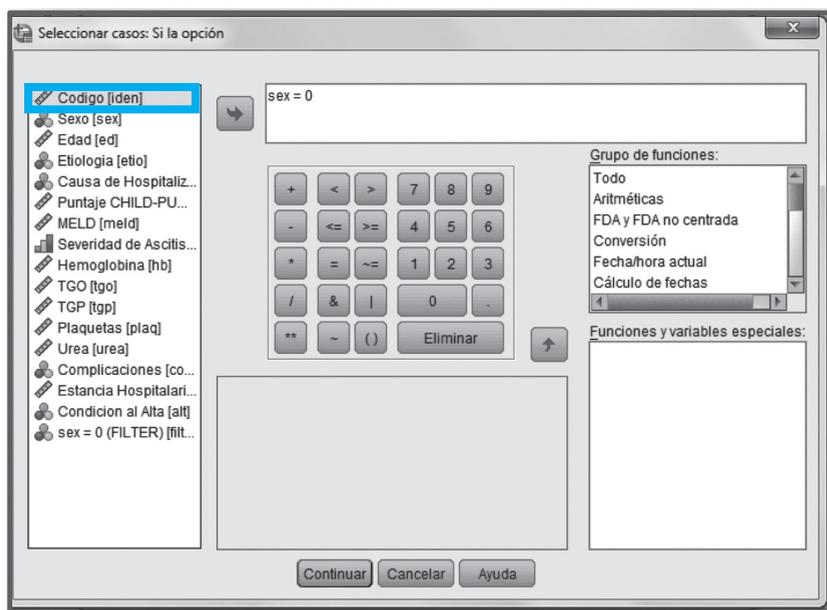


FIGURA A7.2. Cuadro de diálogo 'Seleccionar casos'.



126 FIGURA A7.3. Cuadro de diálogo 'Seleccionar casos: Si la opción'.

	iden	sex	ed	etio	cchp	pcp	meld	asc	hb	tgo
<del>1</del>	1	Mujer	74	VHC	Infeción Bacteriana	8	13	Leve	12.8	75
2	2	Hombre	50	Alcoholica	Hemorragia Digestiva ...	7	9	Leve	9.0	37
3	3	Hombre	51	Alcoholica	Hemorragia Digestiva ...	9	10	Severa	6.8	42
4	4	Hombre	55	No determinada	Hemorragia Digestiva ...	10	15	Severa	5.3	95
5	5	Hombre	46	Alcoholica	Ascitis	9	6	Severa	12.9	33
6	6	Hombre	64	Alcoholica	Hemorragia Digestiva ...	11	21	Moderada	5.8	38
7	7	Hombre	60	Alcoholica	Ascitis	11	14	Severa	5.6	113
8	8	Hombre	61	No determinada	Encefalopatía	7	12	Ausente	8.0	50
<del>9</del>	9	Mujer	50	Alcoholica	Encefalopatía	13	21	Severa	8.0	83
10	10	Hombre	30	Alcoholica	Ascitis	9	13	Moderada	7.1	61
<del>11</del>	11	Mujer	50	Alcoholica	Encefalopatía	13	21	Severa	8.0	83
12	12	Hombre	62	Alcoholica	Encefalopatía	13	20	Severa	11.0	55
13	13	Hombre	58	VHB	Ascitis	7	8	Severa	13.9	106
<del>14</del>	14	Mujer	58	No determinada	Hemorragia Digestiva ...	6	26	Ausente	14.5	43
<del>15</del>	15	Mujer	75	No determinada	Infeción Bacteriana	10	18	Ausente	12.5	102
16	16	Hombre	60	Autoinmune	Ascitis	10	15	Severa	14.0	172
17	17	Hombre	79	Alcoholica	Encefalopatía	14	40	Moderada	8.7	56
<del>18</del>	18	Mujer	54	No determinada	Hemorragia Digestiva ...	8	15	Leve	8.0	64
<del>19</del>	19	Mujer	53	No determinada	Hemorragia Digestiva ...	11	22	Leve	11.4	130

FIGURA A7.4. Vista de datos con datos seleccionados.

- Todos los casos.
- Si se satisface la condición.
- Muestra aleatoria de casos.
- Se basa en el rango de tiempo o de los casos.
- Usar variable de filtro.

En el caso del ejemplo, se selecciona la opción 'Si se satisface la condición'. Luego, se hace clic en el botón 'Si la op...'. SPSS ofrece el cuadro de diálogo 'Seleccionar casos: Si la opción' (Figura A7-2).

Nota: En el ítem 'Resultado', se debe constatar que esté activa la opción 'Descartar casos no seleccionados'.

- En el cuadro de diálogo 'Seleccionar casos: Si la opción' (Figura A7-3), se ingresa la variable 'sex' y los operadores (=) y (0). Se hace clic en el botón 'Continuar'.

El SPSS presenta en la 'Vista de datos', los casos seleccionados no tienen línea de tachado (Figura A7-4).



## Comando Explorar

Si se tiene la necesidad de obtener las medidas de resumen de una variable, considerando otra variable como atributo, se puede utilizar el comando 'Explorar'. Por ejemplo, se quiere evaluar la edad de los pacientes con cirrosis hepática de acuerdo con su sexo.

1. Se despliega el menú 'Analizar' y se selecciona el comando 'Estadísticos descriptivos' y, luego, 'Explorar' (Figura 8.1). Aparece el cuadro de diálogo 'Explorar', en el recuadro 'Lista de dependientes' pasar la variable 'Edad' y en el recuadro 'Lista de factores' se ingresa la variable 'Sexo'. Se puede agregar más de una variable. (Figura 8.2).

