

UNIVERSIDAD AUTÓNOMA DE CIUDAD JUÁREZ



ESTADÍSTICA

PARA COMPARACIONES BÁSICAS
DE GRUPOS: CON USO DE SPSS Y
CALCULADORAS EN LÍNEA

HÉCTOR FRANCISCO PONCE RENOVA

UNIVERSIDAD AUTÓNOMA DE CIUDAD JUÁREZ

JUAN IGNACIO CAMARGO NASSAR

Rector

DANIEL CONSTANDSE CORTEZ

Secretario General

SANTOS ALONSO MORALES MUÑOZ

Director del Instituto de Ciencias Sociales y Administración

JESÚS MEZA VEGA

Director General de Comunicación Universitaria

UNIVERSIDAD AUTÓNOMA DE CIUDAD JUÁREZ

**ESTADÍSTICA
PARA
COMPARACIONES
BÁSICAS DE
GRUPOS:
CON USO
DE SPSS Y
CALCULADORAS
EN LÍNEA**

HÉCTOR FRANCISCO PONCE RENOVA

D. R. © Héctor Francisco Ponce Renova

© Universidad Autónoma de Ciudad Juárez
Avenida Plutarco Elías Calles 1210
Fovissste Chamizal, C. P. 32310
Ciudad Juárez, Chihuahua, México
Tels. +52 (656) 688 2100 al 09

Ponce Renova, Héctor Francisco
Estadística para comparaciones básicas de grupos: con uso de SPSS y calculadoras en línea / Héctor Francisco Ponce Renova. – Primera edición – Ciudad Juárez, Chihuahua, México: Universidad Autónoma de Ciudad Juárez, 2021.

178 páginas; 22 centímetros.

ISBN: 978-607-520-427-7

Contenido: Agradecimientos y reconocimientos.— Introducción.— Objetivos de aprendizaje.— Términos clave.— Comparación de promedios utilizando la distribución normal.— Test de comparación de grupos con la distribución t de estudiante.— Soluciones clave a los problemas del capítulo 1.— Soluciones clave a los problemas del capítulo 2.— Apéndice A: Generalidades del uso de SPSS para las comparaciones de grupo.— Apéndice B: Comparaciones con una sola muestra vs. Una población.— Apéndice C: Test t de dos muestras independientes.— Apéndice D: : Test t de dos muestras dependientes.— Apéndice E: Fórmulas para estimar la magnitud del tamaño del efecto.— Apéndice F: Intervalo de confianza para el d de Cohen.— Referencias.

Estadística matemática
Métodos estadísticos — Procesamiento de datos
SPSS (Programa para computadora)

La edición, diseño y producción editorial de este documento estuvieron a cargo de la Dirección General de Comunicación Universitaria, a través de la Subdirección de Editorial y Publicaciones

Coordinación editorial:
Mayola Renova González
Cuidado editorial:
Subdirección de Editorial y Publicaciones
Diseño de cubierta y diagramación:
Karla María Rascón

Primera edición, 2021
<http://elibros.uacj.mx>



ÍNDICE

Agradecimientos y reconocimientos	11
Introducción.....	13
Resumen	13
Descripción de la obra y su relación con la investigación educativa	14
Dirigido a lectoras y lectores	16
Conocimientos previos	16
Aporte principal de la obra	17
Razón para la creación y publicación de la obra.....	18
Objetivos de aprendizaje.....	19
Términos clave.....	21
Capítulo 1	
Comparación de promedios usando la distribución normal.....	23
a. Preámbulo del Test de Significancia Estadística de la hipótesis nula...24	
b. Lógica del Test de la Hipótesis	26
c. El Test z para comparaciones con una sola muestra	31
d. Errores Tipo I y II para el Test z.....	49
Problemas para resolver del Capítulo 1.....	53
Preguntas para reflexionar	57
Opinión personal.....	58

Capítulo 2

Test de Comparación de Grupos con la distribución t

de estudiante	61
a. Generalidades de la distribución t	62
b. Ejemplo de Test t con una sola muestra.....	68
c. Test t para dos muestras independientes.....	77
d. Supuestos del Test t	81
e. Ejemplo de Test t para muestras independientes y el d de Cohen	86
f. Test t para dos muestras dependientes	95
g. Ejemplo de muestras dependientes usando el Test t y el d de Cohen..	98
Problemas para resolver del Capítulo 2.....	108
Preguntas para reflexionar	114
Opinión personal.....	115

Soluciones clave a los problemas del Capítulo 1.....	117
------------------------------------------------------	-----

Soluciones clave a los problemas del Capítulo 2.....	119
------------------------------------------------------	-----

Apéndices

Apéndice A

Generalidades del uso de SPSS para las comparaciones de grupos.....	133
---------------------------------------------------------------------	-----

Apéndice B

Comparaciones con una sola muestra vs. una población	141
------------------------------------------------------------	-----

Apéndice C

Test t de dos muestras independientes.....	145
----------------------------------------------	-----

Apéndice D

Test t de dos muestras dependientes.....	151
--------------------------------------------	-----

Apéndice E

Fórmulas para estimar la magnitud del tamaño del efecto.....	155
--------------------------------------------------------------	-----

Apéndice F

Intervalo de confianza para el d de Cohen.....	163
--------------------------------------------------	-----

Referencias.....	171
------------------	-----

ÍNDICE DE TABLAS Y FIGURAS

Capítulo 1

<i>Figura 1.1</i> Prueba con el uso del Teorema de Pitágoras.....	27
<i>Figura 1.2</i> Probabilidades de un evento	30
Tabla 1.1 Pasos para poner a prueba una hipótesis nula en comparaciones de grupos	32
<i>Figura 1.3</i> Calculadora en línea para las probabilidades calculadas.....	35
<i>Figura 1.4</i> Test de Significancia Estadística con la curva de la distribución de muestras para dos colas	36
<i>Figura 1.5</i> Intervalos de confianza del 95% para la población y una muestra	37
<i>Figura 1.6</i> Test de Significancia Estadística con la curva de la distribución de muestras para una cola	39
<i>Figura 1.7</i> Intervalos de confianza del 95%.....	40
<i>Figura 1.8</i> Test de Significancia Estadística con la curva de la distribución de muestras para una cola	42
<i>Figura 1.9</i> Intervalos de confianza del 95%.....	43
<i>Figura 1.10</i> Test de Significancia Estadística con la curva de la distribución de muestras para una cola	45
<i>Figura 1.11</i> Intervalos de confianza del 95%.....	46
<i>Figura 1.12</i> Calculadora para el $z = 1.70$	47
<i>Figura 1.13</i> Una vs. dos colas	48

Tabla 1.2 Valores críticos del Test z para una y dos colas usando la curva normal como distribución de muestras.....	49
Figura 1.14 Curva normal como distribución de muestras	50

Capítulo 2

Figura 2.1 Distribuciones t con diferentes grados de libertad (df)	64
Figura 2.2 Distribución t de muestras con sus áreas de rechazo y no-rechazo de la H_0	66
Tabla 2.1 Rechazar o no con el mismo valor $t_{\text{calculado}}$ y diferentes df	67
Tabla 2.2 Pasos para el Test t y el d de Cohen	68
Tabla 2.3 Puntajes y estadísticas de los aspirantes a ingeniería	69
Figura 2.3 Gráfica de barras de los datos de los aspirantes a ingeniería	70
Tabla 2.4 Algunos valores t y sus correspondientes probabilidades en la distribución t	71
Figura 2.4 Test de Significancia Estadística con la curva de la distribución de muestras para una cola	72
Figura 2.5 Calculadora del valor t	73
Figura 2.6 Intervalos de confianza del 95%	75
Figura 2.7 Muestras con homogeneidad y heterogeneidad de varianzas.....	85
Tabla 2.5 Datos de la clase	86
Tabla 2.6 Hoja de resultados de un Test de Levene y Test t	88
Figura 2.8 Calculadora del p para el valor t	90
Figura 2.9 Test de Significancia Estadística	91
Figura 2.10 Intervalos de confianza del 95% para la diferencia entre promedios de las muestras.....	92
Tabla 2.7 Resumen de resultados	94
Tabla 2.8 Datos y estadísticas del pretest y postest	99
Figura 2.11 Test de Significancia Estadística.....	103
Figura 2.12 Calculadora del Test t para muestras dependientes	104
Figura 2.13 Intervalos de confianza del 95%	105
Tabla 2.9 Resumen de resultados	107

Problemas para resolver del Capítulo 2

Tabla Problema 1 del Capítulo 2.....	109
Tabla Problema 2 del Capítulo 2	110
Tabla Problema 3 del Capítulo 2	111
Tabla Problema 4 del Capítulo 2.....	112
Tabla Problema 5 del Capítulo 2	113

Soluciones clave a los problemas del Capítulo 2

Figura 1. Capítulo 2, Problema 1	119
Figura 2. Capítulo 2, Problema 1	120
Figura 1. Capítulo 2, Problema 2.....	121
Figura 2. Capítulo 2, Problema 2.....	122
Figura 1. Capítulo 2, Problema 3.....	124
Figura 2. Capítulo 2, Problema 3.....	125
Figura 3. Capítulo 2, Problema 3.....	125
Figura 1. Capítulo 2, Problema 4	126
Figura 2. Capítulo 2, Problema 4.....	127
Figura 1. Capítulo 2, Problema 5.....	129
Figura 2. Capítulo 2, Problema 5.....	130
Figura 3. Capítulo 2, Problema 5.....	130

Apéndices

Apéndice A.....	133
Figura A-1. Vista de los datos y de las variables en SPSS	134
Figura A-2. Vista de la variable en SPSS	135
Figura A-3. Opciones en la vista de la variable	135
Figura A-4. Algunos de los datos de la tabla del Problema 1 del Capítulo 2.....	137
Figura A-5. Cálculo de estadísticas.....	138
Figura A-6. Tabla de resultados.....	139
Apéndice B.....	141
Figura B-1. Una sola muestra para el Test t	142
Figura B-2. Análisis de los datos de una sola muestra	142
Figura B-3. Entrando la variable y promedio de la población.....	143
Figura B-4. Hoja de resultados	144
Apéndice C.....	145
Figura C-1. Test t para grupos independientes.....	146
Figura C-2. Análisis del Test t para grupos independientes.....	147
Figura C-3. Detalles del análisis del Test t para grupos independientes.....	147
Figura C-4. Hoja de resultados del Test t para muestras independientes.....	149

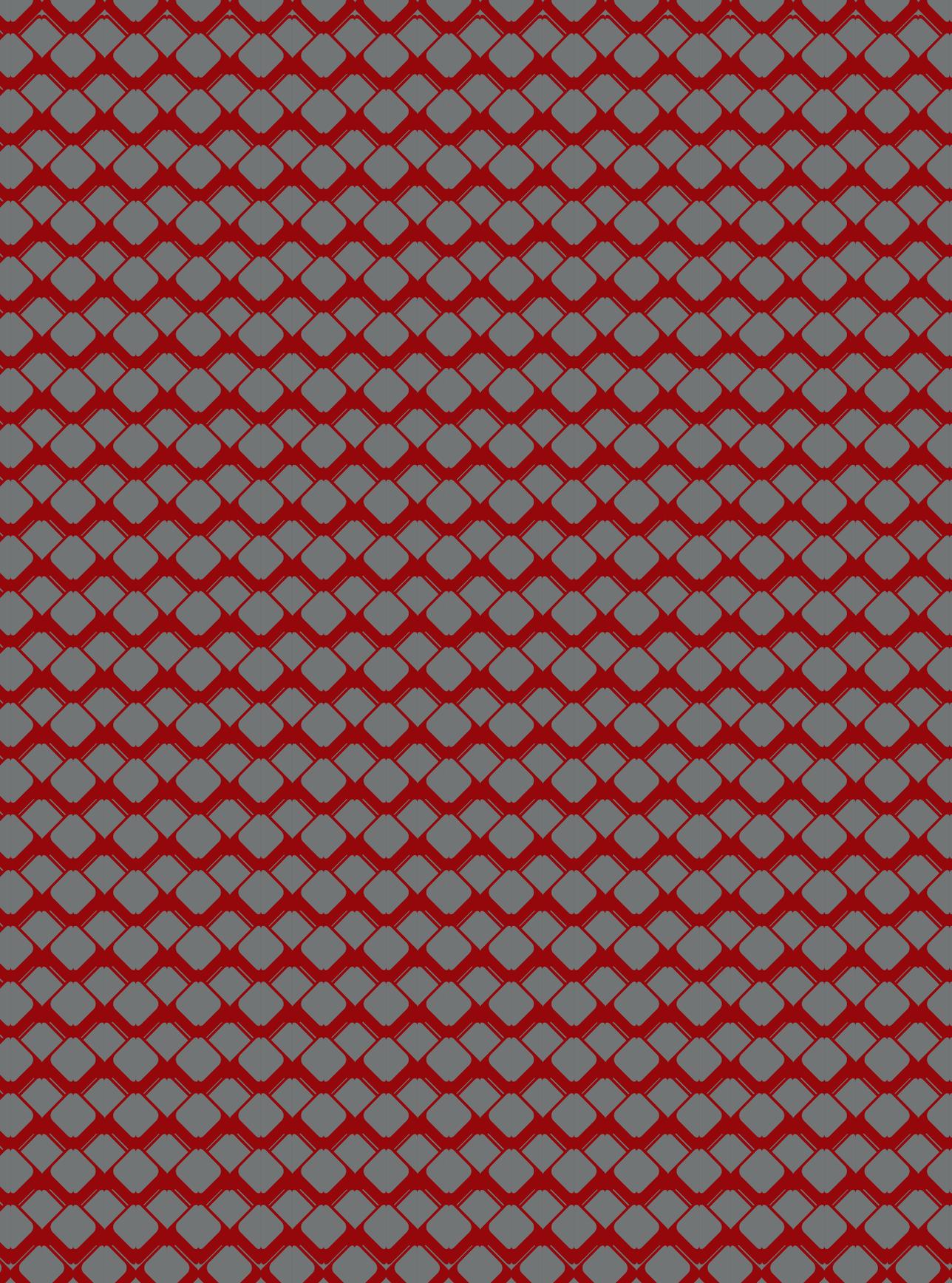
Apéndice D	151
Figura D-1. Entrada de los datos del Test <i>t</i> para muestras dependientes	152
Figura D-2. Análisis del Test <i>t</i> para muestras dependientes	153
Figura D-3. Entrada de las columnas	153
Figura D-4. Hoja de resultados de los Test <i>t</i> para muestras dependientes	154
Apéndice E	155
Tabla E-1. Fórmulas del <i>d</i> de Cohen para un solo grupo o grupos independientes.....	155
Tabla E-2. Fórmulas del <i>d</i> de Cohen para medidas repetidas	157
Tabla E-3. Efectos alternativos al <i>d</i> de Cohen	159
Tabla E-4. Conversión del valor <i>t</i> en el <i>d</i> de Cohen.....	161
Apéndice F	163
Figura F-1. Intervalo de confianza del <i>d</i> de Cohen para una sola muestra	165
Figura F-2. Intervalo de confianza del <i>d</i> de Cohen para muestras independientes.....	166
Figura F-3. Intervalo de confianza del <i>d</i> de Cohen para muestras dependientes: comparación entre calculadoras.....	168

AGRADECIMIENTOS Y RECONOCIMIENTOS

Le doy las gracias a mi familia por el tiempo y apoyo para la creación de esta obra (Isabella, Laura, Boogie, Prathiba y Simón). Asimismo, le agradezco a familiares, amigas y amigos por ser parte de mi vida. Le doy mi agradecimiento más profundo a la Universidad Autónoma de Ciudad Juárez, donde laboro, junto con mis colegas, alumnado y administración, quienes me han ayudado al desarrollo del libro; en especial, a Cely Ronquillo, por creer en hacer este texto, y al resto del equipo editorial de la Institución.

Cuidado con el hombre de un método o un instrumento, ya sea experimental o teórico. Tiende a orientarse hacia los métodos en lugar de orientarse hacia los problemas. El hombre orientado al método está encadenado; el hombre orientado a los problemas está al menos alcanzando libremente lo que es más importante. La inferencia fuerte redirige a un hombre a la orientación hacia el problema, pero requiere que esté dispuesto a dejar de lado repetidamente sus últimos métodos para enseñarse a sí mismo otros nuevos.

J. R. Platt (1964, p. 351)



INTRODUCCIÓN

RESUMEN

El libro contiene dos capítulos para comparar promedios entre grupos con las siguientes secciones en cada uno:

- ⊗ *Problemas para resolver* (se presentan cinco problemas por capítulo para reforzar los conocimientos y habilidades de los mismos, así como para expandir la discusión más allá de esta obra). Se incluyen algunas de las soluciones *clave* a estos problemas: *Soluciones clave a los problemas del capítulo*
- ⊗ *Preguntas para reflexionar* (son interrogantes que tratan de ir más allá del contenido propio del libro)
- ⊗ *Opinión del autor* (trata de explicar y asociar el contenido de cada capítulo a la experiencia y perspectiva de este investigador/profesor)

Además de estas secciones, esta obra muestra cómo resolver los ejemplos de ambos capítulos manualmente, explicando cada paso y ecuación detalladamente. También contiene seis apéndices donde se explica cómo usar el *software spss* y fórmulas para resolver algunos de los problemas. Asimismo, se utilizan calculadoras en línea para estimar

la probabilidad de las estadísticas para el primer capítulo, y, para el segundo, estas calculadoras pueden resolver el problema por completo.

En el Capítulo 1, se comienza hablando de la lógica de la hipótesis nula. Una vez establecida esta lógica, se abarcan las comparaciones con la distribución normal entre las estadísticas de una muestra con los parámetros de una población. En estas comparaciones, se dan una serie de ejemplos aplicados a la investigación educativa que son resueltos manualmente y con ayuda de una calculadora en línea. Finalmente, se repasan los Errores Tipo I y II (véase: Ponce, 2019), que pueden suceder cuando se hacen inferencias de las muestras a las poblaciones.

En el Capítulo 2, se habla de las generalidades de la distribución t . Más al respecto, se hacen comparaciones cuando solo se tiene una sola muestra, así como dos muestras independientes y dos muestras dependientes. En otras palabras, se dan una serie de ejemplos de comparaciones que son resueltos manualmente, con una calculadora en línea y con el empleo de SPSS. Por último, se muestra cómo utilizar el tamaño del efecto del d de Cohen.

DESCRIPCIÓN DE LA OBRA Y SU RELACIÓN CON LA INVESTIGACIÓN EDUCATIVA

En algunas áreas de la investigación educativa, se hace uso de la *medición* del aprendizaje, inteligencia, deserción, motivación, autorregulación y ausentismo, entre muchos otros *fenómenos*, para luego ser analizados por medio de estadísticas. En este caso, estas estadísticas sirven para comparar los promedios de dos grupos. De estos análisis pueden surgir resultados que pueden ser utilizados para crear argumentos (para argumentación en estadística, véase: Hurley, & Watson, 2018; LeMire, 2010) que puedan o no apoyar teorías del aprendizaje, como *Teoría Cognoscitiva Social*, *Teoría del Constructivismo* y *Teoría del Procesamiento de la Información*, entre otras. Con este libro, se trata de poder observar hasta dónde las mediciones del aprendizaje y otros fenómenos educativos poseen datos como una distribución normal para comparar grupos mediante la distribución z y la distribución t . Además, algunas de las preguntas que se tratan de contestar en esta obra son:

- ⊗ ¿Qué es el Test z ?
- ⊗ ¿Existe una diferencia estadísticamente significativa entre el promedio de una muestra y el de una población?
- ⊗ ¿Qué es el Test t ?
- ⊗ ¿Existe una diferencia entre dos muestras independientes?
- ⊗ ¿Existe una diferencia entre el pretest y el posttest?
- ⊗ ¿Cuál es el tamaño del d de Cohen (*effect size*)?

Al dar respuesta a estas interrogantes, se espera que el panorama de las y los lectores se expanda para poder aplicar los conceptos y análisis mostrados en el libro, para llevar a cabo investigación educativa cuantitativa.

Asimismo, el uso de la presente obra también puede ser un inicio para investigadoras e investigadores educativos en ayudar a entender y analizar la enorme cantidad de datos que se generan por medio de las redes sociales, como Facebook, Instagram y Twitter, entre otras fuentes. En otras palabras, este libro y los dos anteriores (Ponce, 2019; y Ponce, 2020) podrían ser un posible principio para hacer análisis de lo que se conoce como *big data*: este concepto engloba bases de datos muy grandes, como las que generan las ya mencionadas redes sociales, que pueden ser analizados por medio de *softwares* que ayuden a encontrar patrones, tendencias y asociaciones, especialmente relacionados con algunos comportamientos humanos e interacciones. Este concepto de *big data* ligado a otro llamado minería de datos (*data mining*: que es la práctica de examinar grandes bases de datos para generar nueva información), pueden ser de gran ayuda en la investigación educativa para resolver preguntas como los siguientes dos ejemplos: ¿existe una relación entre el aprendizaje y ciertos factores (*e. g.*, horas invertidas en las redes sociales)?; y ¿existe una diferencia en el aprendizaje entre el grupo A y el grupo B?

Más precisamente, un ejemplo de *big data*, minería de datos y la presente obra sería examinar si el aprendizaje de distintos grupos de estudiantes es diferente en relación con las horas que pasan en las redes sociales. Se podría esperar que entre más tiempo permanezca un grupo en las redes sociales menor será el aprendizaje; sin embargo, se podría pensar que cierta cantidad de contacto, a través de las

redes sociales, sea beneficioso para el aprendizaje. Esta es una mera especulación, pero más allá de lo que se encuentre en un solo estudio uno de los aspectos recomendables para hacer ciencia y obtener conocimiento es un *metaanálisis*, que reúne los resultados de múltiples estudios para ser analizados estadísticamente (Cumming, 2013). Los resultados de un metaanálisis podrían mostrar que existe un patrón, como un efecto (el d de Cohen) que se repite por medio de los resultados de varios estudios (e. g., una diferencia entre los grupos que pasan diferentes cantidades de tiempo en las redes sociales en cuestión del aprendizaje).

DIRIGIDO A LECTORAS Y LECTORES

Esta obra está dirigida a alumnas y alumnos, investigadoras e investigadores y profesoras y profesores, que deseen entender las comparaciones a través de la distribución z (normal) y la t , así como el tamaño del efecto (d de Cohen). Aunque el libro podría ser utilizado de manera independiente por aquellas personas que estén familiarizadas con ciertos principios de probabilidad y estadística, sería aún mejor que se empleara como un complemento a Ponce (2019) y Ponce (2020).

CONOCIMIENTOS PREVIOS

Se presupone que las y los usuarios de la presente obra tienen ciertos conocimientos de estadística descriptiva, probabilidad a nivel elemental, álgebra y geometría analítica. Aunque se cubre en el libro la formulación de preguntas de investigación, el poder formularlas es un pilar para poder derivar hipótesis y ponerlas a prueba por medio de análisis estadísticos. Se recomienda a Ercikan y Roth (2006), y Ponce, Domínguez y Arriaga (2016), para más información acerca de la creación de interrogantes en la investigación educativa.

La mayoría de los cálculos de la presente obra fueron ejecutados en forma manual, pero en algunos de los casos se usaron calculadoras en línea para la probabilidad calculada y tablas, así como el empleo del *software* SPSS versión 23. El cálculo manual fue para mostrar lo que

está detrás de *softwares* estadísticos, como SPSS, Minitab, SAS y R (Schumacker, & Tomek, 2013), entre otros.

APORTE PRINCIPAL DE LA OBRA

La principal aportación del libro es tratar de aumentar la base probabilística y estadística de Ponce (2019) y Ponce (2020), que sirva como un marco metodológico de la investigación educativa cuantitativa. También, la mayoría de las fuentes fueron publicadas originalmente en el idioma inglés, y, con esta obra, se tradujeron para hacerlas más accesibles a aquellas personas que prefieren el idioma español.

Más al respecto, los conceptos y análisis de la presente obra, junto con los de Ponce (2019) (intervalos de confianza; estadísticas paramétricas y no-paramétricas; variables dependientes e independientes; hipótesis nula y alternativa; Errores Tipo I y II; interpretación de resultados significativos; poder estadístico y d de Cohen) y Ponce (2020) (probabilidad; distribuciones binomial, normal y estandarizada; y correlación), forman una parte de la base de la probabilidad y estadística para entender análisis más sofisticados, como las regresiones lineales, correlaciones canónicas y regresiones logísticas, así como el análisis de la varianza, covarianza y multivariante. Asimismo, esta parte de la base probabilística y estadística sirve para entender análisis multivariados, como ecuaciones estructurales (Byrne, 2009; Jöreskog, & Sörbom, 1996) y modelos lineales jerárquicos (Raudenbush, & Bryk, 2002), entre muchos otros más. En resumen, es fundamental tener el conocimiento de este material para hacer análisis básicos, así como llegar a entender y realizar otros más sofisticados.