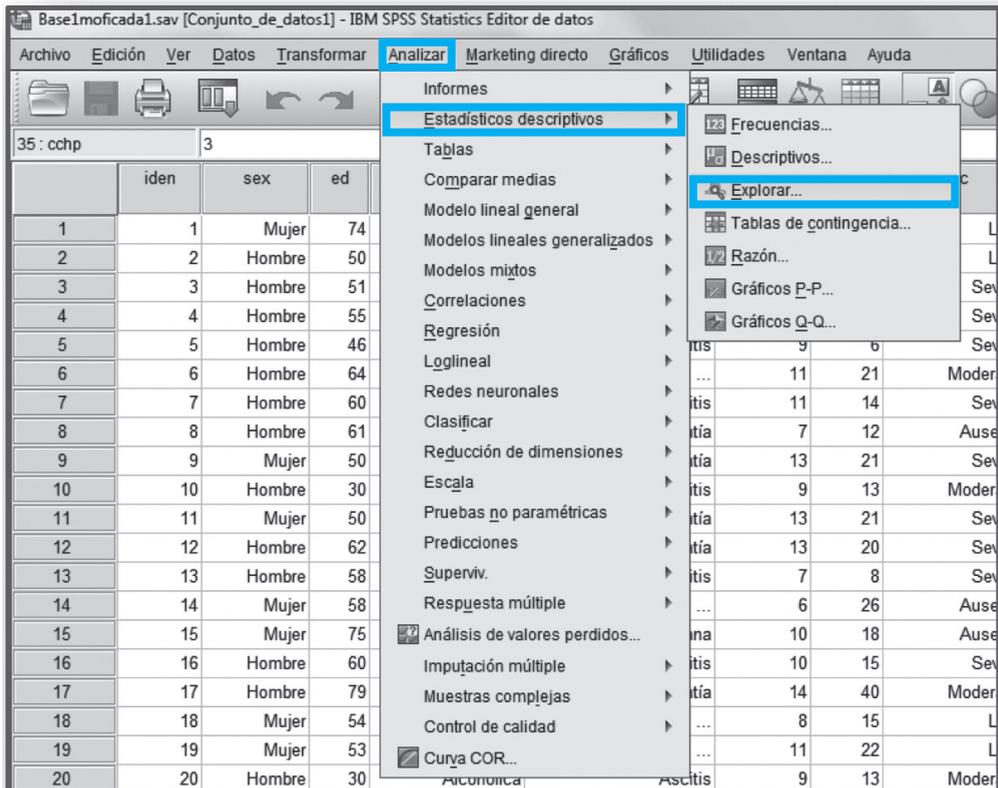


FIGURA A8.1. Comando 'Explorar'.

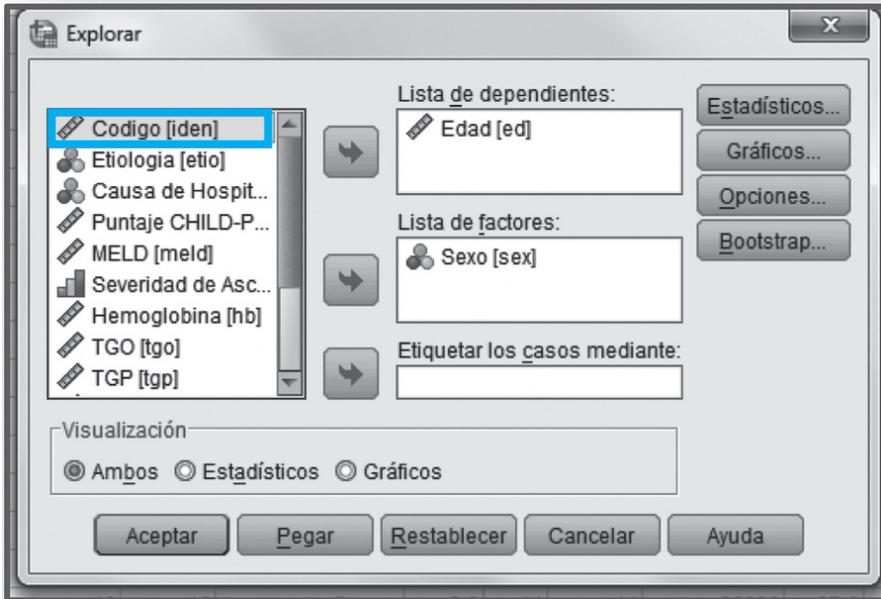


FIGURA A8.2. Cuadro de diálogo 'Explorar'.

2. Se hace clic en el botón 'Estadísticos' y en el cuadro de diálogo 'Explorar: Estadísticos'. Se constata que este activo 'Descriptivos' y, en 'Intervalo de confianza para la media', escribir 95. Finalmente, se hace clic en el botón 'Continuar' (Figura A8.3).
3. Nuevamente en el cuadro de diálogo 'Explorar', se hace clic en botón 'Gráficos', en el cuadro de diálogo 'Explorar: Gráficos'. Se debe constatar que, en el ítem 'Diagramas de caja', esté activo 'Niveles de los factores juntos'; en el ítem 'Descriptivos', se activa 'De tallo y

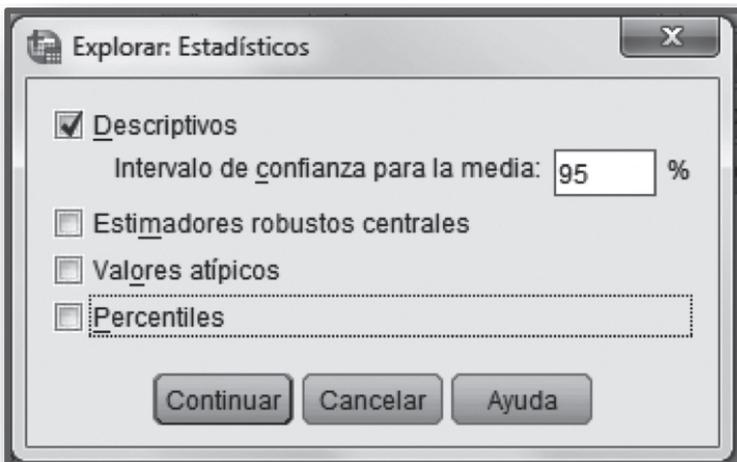


FIGURA A8.3. Cuadro de diálogo 'Explorar: Estadísticos'.