RAZÓN PARA LA CREACIÓN Y PUBLICACIÓN DE LA OBRA

El libro surgió como la tercera parte de Ponce (2019) y Ponce (2020). Las tres obras comenzaron como materiales didácticos de clases de métodos para la investigación educativa y las ciencias sociales. Además, estos materiales didácticos también fueron utilizados en seminarios de tesis de educación a nivel pre y posgrado. Durante casi una década, los materiales didácticos se han ido perfeccionando por las contribuciones de estudiantes y colegas para llegar a las versiones actuales. Ahora, este libro se ha convertido en una obra de divulgación, ya que existen muchos programas educativos que lo podrían usar como referencia para sus tesis y clases.

OBJETIVOS DE APRENDIZAJE

Cuando se comparan grupos puede surgir una serie de preguntas: ¿qué es una comparación de promedios?; ¿cómo se usa?; y ¿cuál es una interpretación? Se trata de responder a esta serie de preguntas. Al darles respuesta, se espera que el panorama se abra para la lectora o el lector, al poder aplicar una serie de conceptos que se desarrollan para los análisis estadísticos.

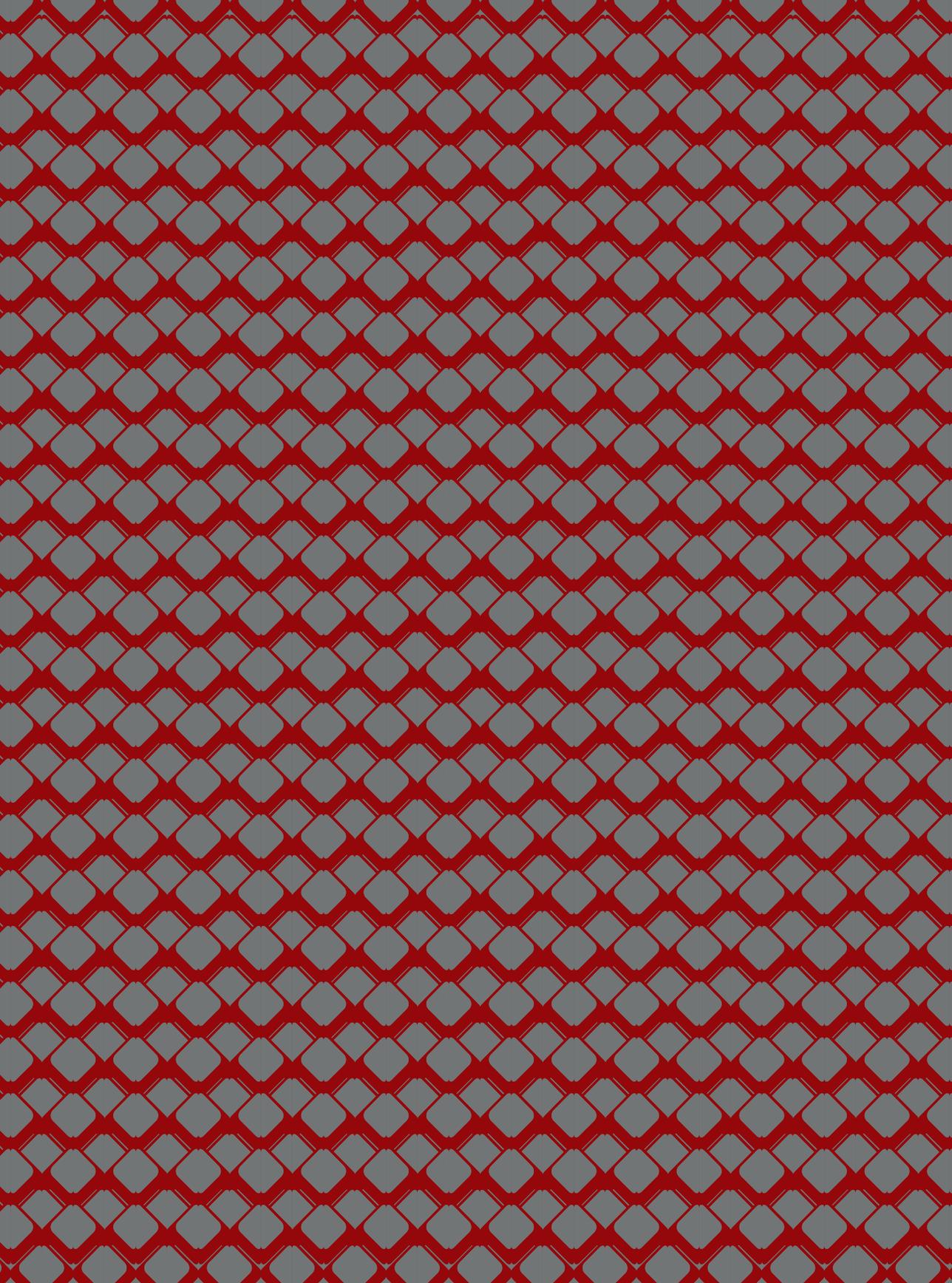
Al terminar, la lectora o el lector será capaz de:

- ⊗ Comprender la lógica del Test de la Hipótesis
- ⊗ Crear preguntas para la comparación de grupos
- ⊗ Formular hipótesis
- ⊗ Llevar a cabo un Test z para la comparación de una muestra y su población
- ⊗ Implementar una serie de pasos para el Test z
- ⊗ Comprender las características de la curva de distribución de muestras para rechazar o no una hipótesis nula
- ⊗ Calcular intervalos de confianza para el promedio
- ⊗ Comprender la congruencia entre un Test de Significancia y un intervalo de confianza
- ⊗ Realizar un Test de Significancia Estadística con una sola cola
- ⊗ Interpretar los resultados de los diferentes test
- ⊗ Aplicar los Errores Tipo I y II

- ⊗ Entender las características de la distribución t
- ⊗ Evaluar el Test de Levene
- ⊗ Comparar dos muestras independientes con Test de Significancia y tamaño del efecto
- ⊗ Comparar dos muestras dependientes con Test de Significancia y tamaño del efecto
- ⊗ Usar fórmulas para calcular los valores para los efectos estadísticos y prácticos
- ⊗ Interpretar el d de Cohen

TÉRMINOS CLAVE

- ⊗ Lógica del Test de la Hipótesis
- ⊗ Contradicción
- ⊗ Preguntas de investigación
- ⊗ Test z para la comparación de grupos
- ⊗ Distribución de muestras
- ⊗ Hipótesis nula y alternativa
- ⊗ Intervalo de confianza
- ⊗ Errores Tipo I y II
- ⊗ Familia de distribuciones t
- ⊗ Error estándar de la diferencia
- ⊗ Error estándar de la diferencia entre promedios de las muestras
- ⊗ Distribución de las diferencias de muestras
- ⊗ Test de Levene
- ⊗ δ = Promedio de las diferencias que existen entre una primera y una segunda muestras de una población
- ⊗ \bar{d} = Promedio de las diferencias que existen entre un pretest y un postest para las muestras
- ⊗ d de Cohen modificado



CAPÍTULO 1

COMPARACIÓN DE PROMEDIOS USANDO LA DISTRIBUCIÓN NORMAL

En algunas ocasiones, en la investigación educativa solo se tiene un set de datos de una sola muestra, es decir, no se cuenta con otra muestra para poder realizar una comparación entre ambas. Entonces, para poder hacer una comparación con algún otro grupo cuando solo se tiene una muestra, se podría pensar en la población de donde se obtuvo dicha muestra. Esta posible comparación sería bajo el supuesto de que se tienen datos de la población de interés. Para poder comparar una población con una muestra, se crea una pregunta de investigación como: ¿existe una diferencia estadísticamente significativa entre el promedio de una muestra y el de la población? Posiblemente, los parámetros (*e. g.*, promedio, varianza, desviación estándar, error estándar) de la población vinieron de la literatura sobre algún tema o de datos en algún reporte, entre otras muchas fuentes. Estos parámetros se consideran hipotéticos y son el punto de comparación con las estadísticas que provienen de muestras representativas (*i. e.*, seleccionar al azar y con características similares

a las de su población). Con la comparación entre ambos, se estima si la diferencia (si la hay) fue probablemente o no al azar. Esto último es el Test de Significancia Estadística.

Nota: El concepto de *diferencia estadísticamente significativa* es distinto a decir *diferencia*. Una diferencia sería simplemente: $4 - 3 = 1$ sin necesidad de usar *estadística inferencial* (véase: Ponce, 2019), y la conclusión sería: existe una diferencia entre los promedios. Por otro lado, una *diferencia estadísticamente significativa* se refiere a que la diferencia no fue al azar: por ejemplo, si $4 - 3 = 1$ con una probabilidad calculada (p) de .0001 comparada con un alfa (α) de .05, usado para rechazar o no una hipótesis nula, la conclusión sería: existe una diferencia estadísticamente significativa entre los promedios. La otra posibilidad es que $4 - 3 = 1$ con $p = .06 > \alpha = .05$. Por lo tanto, la conclusión sería: no existe una diferencia estadísticamente significativa entre los promedios y la diferencia se debió al azar.

A. PREÁMBULO DEL TEST DE SIGNIFICANCIA ESTADÍSTICA DE LA HIPÓTESIS NULA

Varios autores han explicado que la probabilidad y estadística, se dividen en tres grandes ramas:

- a) *Teoría Clásica de la Probabilidad (Classical Probability Theory)*, conocida también como *Teoría de la Frecuencia*. Esta teoría es tratada en la presente obra y fue diseñada para largo plazo, lo cual implica que un solo resultado estadísticamente significativo en un solo estudio no es suficiente para probar un efecto verdadero, porque habría que ver si existe un patrón al realizar una serie de réplicas a través del tiempo (para saber más de replicaciones, se recomienda ver a Cumming, 2013);
- b) La otra rama es llamada en inglés *likelihood* (véase: Rossi, 2018: este libro es para personas con conocimientos avanzados de matemáticas); y
- c) La probabilidad y estadística relacionadas con el Teorema de Bayes (véase: Morris, 2016: esta obra es para principiantes).

Cada una de estas ramas o paradigmas pueden ser utilizados en forma independiente de los demás, pero también pueden ser usados en forma conjunta (véase: Rossi, 2018, con *likelihoods* empleados con el Teorema de Bayes; y Tabachnick y Fidell, 2012, con regresión logística que emplea *likelihood* y la Teoría de la Frecuencia).

Dentro de la *Teoría Clásica de la Probabilidad*, se encuentra el *Test de Significancia Estadística de la Hipótesis Nula* (TSEHN; *Null Hypothesis Statistical Significance Testing*). También, se le conoce comúnmente como *Test de Significancia*. Cuando se usa un TSEHN, se toma una muestra al azar de una población. Bajo este paradigma, se pueden llevar a cabo dos objetivos con los datos de una muestra: a) Estimar los *parámetros* de una población (e. g., promedio [μ] y desviación estándar [σ], entre otros), a partir de las estadísticas de la muestra (e. g., promedio [\bar{x}] y desviación estándar [SD]); y b) Poner a prueba la hipótesis nula (TSEHN; *cfr.* Ponce, 2019). En este libro, se ponen a prueba las hipótesis nulas: i. e., conceptualmente hablando, la hipótesis nula dice para los Test de comparaciones de grupos que *no existe diferencia estadísticamente significativa* entre los promedios de un grupo y de otro (los grupos pueden ser independientes, dependientes o un solo grupo comparado con la población, como se observa más adelante). Por otro lado, una hipótesis nula también puede especificar que un promedio es *mayor* que otro, así como que un parámetro es igual a cierto valor.

La contraparte es la hipótesis alternativa: existe una diferencia estadísticamente significativa entre los promedios de los grupos (en una o dos colas: esto será explicado más adelante). La diferencia cuando no es estadísticamente significativa, se refiere a que puede haber una diferencia entre los promedios de los grupos, pero esta sucedió por mera casualidad (véase: Ponce, 2019, para más detalles). Por otro lado, la diferencia estadísticamente significativa sucede cuando el cambio es tan grande que *no se debió* a una mera casualidad. Cuando existe una diferencia estadísticamente significativa quiere decir que allí puede haber un efecto verdadero. *Grosso modo*, a esta diferencia se le otorga un valor z o t , como en esta obra, y con este valor, se obtiene una probabilidad calculada (p). Esta p se compara con cierto criterio previamente establecido: un alfa (α). Si la $p < \alpha$, se rechaza la hipótesis nula (i. e., conclusión = existe una diferencia estadísticamente signi-

ficativa entre los promedios y puede haber un efecto verdadero); y si la $p \geq \alpha$, no se rechaza la hipótesis nula (conclusión = no existe una diferencia estadísticamente significativa entre los promedios; lo anterior, se puede decir: fallar en rechazar la hipótesis nula). Este tema se explica más adelante con más detalles. Asimismo, Ponce (2019) ahondó en este tópico, así como en los errores al hacer inferencias sobre los resultados estadísticos (Errores Tipo I y II: estos se ven brevemente también en el presente libro). Al respecto, el *poder estadístico* (probabilidad de rechazar una hipótesis nula cuando es falsa) es fundamental para detectar un efecto verdadero (véase: Cohen, 1988; Ponce, 2019).