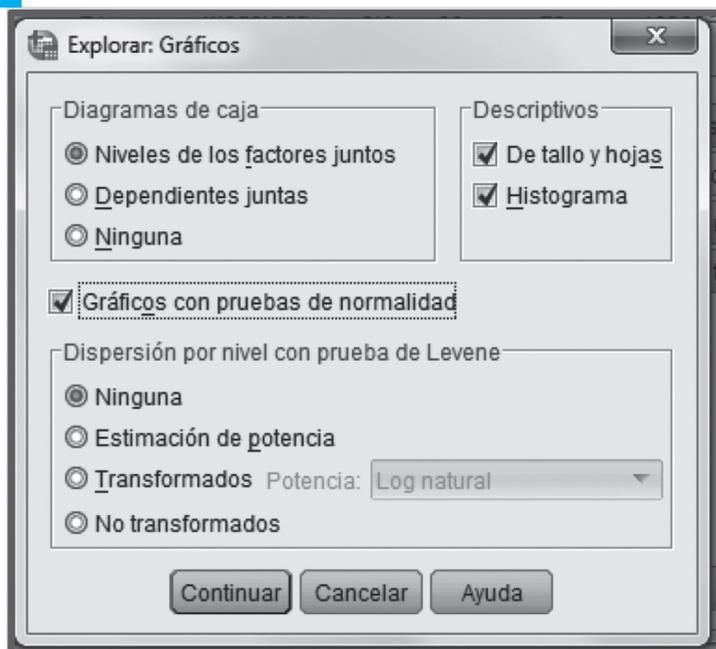


FIGURA A8.4. Cuadro de diálogo 'Explorar: Gráficos'.

hojas' e 'Histograma'; se activa 'Gráficos con pruebas de normalidad'. Se hace clic en botón 'Continuar'.

Finalmente, en el cuadro de diálogo 'Explorar', se hace clic en el botón 'Aceptar' (Figura A8.4).

4. SPSS presenta los resultados en visor de 'Resultados'. Figuras A8.5-A8.16.

		Casos					
		Válidos		Perdidos		Total	
		N	Porcentaje	N	Porcentaje	N	Porcentaje
Edad	Hombre	115	100.0%	0	.0%	115	100.0%
	Mujer	85	100.0%	0	.0%	85	100.0%

FIGURA A8.5. Tabla de resumen: variables 'Edad' y 'sexo'.

Descriptivos					
Sexo			Estadístico	Error típ.	
Edad	Hombre	Media	60.12	1.101	
		Intervalo de confianza para la media al 95%	Límite inferior		57.94
			Límite superior		62.30
		Media recortada al 5%	60.30		
		Mediana	60.00		
		Varianza	139.353		
		Desv. típ.	11.805		
	Mínimo	30			
	Máximo	97			
	Rango	67			
	Amplitud intercuartil	16			
	Asimetría	-.079	.226		
	Curtosis	.662	.447		
	Mujer	Mujer	Media		63.13
Intervalo de confianza para la media al 95%			Límite inferior	60.57	
			Límite superior	65.69	
Media recortada al 5%			63.27		
Mediana			65.00		
Varianza			140.971		
Desv. típ.			11.873		
Mínimo		26			
Máximo		88			
Rango		62			
Amplitud intercuartil		16			
Asimetría		-.283	.261		
Curtosis		.175	.517		

FIGURA A8.6. Tabla de medidas de resumen para 'Edad por sexo'.

Pruebas de normalidad							
Sexo		Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Estadístico	gl	Sig.	Estadístico	gl	Sig.
Edad	Hombre	.069	115	.200 <sup>*</sup>	.982	115	.118
	Mujer	.078	85	.200 <sup>*</sup>	.985	85	.425

a. Corrección de la significación de Lilliefors  
 \*. Este es un límite inferior de la significación verdadera.

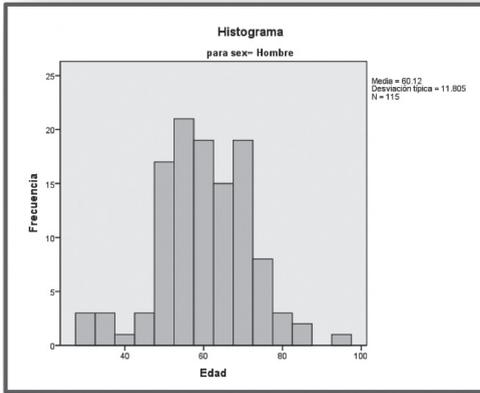


FIGURA A8.8. Histograma de frecuencias de la variable 'edad' en hombres.

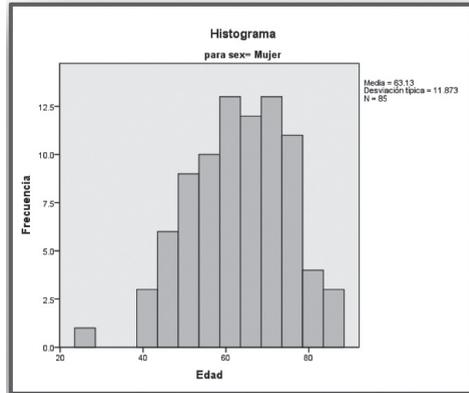


FIGURA A8.9. Histograma de frecuencias de la variable 'edad' en mujeres.

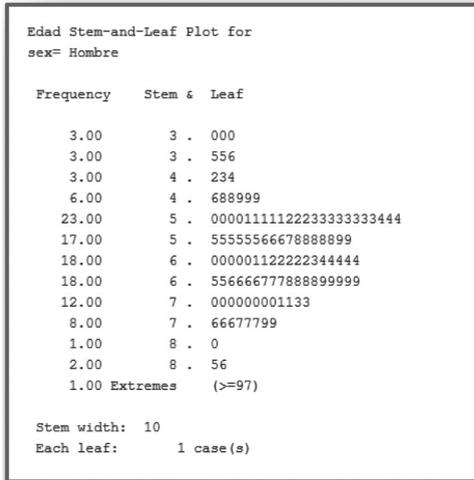


FIGURA A8.10. Gráfica de tallos y hojas de la variable 'edad' en hombres.

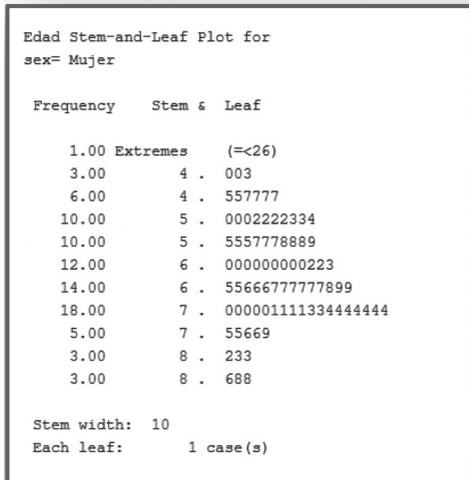


FIGURA A8.11. Gráfica de tallos y hojas de la variable 'edad' en mujeres'.

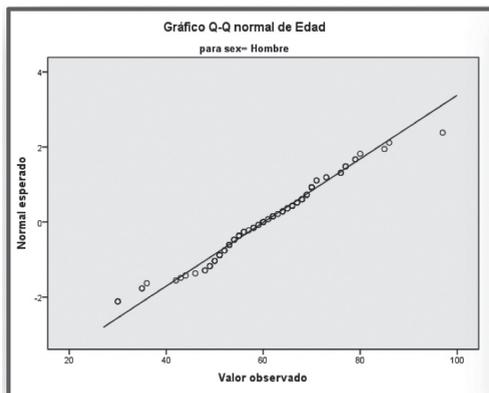


FIGURA A8.12. Gráfico Q-Q normal de 'edad' en hombres.

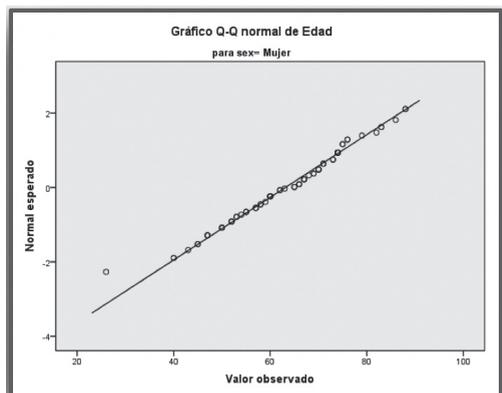


FIGURA A8.13. Gráfico Q-Q normal de 'edad' en mujeres.

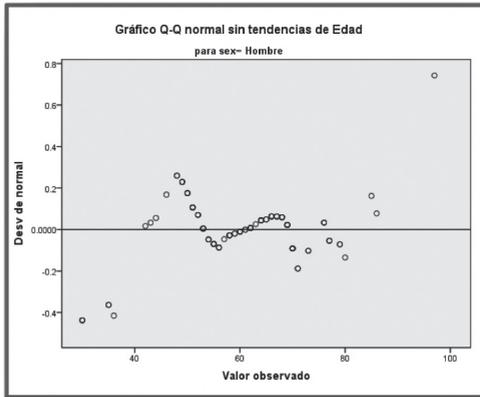


FIGURA A8.14. Gráfico Q-Q normal sin tendencia de edad en hombres.

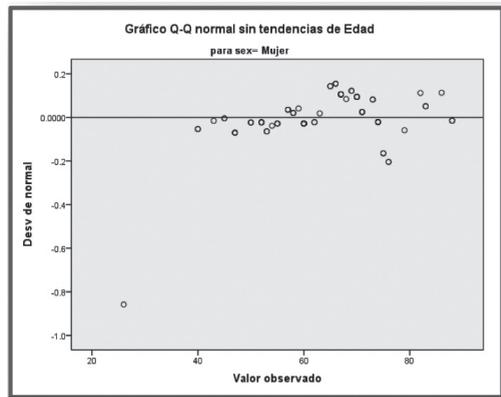


FIGURA A8.15. Gráfico Q-Q normal sin tendencia de edad en mujeres.

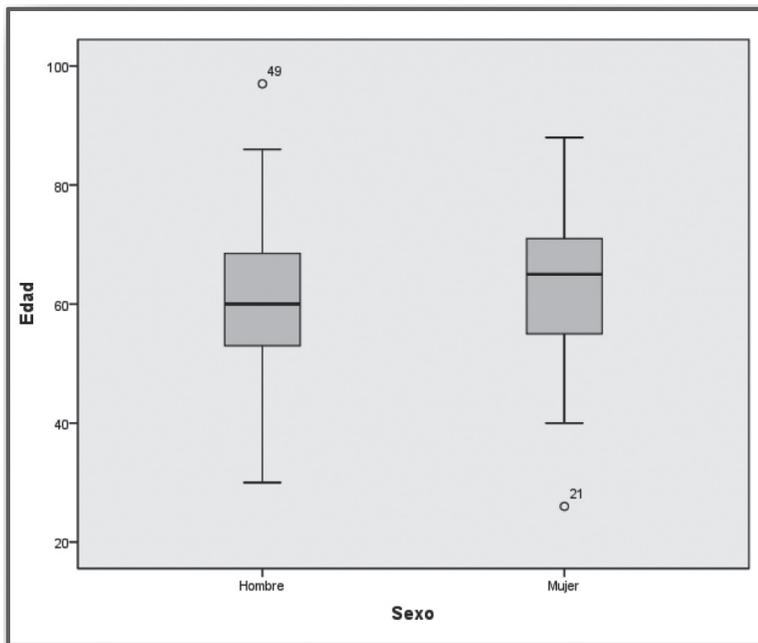


FIGURA A8.16. Gráfica de caja y bigotes de la variable 'edad' en hombres y mujeres.