B. LÓGICA DEL TEST DE LA HIPÓTESIS

El poner a prueba una hipótesis nula (H_0) tiene cierta lógica basada en la lógica matemática, según varias investigadoras e investigadores. Esta lógica matemática sirve para probar teoremas (véase: Devlin, 2012; Hurley, 2014). Con argumentos matemáticos, se puede probar si un teorema es cierto o falso. Un ejemplo de lo anterior es el Teorema de Pitágoras, que puede ser puesto a prueba para ver si es cierto o falso usando la lógica matemática. En el argumento matemático que sigue solo se toma en cuenta un triángulo rectángulo, pero para probar si el Teorema de Pitágoras es cierto o falso, se tendría que poner a prueba para todos los triángulos rectángulos y no solamente para uno, como en este caso.

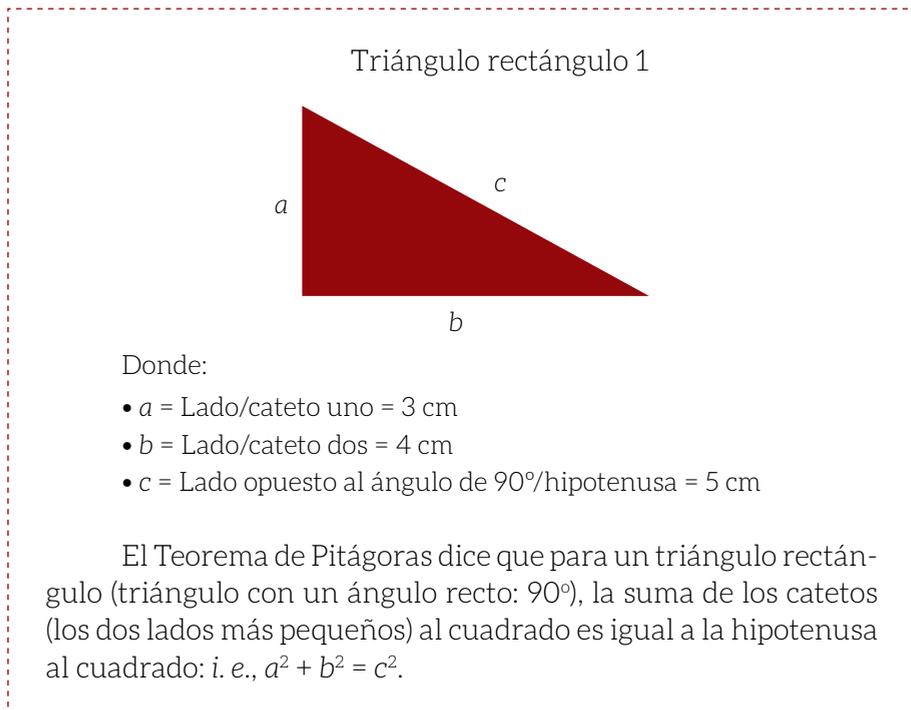
Premisa 1: Existe un triángulo rectángulo (1 en la Figura 1.1) con su ángulo de 90° entre dos de sus lados (catetos: $a = 3$ cm y $b = 4$ cm) y los otros dos ángulos; entre estos catetos e hipotenusa (i. e., lado apuesto al ángulo de 90° : $c = 5$ cm) suman también 90° .

Premisa 2: El Teorema de Pitágoras dice que un triángulo rectángulo es: $a^2 + b^2 = c^2$ (la comprobación de esta premisa, se muestra en la parte inferior de la Figura 1.1).

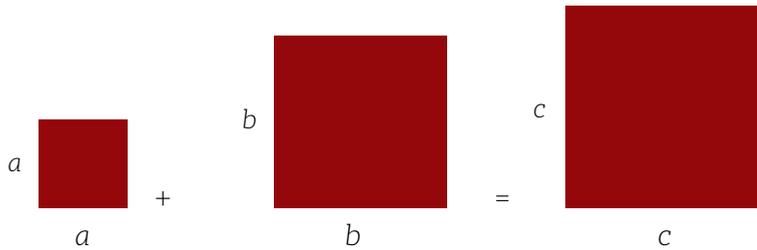
Conclusión: Por lo tanto, la hipotenusa al cuadrado de este triángulo rectángulo (1) es igual a la suma del cuadrado de cada uno de los lados.

En la Figura 1.1, se muestran las figuras y cálculos para probar este argumento matemático. En una narrativa de este argumento anterior, se podría decir como premisa y conclusión: se tiene un triángulo rectángulo y se aplica el Teorema de Pitágoras, y, entonces, el cuadrado de su hipotenusa será igual a la suma del cuadrado de sus lados. Aplicando lo anterior a un argumento de comparación de grupos, se tendría como premisa que existe una diferencia estadísticamente significativa entre los promedios de dos grupos. Como conclusión se tendría: Por lo tanto, esta diferencia *probablemente* no fue al azar. La palabra probablemente alude a la posibilidad de cometer algún error (i. e., Errores Tipo I y II; véase: Cohen, 1988). Toda esta cuestión de desarrollar comparaciones de grupos utilizando argumentos matemáticos, se despliega a lo largo de la presente obra.

Figura 1.1 Prueba con el uso del Teorema de Pitágoras



Continúa...



El área de un cuadrado es el producto de la multiplicación de un lado por el otro. El área de la suma de los cuadrados a y b es igual a la de c . Esto es:

$$\begin{aligned} &= (a \times a) + (b \times b) = (c \times c) \\ \text{Simplificando:} &= a^2 + b^2 = c^2 \\ \text{Sustituyendo:} &= (3 \text{ cm})^2 + (4 \text{ cm})^2 = (5 \text{ cm})^2 \\ \text{Elevando al cuadrado y sumando la parte de los catetos:} &25 \text{ cm}^2 = 25 \text{ cm}^2 \end{aligned}$$

A diferencia de la lógica matemática con sus pruebas en los teoremas, en la probabilidad y estadística (Teoría de la Frecuencia Relativa de la Probabilidad [*Relative Frequency Theory of Probability*]; Hurlley, 2014, p. 570) no se puede comprobar que una hipótesis es cierta o falsa, porque solo se puede concluir si esta se sostiene basándose en la evidencia o no. La razón es que se suele trabajar con muestras para hacer inferencias a una población y esto implica errores, como el error de muestreo, que es la diferencia entre el valor de un parámetro de una población y la estadística de una muestra. En otras palabras, puede pasar que cuando se quiera hacer la inferencia de la muestra a la población se caiga en un error. Por ejemplo, se obtiene un promedio de calificaciones de una escuela primaria (muestra) de 8.5 con una desviación estándar de 1 y se generaliza a la población de escuelas, pero la población tuvo un promedio de 7, así que esta inferencia no fue del todo precisa. También, están los Errores Tipo I y II que impiden hacer una inferencia precisa sobre la hipótesis nula (véase: Ponce, 2019).

En la lógica matemática hay varios métodos para poner a prueba los teoremas. Según Hinkle, Wiersma y Jurs (2003), en la probabilidad y estadística se emplea un método similar al de la *prueba indirecta*, como en la lógica matemática. Con la prueba indirecta, se elimina una premisa cuando se muestra o se lleva a una contradicción de un hecho conocido. Más al respecto, Hurley (2014) dijo que: “En la prueba indirecta, se formula un supuesto dentro de una secuencia de razonamientos de prueba, y cuando la secuencia lleva a una contradicción, se concluye que el supuesto es falso, así la negación es cierta” (p. 443). Por ejemplo, si una persona va por un camino para visitar a un pariente y, de pronto, se bifurca (no recuerda si es a la derecha o a la izquierda, porque ha pasado tiempo desde que visitó a su familiar), entonces decide tomar el camino hacia la izquierda y el supuesto es que este lo llevará a la casa de su pariente. Lo sigue como lo haría en una serie de razonamientos en la lógica matemática. Al cabo de un tiempo, el camino se acaba y no llega a la casa de su familiar; por lo tanto, se cerciora de que el camino hacia la izquierda no era el indicado y, entonces, el camino hacia la derecha es el que lo lleva a la casa.

De una manera similar, se parte de que la hipótesis nula es *cierta* para ver si la evidencia la apoya. La hipótesis nula establece en las comparaciones de grupos que el promedio de un grupo es igual al del segundo grupo. La hipótesis nula es rechazada cuando la probabilidad de la diferencia entre dos promedios (i. e., probabilidad calculada: p_{cal} , valor p o simplemente p) es pequeña relativamente al criterio (alfa: i. e., α) al que se compara bajo el supuesto de una hipótesis nula cierta: e. g., el promedio de una población es igual al promedio de otra (existe una diferencia estadísticamente significativa y esto promete la posibilidad de un efecto verdadero). De este modo, si se rechaza la hipótesis nula, se apoya la alternativa; es decir, la hipótesis alternativa es apoyada al rechazar la hipótesis nula. Lo anterior es similar al lanzar una moneda al aire: puede caer cara o cruz. Esto se resume en que la evidencia apoya o no a la hipótesis nula.

La Figura 1.2 muestra las probabilidades de un evento, que puede ser el lanzar una moneda al aire o comparar el promedio de dos grupos diferentes. Más precisamente, puede ser una diferencia entre el promedio de calificaciones de un grupo de damas y un grupo de varones.

Esto es a manera de recordatorio de los libros de Ponce (2019; 2020). Cuando la probabilidad es *cercana* a cero, se considera que *el evento es poco probable que ocurra*. Tradicionalmente, esta consideración de escasa probabilidad es de menos de .05 (i. e., 5%), que es un criterio empleado para rechazar la hipótesis nula (i. e., a este criterio también se le llama alfa). *Recordando*: un criterio es un área bajo la curva normal y puede tomar los valores tradicionales de 1%, 5% y 10% (véase: Ponce 2019). Por ello, cuando la $p_{calculada}$ ($p_{cal.}$), o simplemente p (i. e., otra área bajo la curva normal), es menor a este criterio, se considera poco probable que el evento en cuestión haya ocurrido al azar cuando la hipótesis nula es cierta (*recordando*: la hipótesis nula es cierta cuando los promedios de dos poblaciones diferentes son iguales, pero esto difícilmente se sabrá, porque se trabaja con muestras). Una advertencia es que el criterio y la $p_{cal.}$ también se encuentran en distribuciones que no son normales: e. g., distribución F (i. e., para el análisis de la variación; véase: Berenson, Levine, & Krehbiel, 2012; Hinkle *et al.*, 2003; Salkind, 2007; 2011), entre otros.

Figura 1.2 Probabilidades de un evento



Otro escenario es que un evento tenga la *misma* probabilidad o no de acontecer: *el evento es tan probable como no probable que suceda*. Volviendo a la $p_{cal.}$, si esta es de .50 (50%), se considera que el evento ocurrió al azar: e. g., lanzar una moneda al aire y tratar de adivinar lo que caerá. Si se usa como criterio el 5%, la $p_{cal.}$ del 50% es, por mucho, más grande que el criterio (Figura 1.2). También, hay eventos que casi siempre acontecen: *el evento es casi seguro que ocurra*. Un ejemplo es cuando el servi-

cio meteorológico pronostica que hay un .9999 (99.99%) de lluvia para la tarde; otras cosas siendo iguales (*ceteris paribus*). Como está cercano al 1 (i. e., 100%) es casi seguro o muy probable que suceda. También, en este caso, la $p_{cal.}$ (99.99%) es mucho mayor que el criterio del 5%. Un recordatorio es que cuando la $p_{cal.}$ es menor al criterio existe significancia estadística (i. e., la diferencia *probablemente* no se debió al azar). Por otro lado, cuando la $p_{cal.}$ es mayor o igual al criterio no existe significancia estadística (i. e., la diferencia *probablemente* se debió al azar).