## Base de datos ficticia de cirrosis

iden	sex	Ed	etio	cchp	pcp	meld	asc	Hb	tgo	tgp	urea	comp	esthosp	Alta
1	1	74	3	4	8	13	1	12,8	75	39	36,0	1	6	0
2	0	50	1	3	7	9	1	9,0	37	30	75,0	0	8	0
3	0	51	1	3	9	10	3	6,8	42	27	24,0	1	19	0
4	0	55	8	3	10	15	3	5,3	95	15	161,0	0	4	0
5	0	46	1	0	9	6	3	12,9	33	29	27,0	1	7	0
6	0	64	1	3	11	21	2	5,8	38	23	57,0	0	5	0
7	0	60	1	0	11	14	3	5,6	113	66	29,0	0	6	0
8	0	61	8	1	7	12	0	8,0	50	28	46,0	0	4	0
9	1	50	1	1	13	21	3	8,0	83	47	24,0	1	2	1
10	0	30	1	0	9	13	2	7,1	61	17	37,0	1	26	0
11	1	50	1	1	13	21	3	8,0	83	47	24,0	0	2	1
12	0	62	1	1	13	20	3	11,0	55	22	20,0	0	7	0
13	0	58	2	0	7	8	3	13,9	106	64	20,0	0	13	0
14	1	58	8	3	6	26	0	14,5	43	63	176,0	0	11	0
15	1	75	8	4	10	18	0	12,5	102	79	109,0	1	3	1
16	0	60	5	0	10	15	3	14,0	172	111	31,0	0	3	0
17	0	79	1	1	14	40	2	8,7	56	32	193,0	1	3	1
18	1	54	8	3	8	15	1	8,0	64	32	54,0	0	7	0
19	1	53	8	3	11	22	1	11,4	130	113	28,0	1	15	0
20	0	30	1	0	9	13	2	7,1	61	17	37,0	1	26	0
21	1	26	8	0	9	15	3	11,9	49	11	59,0	0	9	0
22	0	53	1	0	10	16	3	5,0	44	14	27,0	0	10	0
23	1	57	8	0	8	11	3	5,0	47	13	23,0	0	15	0
24	0	53	1	1	11	18	2	6,8	98	54	98,0	0	14	0
25	0	79	1	1	14	40	2	5,0	56	32	193,0	1	3	1
26	0	70	3	3	7	9	2	11,4	70	43	23,0	1	11	0
27	0	48	2	3	7	10	0	5,9	243	147	29,0	0	5	0
28	1	66	8	3	6	8	0	4,5	20	11	85,0	0	6	0
29	0	57	1	3	9	14	3	7,3	62	26	31,0	0	6	0
30	1	71	3	1	13	24	2	8,0	30	22	110,0	0	3	0
31	0	67	8	3	8	17	0	11,5	38	17	22,0	0	12	0
32	0	68	1	0	9	15	3	9,2	50	30	46,0	1	9	0
33	1	88	8	0	9	14	3	9,0	95	35	32,0	0	7	0
34	0	48	2	3	7	10	0	5,9	243	147	29,0	0	5	0
35	1	66	8	3	6	8	0	4,5	20	11	85,0	0	6	0
36	1	45	1	0	9	14	3	11,0			20,0	0	13	0
37	0	69	1	0	10	11	3	13,6	87	44	21,0	0	10	0
38	0	62	1	1	8	9	0	11,0	112	113	46,0	1	13	0
39	1	69	8	3	8	8	2	7,0	50	36	66,0	0	6	0
40	1	57	8	0	8	11	3	5,5	47	13	23,0	0	15	0
41	0	53	1	1	11	18	2	6,8	98	54	98,0	0	14	0
42	1	71	2	0	9	17	3	12,6	35	19	61,0	0	13	0
43	1	67	8	3	9	10	3	7,7	53	43	77,0	0	10	0
44	0	70	1	4	8	9	3	5,5	71	35	13,0	0	4	0
45	0	53	1	3	6	10	0	9,5	54	17	41,0	0	5	0
46	1	73	7	1	10	27	3	11,8			213,0	1	3	1
47	1	88	8	0	9	14	3	9,0	95	35	32,0	0	7	0
48	0	76	8	4	6	10	0	13,5	64	36	39,0	1	24	0
49	0	97	8	0	10	19	2	11,0	22	15	65,0	0	11	0
50	0	63	8	3	6	10	0	12,3	30	15	25,0	0	11	0