C. EL TEST Z PARA COMPARACIONES CON UNA SOLA MUESTRA

Usualmente en la investigación educativa, se tienen preguntas de investigación que involucran comparaciones de grupos, observaciones u objetos: e. g., ¿existe una diferencia estadísticamente significativa entre el promedio de una población y el de una muestra?; ¿existe una diferencia estadísticamente significativa entre los promedios de las y los estudiantes universitarios?; ¿existe una diferencia estadísticamente significativa entre el pretest y el posttest?; y ¿existe una diferencia estadísticamente significativa entre el promedio del grupo experimental y el de control? Para dar respuesta a cada una de estas interrogantes, se pueden utilizar varios tipos de análisis (e. g., Test z, Test t y Test F, entre otros). Uno de ellos es el Test z, que implica saber cuatro aspectos para llevarlo a cabo (*cfr.* Hinkle *et al.*, 2003):

1. Promedio de la población (μ).
2. Desviación estándar de la población (σ).
3. Promedio de la muestra (\bar{x}).
4. Tamaño de la muestra (n).

Una vez que se tienen estos datos, se puede pasar a la fórmula del Test z (Ecuación 1.1):

$$Z_{calculado} = \frac{(\bar{x} - \mu)}{\sigma_{\bar{x}}} \quad (\text{Ecuación 1.1})$$

Donde:

- ⊗ $z_{\text{calculado}}$ = Coeficiente en la curva normal (distribución de muestras) para ser comparado con el valor z_{critico} y decidir si la H_0 , se rechaza o no. También, se le conoce en este contexto como estadística de prueba (*Test Statistic*)
- ⊗ \bar{x} = Promedio de la muestra
- ⊗ μ = Promedio de la población
- ⊗ $\sigma_{\bar{x}}$ = Error estándar de la distribución de muestras: también conocido para las muestras como $SE_{\bar{x}}$

La Ecuación 1.2 muestra cómo calcular el error estándar de la distribución de muestras:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad (\text{Ecuación 1.2})$$

Donde:

- ⊗ σ = Desviación estándar de la población
- ⊗ n = Tamaño de la muestra

Recordando de Ponce (2019) y Ponce (2020), se muestran los pasos para poner a prueba una hipótesis nula (Tabla 1.1).

Tabla 1.1 Pasos para poner a prueba una hipótesis nula en comparaciones de grupos

1.º Establecimiento de la o las preguntas de investigación.
2.º Derivar las hipótesis de la pregunta de investigación (una vs. dos colas).
3.º Establecer el criterio para rechazar la hipótesis nula: el nivel del alfa (α); e. g., $\alpha = .01$; $\alpha = .05$; o $\alpha = .10$.
4.º Calcular el valor de la prueba estadística (<i>Test Statistic</i>); i. e., en este caso, el $z_{\text{calculado}}$ y su correspondiente probabilidad <small>calculada</small> .
5.º Llevar a cabo un intervalo de confianza.
6.º Interpretar los resultados.

Nota: También, es recomendable calcular las estadísticas descriptivas y cerciorarse de la distribución normal de los datos (véase: Ponce, 2019).

Ejemplo uno de comparación de grupos con una sola muestra

Un ejemplo heurístico de lo anterior es una investigación en la cual se plantea la siguiente interrogante (1.º *Establecimiento de la o las preguntas de investigación*): ¿existe una diferencia estadísticamente significativa entre el promedio del coeficiente intelectual de una muestra y el de una población? Además, se sabe que el coeficiente intelectual de la población es de un promedio de 100 puntos con una desviación estándar (σ) de 15 puntos en el Test de Inteligencia de Wechsler (véase: Wechsler, 1997). Se toma una muestra ($n = 31$) de la población al azar y esta resulta que tiene un coeficiente intelectual de 115 puntos.

De la pregunta de investigación, se derivan las hipótesis (2.º *Derivar las hipótesis de la pregunta de investigación* [una vs. dos colas]). La hipótesis nula se puede expresar como que el promedio de la población y el de la muestra son iguales: i. e., $H_0: \mu = \bar{x}$. Otra versión de la hipótesis nula es que, al restar el promedio de la población y el de la muestra, resulta en cero: $H_0: \mu - \bar{x} = 0$. Otra versión de la hipótesis nula es: $H_0: \mu = 100$. La hipótesis alternativa con dos colas muestra que el promedio de la población y el de la muestra *no* son iguales: $H_A: \mu \neq \bar{x}$. También, se puede expresar como que la diferencia entre ambos promedios no es igual a cero: i. e., $H_A: \mu - \bar{x} \neq 0$. Una última opción en este caso es: $H_A: \mu \neq 100$. En síntesis, las hipótesis quedan de la siguiente manera:

Nulas:

$$H_0: \mu = \bar{x}$$

$$H_0: \mu - \bar{x} = 0$$

$$H_0: \mu = 100$$

Alternativas:

$$H_A: \mu \neq \bar{x}$$

$$H_A: \mu - \bar{x} \neq 0$$

$$H_A: \mu \neq 100.$$

Ahora, se puede establecer el criterio para rechazar o no la hipótesis nula (3.º *Establecer el criterio para rechazar la hipótesis nula: el nivel del alfa* [α]; e. g., $\alpha = .01$; $\alpha = .05$; o $\alpha = .10$). Lo usual es un Coeficiente Alfa

de .05, así que este se adopta para este ejemplo. Habría que recordar que antes de que se tomara la muestra de la población, no se sabía si la muestra iba a tener un promedio mayor, igual o menor al de la población. Por tal motivo, se toma el Test de dos colas, porque no es necesario especificar en la hipótesis alternativa cuál promedio es mayor que el otro. Como el test es de dos colas, este valor del alfa se divide entre dos: $.05 / 2 = .025$ por cola.

Para el Test de Significancia Estadística, se va a una Tabla de Valores z (acceder al siguiente enlace para usar una Tabla z, así como una calculadora: <http://www.z-table.com/>) y se encuentra que el valor correspondiente es: $z_{\text{critico}} = 1.96$ (4.º *Calcular el valor de la prueba estadística [Test Statistic]*; i. e., en este caso, el $z_{\text{calculado}}$ y su correspondiente probabilidad $_{\text{calculada}}$). Antes de poder emplear la fórmula para el $z_{\text{calculado}}$, se calcula el valor del error estándar de la distribución de muestras ($\sigma_{\bar{x}} = \sigma / \sqrt{n}$) de la siguiente manera:

$$\begin{aligned}\sigma_{\bar{x}} &= 15 / \sqrt{31} \\ \sigma_{\bar{x}} &= 15 / 5.57 \\ \sigma_{\bar{x}} &= 2.69.\end{aligned}$$

Una vez que se calculó el error estándar de la distribución de muestras, se utiliza la fórmula: $z_{\text{calculado}} = (\bar{x} - \mu) / \sigma_{\bar{x}}$; con los siguientes datos: $\bar{x} = 110$ y $\mu = 100$:

$$\begin{aligned}z_{\text{calculado}} &= 110 - 100 / 2.69 \\ z_{\text{calculado}} &= 10 / 2.69 = 3.72.\end{aligned}$$

La comparación indica que el $z_{\text{calculado}} = 3.72 > z_{\text{critico}} = 1.96$. Usando una Tabla z, se puede aproximar el valor de la probabilidad $_{\text{calculada}}$. Sin embargo, esta probabilidad $_{\text{calculada}}$ se puede calcular empleando una calculadora, como la siguiente: <http://www.socscistatistics.com/pvalues/normaldistribution.aspx>, con los datos de un $\alpha = .05$ y dos colas (Figura 1.3).

Figura 1.3 Calculadora en línea para las probabilidades calculadas

P Value from Z Score Calculator

This is very easy: just stick your Z score in the box marked Z score, select your significance level and whether you're testing a one or two-tailed hypothesis (if you're not sure, go with the defaults), then press the button!

If you need to derive a Z score from raw data, [you can find a Z test calculator here](#).

Z score:

Significance Level:

- 0.01
 0.05
 0.10

One-tailed or two-tailed hypothesis?:

- One-tailed
 Two-tailed

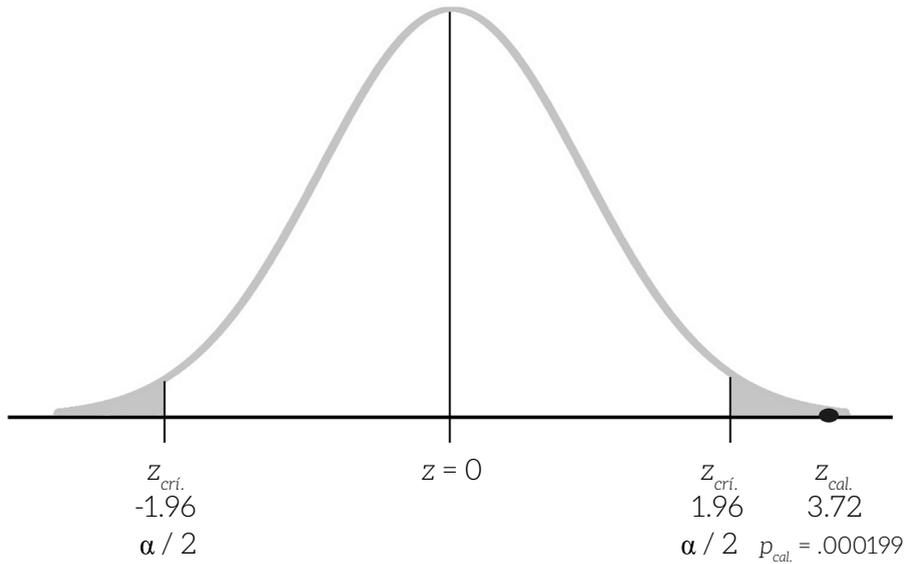
The P-Value is .000199.

The result is significant at $p < .05$.

Fuente: Social Science Statistics.

Con esta calculadora, se puede convertir un valor z en una $p_{cal.}$. Por ejemplo, el $z_{calculado} = 3.72$ se convierte en una $p_{cal.} = .000199$. También, esta $p_{cal.}$ sirve para el Test de Significancia Estadística: i. e., $p_{cal.} = .000199 < .025$ (i. e., $\alpha = .05 / 2$; véase Figura 1.4). Esto indica que la probabilidad de que el promedio de la muestra (110) ocurra al azar es menor a $\alpha / 2$ si la hipótesis nula era cierta ($H_0: \mu = 100$). *Recordando otra interpretación:* la hipótesis nula es cierta cuando el promedio de ambas poblaciones es la misma (véase: Ponce, 2019). En este caso en particular no se tienen dos poblaciones, sino una población y una muestra donde la hipótesis nula indica que los promedios de estas dos son iguales.

Figura 1.4 Test de Significancia Estadística con la curva de la distribución de muestras para dos colas



Nota: Valor $z_{\text{crítico}} = z_{\text{cri.}}$. La distribución de muestras tiene un promedio de un valor $z = 0$ y el error estándar = 1. El área de rechazo de la hipótesis nula es la sombreada. Esta área representa valores poco probables cuando la hipótesis nula es cierta. Usualmente para las dos colas, el alfa es de .05, que se divide entre dos (i. e., $\alpha / 2 = .025$), y su valor $z_{\text{crítico}} = |1.96|$ (las líneas verticales indican un valor absoluto). Por ejemplo, $|x|$ es igual a $-x$ y $+x$.

Además de poner a prueba la hipótesis nula, se puede estimar un intervalo de confianza (IC) (5.º Llevar a cabo un intervalo de confianza), que permite estimar el rango en el que estaría cierto porcentaje de las muestras. También, Cumming (2013) ha enfatizado el uso de los intervalos para considerar la magnitud de los errores de muestreo. Se puede obtener un IC para el promedio de la población de la siguiente manera (Ecuación 1.3 y Ecuación 1.4):

Límite bajo (*lower limit*): $\bar{x} - \text{Valor } z_{\text{crítico}} \times \text{Error estándar}$ (Ecuación 1.3)

Límite alto (*upper limit*): $\bar{x} + \text{Valor } z_{\text{crítico}} \times \text{Error estándar}$ (Ecuación 1.4)

Nota: Recordando: un intervalo de confianza se especifica empleando corchetes y una coma entre los límites: e. g. [límite bajo, límite alto].