iden	sex	Ed	etio	cchp	pcp	meld	asc	Hb	tgo	tgp	urea	comp	esthosp	Alta
51	0	70	1	1	8	14	0	5,0	87	73	34,0	0	5	0
52	0	51	1	3	9	15	1	8,7	98	44	11,0	1	9	0
53	0	58	1	1	8	13	0	15,0	73	25	29,0	0	6	0
54	1	76	7	4	9	14	3	6,4	93	49	16,0	0	11	0
55	0	70	1	1	8	13	1	6,0	55	27	21,0	1	15	0
56	0	54	1	0	7	10	3	13,4	34	38	76,0	0	7	0
57	0	55	8	0	11	18	3	5,5	80	28	39,0	0	12	0
58	0	58	1	1	8	11	0	11,7	70	32	40,0	0	7	0
59	1	58	8	3	6	8	2	10,0	42	42	16,0	0	4	0
60	0	69	1	1	7	17	3	6,0	44	34	29,0	0	4	0
61	1	75	8	3	11	24	2	8,3	28	11	72,0	0	7	0
62	0	56	1	3	7	15	0	8,0	78	102	32,0	1	11	0
63	1	67	8	1	8	14	0	6,0	70	31	19,0	0	3	0
64	1	67	8	0	9	22	3	8,0	50	22	93,0	0	5	0
65	0	56	1	1	11	18	1	8,4	76	42	33,0	1	6	0
66	0	85	1	3	8	11	3	7,7	96	32	97,0	0	13	0
67	0	50	1	3	8	11	0	12,4	42	34	26,0	0	2	0
68	0	61	8	0	9	8	3	11,8	90	99	23,0	0	7	0
69	0	44	1	3	6	14	1	4,9	27	16	29,0	0	5	0
70	0	54	1	1	14	28	3	8,7	90	11	65,0	1	12	1
71	1	63	1	0	12	30	3	6,9	77	40	51,0	1	9	0
72	1	58	8	3	6	8	2	10,0	42	42	16,0	0	4	0
73	0	59	1	1	11	21	3	8,9	102	55	36,0	1	7	1
74	0	69	1	0	10	11	3	13,6	87	44	21,0	0	10	0
75	0	36	8	3	12	15	2	6,8	92	53	56,0	1	25	0
76	0	49	1	0	11	17	3	3,8	52	10	10,0	0	6	0
77	0	65	1	3	6	12	0	7,3	47	47	76,0	0	7	0
78	0	60	1	1	13	22	3	10,4	94	59	36,0	0	7	0
79	0	56	1	3	7	15	0	8,7	78	102	32,0	1	11	0
80	0	77	5	1	7	8	0	9,5	133	100	28,0	0	7	0
81	0	65	1	3	6	14	0	7,6	80	23	46,0	0	9	0
82	0	42	1	4	15	31	3	8,6	199	21	28,0	1	3	1
83	0	54	3	3	8	14	0	6,2	157	217	74,0	0	6	0
84	0	73	1	4	5	15	0	13,0	46	24	43,0	0	8	0
85	0	70	1	4	8	9	3	10,0	71	35	13,0	0	4	0
86	0	53	1	3	6	10	0	9,5	54	17	41,0	0	5	0
87	1	52	8	4	11	27	2	8,5	137	44	32,0	0	12	0
88	1	73	3	3	9	12	0	6,2	28	26	92,0	1	9	0
89	1	86	8	4	9	15	0	10,0	60	56	44,0	1	15	1
90	1	70	8	0	7	11	3	11,4	18	30	154,0	0	5	0
91	1	60	5	3	7	11	0	9,1	79	58	66,0	0	9	0
92	0	67	8	3	11	29	3	10,2	314	143	145,0	0	4	0
93	1	68	8	3	6	7	0	7,6	28	21	58,0	0	9	0
94	0	59	1	3	5	11	0	11,3	168	131	38,0	0	9	0
95	1	67	8	1	6	11	0	14,5	64	43	37,0	0	4	0
96	0	71	8	1	7	14	0	11,8	57	41	27,0	0	3	0
97	0	49	1	1	11	25	3	10,8	196	17	48,0	1	16	0
98	0	70	1	3	6	12	0	12,9	48	20	64,0	0	7	0
99	0	58	1	3	8	8	1	6,8	54	30	108,0	0	9	0
100	1	47	6	3	7	9	0	10,5	29	20	17,0	0	8	0