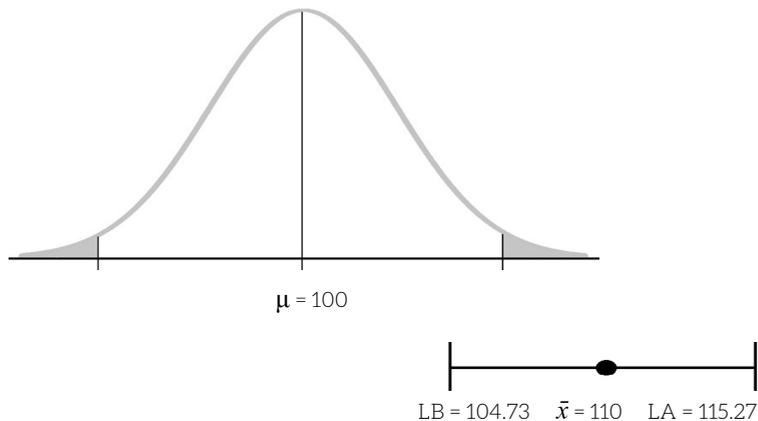
Con estas fórmulas, se puede calcular un $IC_{95\%}$ con el promedio de la muestra. Esto es:

$$\text{Límite bajo (lower limit): } 110 - 1.96 \times 2.69 = 104.73$$

$$\text{Límite alto (upper limit): } 110 + 1.96 \times 2.69 = 115.27.$$

El IC del 95% [104.73, 115.27] indica que, dadas las muestras de 31 participantes, unas 95 de 100 muestras estarían entre estos dos límites. La Figura 1.5 muestra estos dos IC. Como se puede apreciar, el IC de la muestra no captura el promedio hipotético de la población ($\mu = 100$).

Figura 1.5 Intervalos de confianza del 95% para la población y una muestra



En conclusión, se rechaza la hipótesis nula y se dice que el promedio de la muestra no es igual a 100 (6.º *Interpretar los resultados*). Este valor del test estadístico indica que el promedio de la muestra ($\bar{x} = 110$) está a 3.72 errores estándar por encima del promedio de la población ($\mu = 100$). En otras palabras, se considera que el promedio de la muestra es significativamente diferente del valor hipotético de la po-

blación a un nivel de significancia del $\alpha / 2$. En resumen, es muy poco probable que ocurra esta diferencia entre la población y el promedio al azar cuando la hipótesis nula es cierta. La conclusión anterior es apoyada por el $IC_{95\%}$ de la muestra, porque no incluye al parámetro de la población.

Ejemplo dos de comparación de grupos con una sola muestra

Otro ejemplo heurístico similar al anterior es cuando se tiene una población con el mismo nivel de coeficiente intelectual promedio de 100 puntos con una desviación estándar de 15 puntos, pero, en este caso, la muestra ($n = 31$) tiene un promedio de 105 puntos. Estos son los pasos para poner a prueba la hipótesis nula:

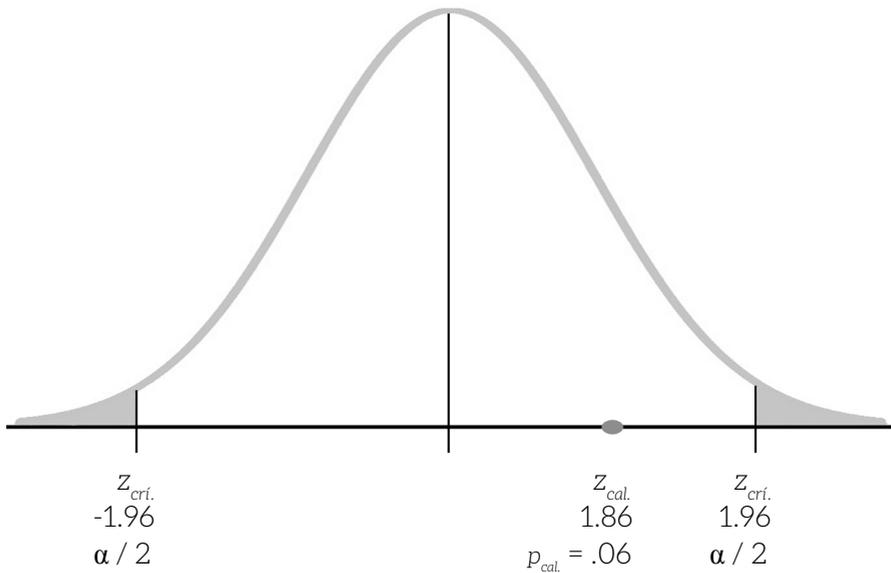
1. *La pregunta de investigación es la misma: ¿existe una diferencia estadísticamente significativa entre el promedio del coeficiente intelectual de una muestra y el de una población?*
2. *Establecer las hipótesis: nula y alternativa (una vs. dos colas): se adopta en un Test de dos colas, para que vaya de acuerdo con el ejemplo anterior:*

$$H_0: \mu = \bar{x}$$

$$H_A: \mu \neq \bar{x}$$

3. *Establecer el criterio para rechazar la hipótesis nula. El criterio es el tradicional: $\alpha = .05$.*
4. *Calcular el valor del test estadístico: i. e., en este caso, el $z_{\text{calculado}}$ y su correspondiente p_{cal} . El error estándar es el mismo del ejemplo anterior, porque se mantuvieron los mismos valores: $\sigma = 15$ y $n = 31$; entonces, el $\sigma_{\bar{x}} = 15 / \sqrt{31} = 2.69$. Usando la fórmula del valor $z_{\text{calculado}}$ (i. e., $z_{\text{calculado}} = [\bar{x} - \mu] / \sigma_{\bar{x}}$), resulta: $z_{\text{calculado}} = 105 - 100 / 2.69 = 1.86$. Empleando la calculadora (<http://www.socscistatistics.com/pvalues/normaldistribution.aspx>) para obtener la probabilidad $p_{\text{calculada}}$ del $z_{\text{calculado}} = 1.86$ es igual: .06. La Figura 1.6 muestra estos valores.*

Figura 1.6 Test de Significancia Estadística con la curva de la distribución de muestras para una cola



Nota: Valor z crítico = $z_{\text{cri.}}$. La distribución de muestras tiene un promedio de un valor $z = 0$ y el error estándar = 1.

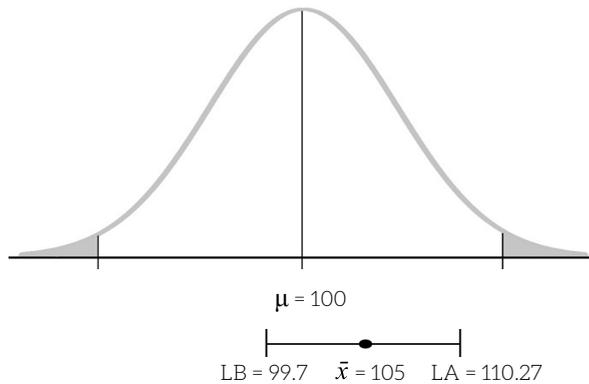
5. Llevar a cabo un intervalo de confianza:

Límite bajo (*lower limit*): $105 - 1.96 \times 2.69 = 99.73$

Límite alto (*upper limit*): $105 + 1.96 \times 2.69 = 110.27$.

El $IC_{95\%} = [99.73, 110.27]$. Este intervalo de confianza aparece en la Figura 1.7.

Figura 1.7 Intervalos de confianza del 95%



6. Decidir con base en la evidencia si se rechaza o no la hipótesis nula. Los resultados indican que el $z_{\text{calculado}} = 1.86 < z_{\text{critico}} = 1.96$; asimismo, la $p_{\text{cal.}} = .06 > \alpha / 2 = .05 / 2 = .025$. En otras palabras, el promedio de la muestra de 105 excede al de la población de 100 puntos en 1.86 error estándar, pero es menor al punto crítico de 1.96. Esta diferencia no cae en el área sombreada de rechazo a la hipótesis nula; es decir, la diferencia entre el promedio de la muestra y el de la población no es lo suficientemente grande como para rechazar la hipótesis nula. Esta diferencia podría ser atribuida a un error de muestreo; esto es, esta diferencia probablemente apareció al azar. Entonces, se falla en rechazar o, en otras palabras, no se rechaza la hipótesis nula.

El IC del 95% [99.73, 110.27] indica que, dadas las muestras de 31 participantes, unas 95 de 100 muestras capturarían el valor verdadero de la población. De hecho, este $IC_{95\%}$ captura al promedio de 100 entre sus límites (véase Figura 1.7).

Ejemplo tres de comparación de grupos con una sola cola y una sola muestra

Los ejemplos anteriores tienen dos colas para el Test z. En los siguientes dos ejemplos, se emplea una sola cola de la distribución de muestras, porque se espera que, en uno de los casos, el promedio de la muestra sea mayor al de la población. En el otro caso, se espera que el promedio de la muestra sea menor al de la población. Lo anterior se explica en pocas palabras:

Para la cola derecha: cuando se espera que el promedio de la muestra sea *mayor* al de la población: $\bar{x} > \mu$.

Para la cola izquierda: cuando se espera que el promedio de la muestra sea *menor* al de la población: $\bar{x} < \mu$.

Por ejemplo, se tiene un grupo de estudiantes que se sospecha que tienen *aptitudes sobresalientes*. Por esta sospecha, se asume que sus coeficientes intelectuales deben estar por encima del promedio. Los pasos para poner a prueba la hipótesis nula son:

1. *Establecimiento de la o las preguntas de investigación:* ¿existe una diferencia estadísticamente significativa entre el promedio de una muestra de alumnas y alumnos con supuestas aptitudes sobresalientes y el de una población? Los datos para poner a prueba la hipótesis nula son los siguientes:
 1. Promedio de la población: $\mu = 100$ puntos.
 2. Desviación estándar de la población: $\sigma = 15$.
 3. Promedio de la muestra: $\bar{x} = 108$.
 4. Tamaño de la muestra: $n = 100$.
2. *Derivar las hipótesis de la pregunta de investigación* (en este caso, es de una sola cola):

$$H_0: \mu = \bar{x}$$

$$H_A: \mu < \bar{x}.$$

3. Establecer el criterio para rechazar la hipótesis nula: el nivel del alfa (α); $\alpha = .05$. Buscando en una Tabla z bajo una cola y un alfa = .05, le corresponde: $z_{\text{crítico}} = 1.64$.
4. Calcular el valor de la prueba estadística (Test Statistic); i. e., en este caso, el $z_{\text{calculado}}$ y su correspondiente p_{cal} . Se calcula, primero, el error estándar de la distribución de muestras:

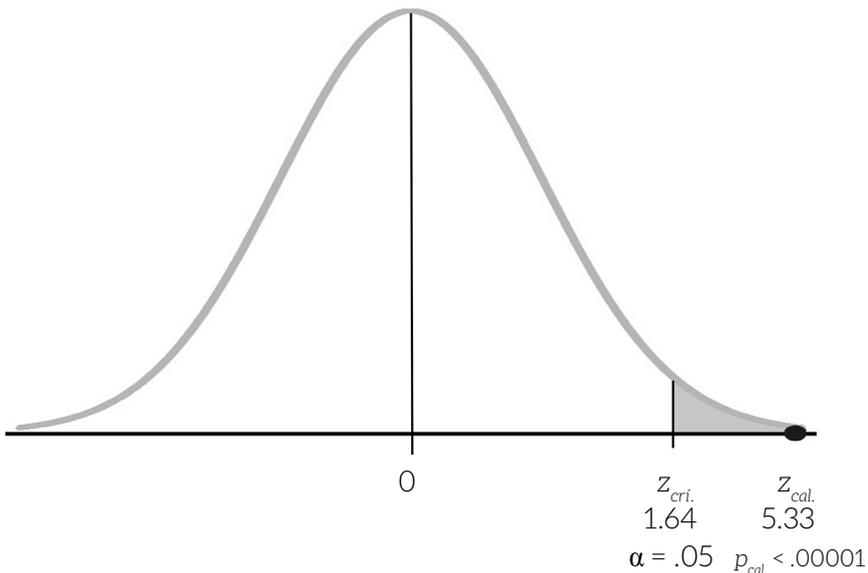
$$\sigma_{\bar{x}} = 15 / \sqrt{100} = 1.5.$$

Luego, se aplica a la fórmula:

$$z_{\text{calculado}} = 108 - 100 / 1.5 = 5.33.$$

Al valor $z_{\text{calculado}} = 5.33$, le corresponde una $p_{\text{cal}} < .00001$ al usar la calculadora (<http://www.socscistatistics.com/pvalues/normaldistribution.aspx>). La Figura 1.8 contiene una distribución de muestras con los valores antes calculados.

Figura 1.8 Test de Significancia Estadística con la curva de la distribución de muestras para una cola



Nota: Valor z crítico = $z_{\text{cri.}}$. La distribución de muestras tiene un promedio de un valor $z = 0$ y el error estándar = 1.

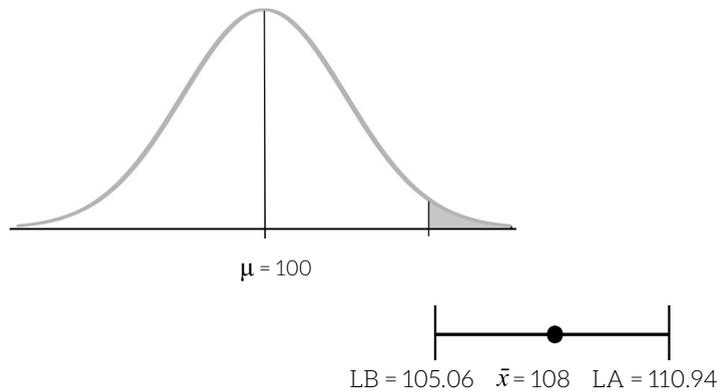
5. Llevar a cabo un intervalo de confianza:

Límite bajo (*lower limit*): $108 - 1.96 \times 1.50 = 105.06$

Límite alto (*upper limit*): $108 + 1.96 \times 1.50 = 110.94$.

El $IC_{95\%} = [105.06, 110.94]$. Este intervalo de confianza aparece en la Figura 1.9.

Figura 1.9 Intervalos de confianza del 95%



6. Interpretar los resultados. En otras palabras, se trata de decidir con base en la evidencia si se rechaza o no la hipótesis nula.

Los resultados del test indican que el $z_{\text{calculado}} = 5.33 > z_{\text{crítico}} = 1.64$; también, la $p_{\text{cal.}} < .00001 < \alpha = .05$. Por lo tanto, se rechaza la hipótesis nula y los resultados del test apoyan la hipótesis alternativa: el promedio de la muestra es mayor que el de la población. Se puede adicionar que la probabilidad de encontrar una muestra con un promedio de 108 puntos de coeficiente intelectual o más puntaje al azar, sería de menos de .00001 cuando la hipótesis nula es cierta: $H_0: \mu = \bar{x}$. En pocas palabras, la interpretación sería que existe una probabilidad de .00001 de

encontrar un promedio igual o más grande que 108 cuando la hipótesis nula es cierta.

El intervalo de confianza del 95% [105.06, 110.94] indica que, dadas las muestras de 100 participantes, unas 95 de 100 muestras capturarían el valor verdadero de la población. Este $IC_{95\%}$ no captura el promedio de 100 entre sus límites (véase Figura 1.9).

Otro ejemplo heurístico de una cola es cuando se espera que el promedio de una muestra sea menor al de una población. Esto podría pasar con algún grupo de alumnas y alumnos que, por algún motivo, se han quedado rezagados uno o dos grados escolares y un ejemplo podría ser por razones académicas únicamente.

1. *Establecimiento de la o las preguntas de investigación:* ¿existe una diferencia estadísticamente significativa entre el promedio de una muestra de estudiantes con *rezago escolar* y el de una población? Los datos para poner a prueba la hipótesis nula son los siguientes:

1. Promedio de la población: $\mu = 100$ puntos.
2. Desviación estándar de la población: $\sigma = 15$.
3. Promedio de la muestra: $\bar{x} = 95$.
4. Tamaño de la muestra: $n = 32$.

Los pasos para poner a prueba la hipótesis nula son:

2. *Derivar las hipótesis de la pregunta de investigación* (en este caso, es de una sola cola):

$$H_0: \mu = \bar{x}$$

$$H_A: \mu > \bar{x}.$$

3. *Establecer el criterio para rechazar la hipótesis nula:* el nivel del alfa (α); $\alpha = .05$. Buscando en una Tabla z bajo una cola y un alfa = .05, le corresponde: $z_{\text{crítico}} = -1.64$. En este caso, es un valor crítico negativo, porque está del lado izquierdo de la distribución de muestras.

4. Calcular el valor de la prueba estadística (Test Statistic); i. e., en este caso, el $z_{\text{calculado}}$ y su correspondiente p_{cal} . Se calcula, primero, el error estándar de la distribución de muestras:

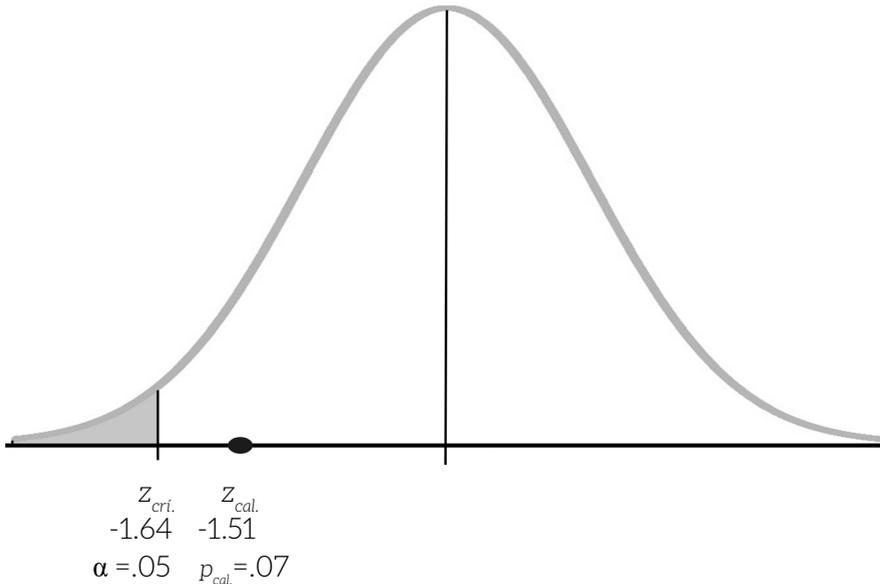
$$\sigma_{\bar{x}} = 15 / \sqrt{32} = 2.65.$$

Luego, se aplica a la fórmula:

$$z_{\text{calculado}} = 96 - 100 / 2.65 = -1.51.$$

Del $z_{\text{calculado}} = -1.51$, se obtiene una $p_{\text{cal}} = .07$. al usar la calculadora en línea (<http://www.socscistatistics.com/pvalues/normaldistribution.aspx>). La Figura 1.10 contiene una distribución de muestras con los valores antes calculados.

Figura 1.10 Test de Significancia Estadística con la curva de la distribución de muestras para una cola



Nota: Valor $z_{\text{crítico}} = z_{\text{cri.}}$ y probabilidad $p_{\text{calculada}} = p_{\text{cal.}}$. La distribución de muestras tiene un promedio de un valor $z = 0$ y el error estándar = 1.

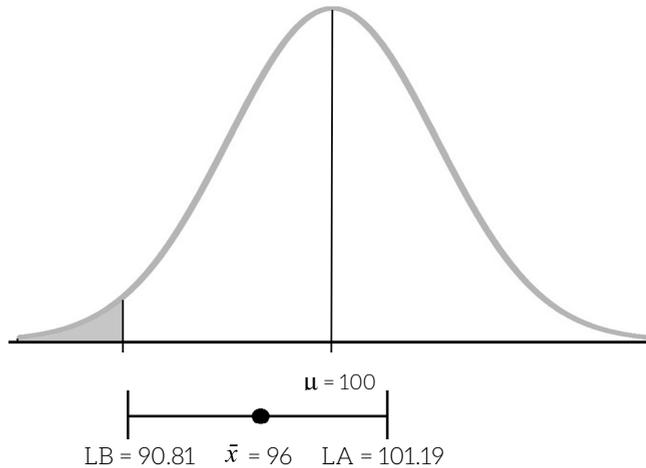
5. Llevar a cabo un intervalo de confianza:

Límite bajo (*lower limit*): $96 - 1.96 \times 2.65 = 90.81$

Límite alto (*upper limit*): $96 + 1.96 \times 2.65 = 101.19$.

El $IC_{95\%} = [90.81, 101.19]$. Este intervalo de confianza aparece en la Figura 1.11.

Figura 1.11 Intervalos de confianza del 95%



6. Interpretar los resultados: *i. e.*, decidir con base en la evidencia si se rechaza o no la hipótesis nula. Los resultados del test indican que el $z_{\text{calculado}} = -1.51 > z_{\text{crítico}} = -1.64$ (*i. e.*, el $z_{\text{calculado}}$ no cae en la zona de rechazo); la $p_{\text{cal.}} = .07 > \alpha = .05$. Por lo tanto, se falla en rechazar la hipótesis nula y los resultados del test no apoyan la hipótesis alternativa: el promedio de la muestra es mayor que el de la población. Se puede adicionar que la probabilidad de encontrar una muestra con un promedio de 96 puntos o menos al azar, sería de .07 cuando la hipótesis nula es cierta: *i. e.*, $H_0: \mu = \bar{x}$.

El intervalo de confianza del 95% [90.81, 101.19] indica que, dadas las muestras de 32 participantes, unas 95 de 100 muestras captu-

rarían el valor verdadero de la población. Este $IC_{95\%}$ captura el promedio de 100 entre sus límites (véase Figura 1.11).

En resumen, el utilizar una o dos colas tiene una relación con los valores críticos z y el alfa; es decir, cuando se emplean dos colas, el alfa se divide entre dos, pero cuando se usa una sola cola (la del lado derecho o la del lado izquierdo) no se divide. Al dividir el alfa en dos, el valor crítico aumenta (e. g., con un alfa de .05 para una cola, el valor crítico es de 1.64; y para las dos colas es de 1.96). Al aumentar el valor crítico en términos absolutos, el valor $z_{\text{calculado}}$ tiene que ser más grande en términos absolutos para rechazar la hipótesis nula. Por ejemplo, en un Test z se obtiene un $z_{\text{calculado}}$ de 1.70 con una $p_{\text{cal.}}$ de .044. En el Test z para una sola cola, este resultado amerita rechazar la hipótesis nula, porque: $z_{\text{calculado}} = 1.70 > z_{\text{crítico}} = 1.64$; y la probabilidad $\text{calculada} = .044 < \text{alfa} = .05$. Por otro lado, y con los mismos datos, no se rechaza la hipótesis nula cuando se usa un test para dos colas, porque: $z_{\text{calculado}} = 1.70 < z_{\text{crítico}} = 1.96$; y la $p_{\text{cal.}} = .088 > \text{alfa} = .05 / 2 = .025$. La probabilidad calculada es la suma de las dos colas: $.044 + .044 = .88$ y, por ello, es más grande que la de una cola. Para demostrar lo anterior, se recurrió a una calculadora en línea que apoyó la conclusión anterior (Figura 1.12).

Figura 1.12 Calculadora para el $z = 1.70$

P Value from Z Score Calculator

This is very easy: just stick your Z score in the box marked Z score, select your significance level and whether you're testing a one or two-tailed hypothesis (if you're not sure, go with the defaults), then press the button!

If you need to derive a Z score from raw data, [you can find a Z test calculator here.](#)

Z score:

Significance Level:

- 0.01
- 0.05
- 0.10

One-tailed or two-tailed hypothesis?:

- One-tailed
- Two-tailed

The P-Value is .044565.

The result is significant at $p < .05$.

P Value from Z Score Calculator

This is very easy: just stick your Z score in the box marked Z score, select your significance level and whether you're testing a one or two-tailed hypothesis (if you're not sure, go with the defaults), then press the button!