iden	sex	Ed	etio	cchp	pcp	meld	asc	Hb	tgo	tgp	urea	comp	esthosp	Alta
101	1	47	8	4	7	8	0	10,5	29	20	17,0	0	8	0
102	1	62	8	3	10	10	3	9,4	15	14	48,0	0	6	0
103	1	70	8	0	11	16	2	7,4	121	59	82,0	1	12	0
104	0	66	1	1	9	15	0	10,5	26	15	39,0	0	12	0
105	1	60	8	1	10	14	2	13,4	37	16	29,0	0	6	0
106	1	70	5	0	8	10	3	9,5	38	33	26,0	0	4	0
107	1	74	2	3	6	12	0	10,0	60	38	20,0	0	4	0
108	1	83	8	3	10	17	2	6,4	55	29	82,0	0	9	0
109	1	74	8	0	11	19	3	10,0	68	31	51,0	0	10	0
110	1	65	2	3	7	11	0	8,3	47	31	39,0	0	11	0
111	0	67	5	1	6	13	0	12,3	145	106	25,0	0	4	0
112	1	71	8	3	9	12	3	5,3	67	33	39,0	1	6	0
113	0	70	1	3	11	19	3	4,0	95	16	87,0	1	11	0
114	1	52	3	4	11	26	3	8,7	77	69	138,0	1	13	0
115	0	30	2	3	10	15	3	6,0	45	86	51,0	0	11	0
116	0	55	8	0	9	12	3	12,5	41	29	37,0	0	5	0
117	0	51	1	3	7	10	0	4,0	157	65	18,0	0	7	0
118	0	69	8	3	8	22	2	8,2	42	43	77,0	0	13	0
119	1	82	8	1	7	11	0	11,0	79	84	24,0	0	11	0
120	1	76	8	0	9	13	3	11,2	45	27	46,0	1	2	1
121	0	73	1	1	12	18	3	10,6	43	40	20,0	0	3	1
122	0	86	1	1	9	14	2	10,0	78	50	74,0	0	7	0
123	1	55	6	3	5	8	0	8,9	34	18	68,0	0	6	0
124	0	68	1	3	12	20	3	12,0	45	55	32,0	1	8	0
125	0	52	1	1	12	17	3	8,7	141	60	48,0	1	26	1
126	1	53	2	3	12	28	3	8,7	46	34	44,0	1	22	1
127	1	60	1	0	15	36	2	11,0	57	20	232,0	1	9	1
128	0	55	1	3	6	14	0	10,0	47	31	36,0	0	4	0
129	1	43	5	5	9	15	3	13,5	76	52	24,0	0	7	0
130	1	79	8	1	12	17	3	12,4	45	28	26,0	1	6	0
131	0	43	1	1	13	31	3	10,0	141	99	169,0	1	1	1
132	0	64	5	0	11	26	3	5,8	43	18	83,0	1	19	0
133	0	66	8	0	10	16	3	9,1	48	15	24,0	0	4	0
134	0	35	8	0	10	14	3	9,2	75	32	36,0	1	24	1
135	0	76	4	3	13	18	1	14,0	117	78	64,0	1	21	0
136	0	50	3	3	8	12	3	9,4	198	133	83,0	1	11	0
137	0	55	8	4	8	14	3	9,6	17	26	33,0	0	16	0
138	1	52	3	4	10	27	1	13,8	30	24	75,0	0	4	0
139	0	53	2	0	13	22	3	8,7	56	30	33,0	1	14	0
140	1	60	8	4	6	9	0	8,1	27	17	28,0	0	7	0
141	0	52	4	0	13	20	3	11,6	77	36	33,0	0	12	0
142	0	60	1	4	6	16	0	13,5	41	31	58,0	0	4	0
143	1	40	6	4	10	17	3	12,0	393	255	87,0	0	9	0
144	1	55	3	3	8	15	0	12,6	72	56	47,0	0	5	0
145	1	74	7	0	8	19	3	8,7	21	16	127,0	1	13	0
146	1	60	3	1	9	10	0	12,4	80	28	18,0	0	12	0
147	0	77	5	0	8	7	2	14,0	24	21	37,0	0	8	0
148	0	64	1	1	9	18	0	7,4	37	15	83,0	0	7	0
149	0	62	2	4	14	36	3	12,0	217	169	200,0	1	3	1
150	0	68	1	3	8	11	3	9,4	35	52	80,0	0	7	1

iden	sex	Ed	etio	cchp	pcp	meld	asc	Hb	tgo	tgp	urea	comp	esthosp	Alta
151	0	51	1	3	6	14	0	7,8	49	39	70,0	0	8	0
152	1	67	1	3	9	16	1	13,0	136	75	31,0	0	6	0
153	1	60	8	3	5	28	0	3,6	21	19	29,0	0	6	0
154	1	50	8	3	7	12	0	8,6	37	24	45,0	0	14	0
155	0	80	1	3	9	13	2	11,4	22	15	25,0	0	11	0
156	1	59	1	4	12	20	3	6,9	49	12	23,0	1	26	0
157	0	64	5	0	11	26	3	5,8	43	18	83,0	1	19	0
158	0	66	8	0	10	16	3	9,1	48	15	24,0	0	4	0
159	0	35	8	0	10	14	3	9,2	75	32	36,0	1	24	1
160	0	76	4	3	13	18	1	14,0	117	78	64,0	1	21	0
161	0	50	3	3	8	12	3	9,4	198	133	83,0	1	11	0
162	0	55	8	4	8	14	3	9,6	17	26	33,0	0	16	0
163	1	52	3	4	10	27	1	13,8	30	24	75,0	0	4	0
164	0	53	2	0	13	22	3	10,0	56	30	33,0	1	14	0
165	1	60	8	4	6	9	0	8,1	27	17	28,0	0	7	0
166	0	52	4	0	13	20	3	11,6	77	36	33,0	0	12	0
167	0	60	1	4	6	16	0	13,5	41	31	58,0	0	4	0
168	1	40	6	4	10	17	3	12,0	393	255	87,0	0	9	0
169	1	55	3	3	8	15	0	12,6	72	56	47,0	0	5	0
170	1	74	7	0	8	19	3	8,7	21	16	127,0	1	13	0
171	1	60	3	1	9	10	0	12,4	80	28	18,0	0	12	0
172	0	77	5	0	8	7	2	14,0	24	21	37,0	0	8	0
173	0	64	1	1	9	18	0	7,4	37	15	83,0	0	7	0
174	0	62	2	4	14	36	3	12,0	217	169	200,0	1	3	1
175	0	68	1	3	8	11	3	9,4	35	52	80,0	0	7	1
176	0	51	1	3	6	14	0	7,8	49	39	70,0	0	8	0
177	0	71	8	1	7	14	0	11,8	57	41	27,0	0	3	0
178	0	49	1	1	11	25	3	10,8	196	17	48,0	1	16	0
179	0	70	1	3	6	12	0	12,9	48	20	64,0	0	7	0
180	0	58	1	3	8	8	1	6,8	54	30	108,0	0	9	0
181	1	47	6	3	7	9	0	10,5	29	20	17,0	0	8	0
182	1	47	8	4	7	8	0	10,5	29	20	17,0	0	8	0
183	1	62	8	3	10	10	3	9,4	15	14	48,0	0	6	0
184	1	70	8	0	11	16	2	7,4	121	59	82,0	1	12	0
185	0	66	1	1	9	15	0	10,5	26	15	39,0	0	12	0
186	1	60	8	1	10	14	2	13,4	37	16	29,0	0	6	0
187	1	70	5	0	8	10	3	9,5	38	33	26,0	0	4	0
188	1	74	2	3	6	12	0	10,0	60	38	20,0	0	4	0
189	1	83	8	3	10	17	2	6,4	55	29	82,0	0	9	0
190	1	74	8	0	11	19	3	10,0	68	31	51,0	0	10	0
191	1	65	2	3	7	11	0	8,3	47	31	39,0	0	11	0
192	1	66	8	3	6	8	0	4,5	20	11	85,0	0	6	0
193	1	45	1	0	9	14	3	11,0			20,0	0	13	0
194	0	69	1	0	10	11	3	13,6	87	44	21,0	0	10	0
195	0	62	1	1	8	9	0	11,0	112	113	46,0	1	13	0
196	1	69	8	3	8	8	2	7,0	50	36	66,0	0	6	0
197	1	57	8	0	8	11	3	10,1	47	13	23,0	0	15	0
198	0	53	1	1	11	18	2	6,8	98	54	98,0	0	14	0
199	1	71	2	0	9	17	3	12,6	35	19	61,0	0	13	0
200	1	67	8	3	9	10	3	7,7	53	43	77,0	0	10	0

# Base de datos Cáncer

# A 10

Caso	GC
1	2
2	3
3	1
4	2
5	3
6	1
7	2
8	5
9	2
10	1
11	4
12	1
13	2
14	1
15	4
16	3
17	3
18	1
19	4
20	4
21	3
22	4
23	3
24	1
25	1
26	4
27	3
28	3
29	3
30	3

Caso	GC
31	3
32	1
33	5
34	1
35	1
36	4
37	5
38	4
39	2
40	5
41	3
42	1
43	4
44	3
45	1
46	3
47	2
48	1
49	2
50	3
51	5
52	4
53	5
54	2
55	2
56	2
57	3
58	2
59	4
60	3

Caso	GC
61	2
62	1
63	2
64	2
65	4
66	2
67	5
68	3
69	5
70	1
71	4
72	4
73	4
74	1
75	3
76	1
77	4
78	1
79	3
80	4
81	2
82	4
83	2
84	3
85	4
86	2
87	1
88	2
89	1
90	3

Caso	GC
91	3
92	3
93	3
94	2
95	3
96	4
97	3
98	3
99	3
100	3
101	3
102	3
103	3
104	1
105	1
106	4
107	4
108	5
109	3
110	3
111	1
112	5

GC: grado de cáncer.



## Base de datos IMC

A  
11

Caso	Peso	Talla
1	81	1,67
2	90	1,72
3	77	1,62
4	68	1,71
5	99	1,72
6	65	1,67
7	75	1,66
8	83	1,77
9	80	1,65
10	78	1,7
11	76	1,61
12	121	1,73
13	101	1,7
14	84	1,65
15	87	1,69
16	82	1,74
17	86	1,65
18	119	1,72
19	95	1,6
20	65	1,71
21	95	1,72
22	67	1,67
23	79	1,63
24	87	1,77
25	83	1,65

Caso	Peso	Talla
26	70	1,73
27	71	1,62
28	78	1,71
29	107	1,7
30	85	1,62
31	89	1,61
32	125	1,72
33	86	1,67
34	93	1,72
35	77	1,62
36	68	1,71
37	92	1,72
38	99	1,67
39	70	1,66
40	88	1,77
41	89	1,65
42	71	1,7
43	79	1,61
44	72	1,73
45	74	1,7
46	66	1,65
47	94	1,69
48	86	1,74
49	84	1,65
50	128	1,72



## Referencias bibliográficas

1. Triola M. Estadística. 12.ª edición. Ciudad de México: Editorial Pearson Educación; 2018.
2. Hernández R, Fernández C, Baptista P. Metodología de la investigación. 6.ª edición. México D. F.: McGraw Hill; 2014.
3. López F. Manual de medicina basada en la evidencia. México D. F.: Editorial El Manual Moderno; 2004.
4. Celis A. Bioestadística. 3.ª edición. Ciudad de México: Editorial El Manual Moderno; 2014.
5. Posada Hernández GJ. Elementos básicos de estadística descriptiva para el análisis de datos. Medellín: Funlam; 2016.
6. Argimon J, Jiménez J. Métodos de investigación clínica y epidemiológica. 5.ª edición. Barcelona: Elsevier; 2019.
7. Hernández M. Epidemiología, diseño y análisis de estudio. 2.ª edición. México D. F.: Editorial Medica Panamericana; 2014.
8. MacDonald JH. Estadísticas biológicas. University of Delaware; 2023. (<https://LibreTexts.org>)
9. Pagano M, Gauvreau K. Fundamentos de bioestadística. 2.ª edición. México D. F.: Thomson Learning; 2002.
10. Pérez C. Técnicas de análisis de datos con SPSS 15. Madrid: Pearson Educación S. A. 2009.
11. Londoño J. Metodología de la investigación Epidemiología. 5.ª edición. Bogotá: Editorial El Manual Moderno; 2014
12. Dawson B, Trapp R. Bioestadística médica. 4.ª edición. México D. F.: Editorial El Manual Moderno; 2002.
13. Juez P. Herramientas estadísticas para la investigación en medicina y economía de la salud. Madrid: Editorial Centro De Estudios Ramón Areces; 2004.
14. Daniel W. Bioestadística. 4.ª edición. México D. F.: Editorial Limusa Wiley; 2001.
15. Kamath PS, Kim WR and Advanced Liver Disease Study Group. The model for end-stage liver disease (MELD). Hepatology (Baltimore, Md.), 45(3), 797-805. 2007. <https://doi.org/10.1002/hep.21563>.
16. Greenberg R, Daniels S, Flanders W, Boring J, Eley J. Epidemiología médica. 3.ª edición. México D. F.: Editorial El Manual Moderno; 2003.
17. Del Rey J, Gil A, Herruzo R, Rodríguez F. Fundamentos de epidemiología. Madrid: Editorial Universitaria Ramon Areces; 2007.





FONDO  
EDITORIAL  
COMUNICACIONAL

ISBN: 978-612-49026-8-0



9 786124 902680