If you need to derive a Z score from raw data, [you can find a Z test calculator here.](#)

Z score:

Significance Level:

- 0.01
- 0.05
- 0.10

One-tailed or two-tailed hypothesis?:

- One-tailed
- Two-tailed

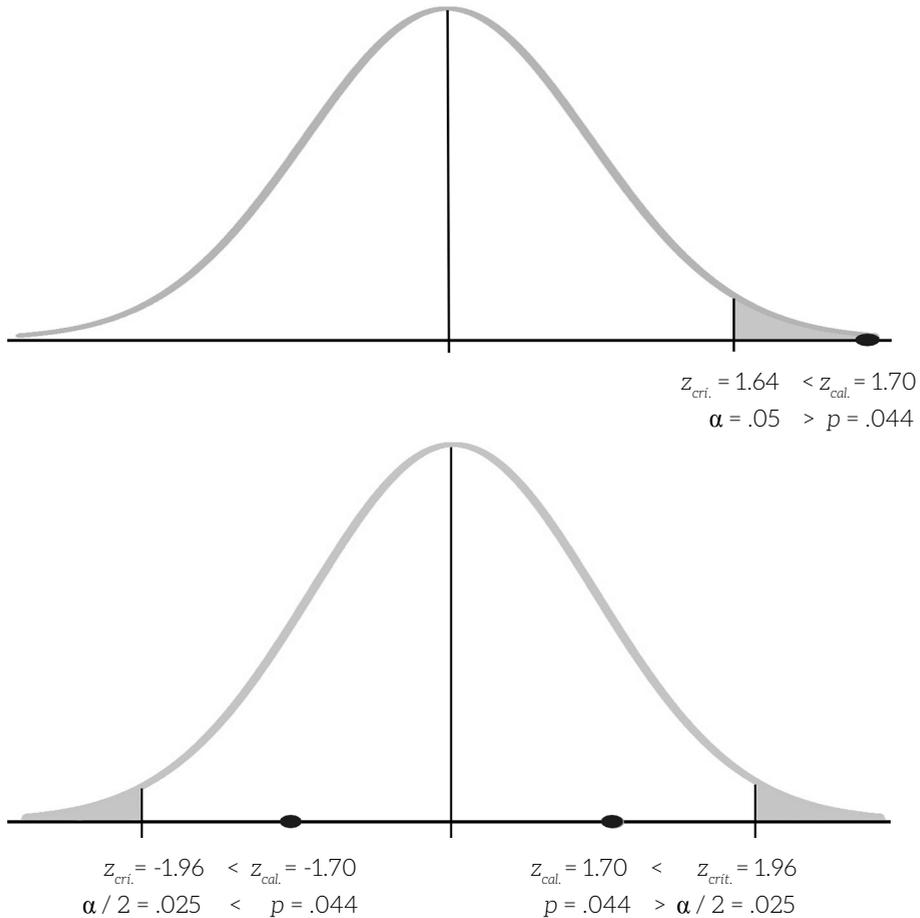
The P-Value is .089131.

The result is *not* significant at $p < .05$.

Fuente: Social Science Statistics.

Utilizando distribuciones normales, la Figura 1.13 muestra la comparación de una vs. dos colas cuando se emplean los datos del ejemplo anterior.

Figura 1.13 Una vs. dos colas



La Tabla 1.2 muestra los valores críticos del Test z cuando se utiliza la curva normal como la distribución de muestras con una y dos colas. Un ejemplo de esta tabla es el Test z con dos colas y un alfa = .10, el cual se divide entre dos: $.10 / 2 = .05$. Este alfa entre dos (.05) es

equivalente al alfa de una cola de .05 y tiene el mismo $z_{\text{crítico}}$ en el valor absoluto de 1.64.

Tabla 1.2 Valores críticos del Test z para una y dos colas usando la curva normal como distribución de muestras

Nivel de Significancia para el Test de dos colas $\alpha / 2c$	Nivel de Significancia para el Test de una cola A	Valores críticos en forma absoluta
.10 / 2	.05	1.64
.05 / 2	.025	1.96
.01 / 2	.005	2.58

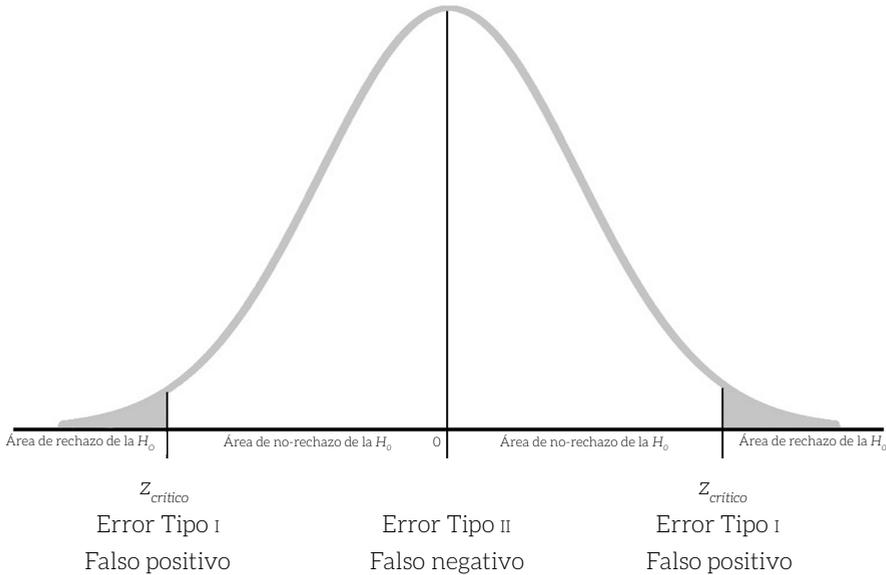
D. ERRORES TIPO I Y II PARA EL TEST Z

El siguiente pasaje es un recordatorio de Ponce (2019). Al poner a prueba la hipótesis nula con un test estadístico como el z, se pueden cometer errores del Tipo I y II. Recordando: el Error Tipo I es la *probabilidad* de rechazar una hipótesis nula cierta. También, se le conoce como el *falso positivo*. En este caso, la hipótesis nula establece que el promedio de una población es igual al de otra población. Puede pasar que el $z_{\text{calculado}}$ del test estadístico aparezca al azar en el área sombreada de la distribución de muestras y, por ello, se rechace la hipótesis nula (Figura 1.14). Sin embargo, al tomar una serie de muestras al azar, estas podrían aparecer en el área de no rechazar. Lo anterior indicaría que el promedio de la población y el de las muestras serían iguales y esto apoyaría la hipótesis nula. Volviendo al valor $z_{\text{calculado}}$ en el área sombreada, si se rechazó la hipótesis nula en este caso, se cometió el Error Tipo I, porque la hipótesis nula era cierta (los promedios de las poblaciones eran iguales): la diferencia entre ambos promedios, se debió a un error de muestreo.

También, cabe la posibilidad (*i. e.*, la probabilidad) de no rechazar una hipótesis nula cuando es falsa (Error Tipo II). A esto se le conoce como *falso negativo*. Esto puede suceder cuando el promedio de una muestra y el de una población enseñan un valor $z_{\text{calculado}}$ que es menor al valor $z_{\text{crítico}}$ (Figura 1.14: área sin sombrear). Sin embargo, al usar más muestras, sus valores $z_{\text{calculados}}$ caerían en las áreas sombreadas y, por lo

tanto, el promedio de las muestras y el de la población no serían iguales. Entonces, se hubiera cometido el Error Tipo II cuando no se rechazó una hipótesis nula falsa (i. e., promedios diferentes), porque se debió a un error de muestreo, ya que el promedio de una muestra y el de la población resultaron iguales.

Figura 1.14 Curva normal como distribución de muestras



En resumen, no se debe rechazar una hipótesis nula cierta: pasa cuando los promedios son iguales a nivel de las poblaciones. Hay que evitar el Error Tipo I cuando a nivel de muestra y población ocurre una diferencia estadísticamente significativa, pero, a largo plazo, otras muestras no tendrían una diferencia estadísticamente significativa. Ahora bien, se debe rechazar una hipótesis nula falsa: sucede cuando los promedios en cuestión no son iguales. No hay que caer en el Error Tipo II cuando a nivel de muestra y población no acontece una diferencia estadísticamente significativa en una sola ocasión. Se debe hacer Test de Significancia Estadística, a largo plazo, para ver si surge un patrón, a través de múltiples estudios (véase a Cumming, 2013, bajo el tema de metaanálisis).

Importancia de cometer uno o el otro error

La importancia de cometer el Error Tipo I o el II es relativa a la investigación que se esté llevando a cabo. Por ejemplo, el Error Tipo I (*falso positivo*) podría ocurrir cuando se compara el promedio de una muestra con el de una población y se concluye que hay una diferencia estadísticamente significativa cuando en realidad no existe esta diferencia. Un ejemplo similar a lo antes descrito en este capítulo, es que se tiene un grupo de estudiantes que se sospecha tienen aptitudes sobresalientes y la evidencia de los análisis de comparaciones de grupos indica que hay una diferencia estadísticamente significativa respecto al promedio. Por lo tanto, se concluye que este grupo tiene aptitudes sobresalientes cuando en realidad estas no existen, porque fue un error de muestreo. Siguiendo con este ejemplo heurístico, el haber sido identificado como una persona con aptitudes sobresalientes podría implicar el tener que recibir asignaturas de las clases más complejas, pero sin afectar negativamente en lo físico o psicológico, entre otros múltiples aspectos, a la alumna o alumno; *ceteris paribus*. En este caso del Error Tipo I no tiene mucha importancia cometerlo, porque no tiene implicaciones negativas; *ceteris paribus*. En contraparte, otro ejemplo heurístico del Error Tipo I es que se identificó a un grupo de estudiantes como personas con retraso mental cuando en realidad esto no era cierto. El supuesto del ejemplo es que esta etiqueta podría implicar estigmatización y falta de atención a ellas y ellos por parte de la comunidad escolar, porque no se espera mucho de estas y estos alumnos. En otras palabras, el haber cometido el Error Tipo I significa una serie de consecuencias negativas. En resumen, estos dos ejemplos del Error Tipo I muestran que la importancia de este error radica en las consecuencias que trae consigo. Expresado de otro modo: las consecuencias del Error Tipo I serían relativas al contexto en el que estén las y los estudiantes.

De una manera similar, la importancia del Error Tipo II depende de las implicaciones que trae consigo. Más precisamente, si no se encontró diferencia estadísticamente significativa entre el promedio de un grupo de alumnas y alumnos con el de la población cuando sí la había, se cometió este último tipo de error. Si las implicaciones serían

que no se van a llevar a cabo asignaturas más complejas, esto podría traer consigo que la o el estudiante no se vea retada o retado académicamente; *ceteris paribus*. Tal vez, esto no tenga otra consecuencia negativa para el desarrollo de la alumna o alumno. De este modo, el haber cometido el Error Tipo II no tendría mucho peso en este ejemplo heurístico. Por otro lado, y siguiendo los ejemplos anteriores, un grupo de estudiantes con retraso mental no es identificado como tal. Por lo tanto, este grupo no recibe la educación especial que necesita y, como consecuencia, no logra académicamente lo que podría si la atención necesaria no es provista. Una posible solución para el Error Tipo II es tener el poder estadístico del 80% (Cohen, 1988), para poder rechazar la hipótesis nula cuando es falsa. La recomendación es usar el *software* de *G*Power*, para estimar el tamaño de la muestra en términos de este nivel de poder estadístico (véase: Ponce, 2019, para detalles acerca de este *software*).

Repaso del cálculo del error estándar

Hinkle *et al.* (2003, p. 189) explicaron que cuando la varianza es conocida, la distribución de muestras del promedio de una población se distribuye normalmente; *i. e.*, se asume que μ es también el promedio de la distribución de muestras y la desviación estándar es el error estándar del promedio ($\sigma_{\bar{x}} = \sigma / \sqrt{n}$). Por otro lado, la desviación estándar de la población es muchas veces desconocida, así que el error estándar se puede estimar de la desviación estándar de una muestra. En corto, la desviación estándar de la muestra es usada para calcular el error estándar del promedio de la distribución de muestras. Entonces, el error estándar de la distribución de muestras es:

$$SE_{\bar{x}} = SD / \sqrt{n}$$

Donde:

- ⊗ $SE_{\bar{x}}$ = Error estándar del promedio de la distribución de muestras
- ⊗ SD = Desviación estándar de la muestra
- ⊗ n = Tamaño de la muestra

PROBLEMAS PARA RESOLVER DEL CAPÍTULO 1

Para resolver los siguientes problemas, se recomienda seguir los pasos de la Tabla 1.1 para desarrollar un procedimiento verificable y, con ello, crear argumentos lógicos y basados en la evidencia encontrada. Estos problemas se resuelven manualmente y para el p , con la ayuda de una calculadora* en línea:

- a) Crear la pregunta de investigación en los casos donde sea necesario (en algunos de los problemas, se dan las interrogantes para que la lectora o el lector siga esta lógica y las desarrolle para el resto de los problemas).
- b) Derivar las hipótesis nula y alternativa de la pregunta de investigación para una o dos colas.
- c) Pensar y argumentar la razón para usar una o dos colas.
- d) Establecer el criterio para rechazar la hipótesis nula con un nivel del alfa (α) = .05.
- e) Pensar y argumentar otros posibles criterios a utilizar.
- f) Calcular el valor de la prueba estadística (*Test Statistic*); i. e., en este caso, el $z_{\text{calculado}}$ y su correspondiente probabilidad $_{\text{calculada}}$, empleando para este último la calculadora* (<http://www.socscistatistics.com/pvalues/normaldistribution.aspx>)
- g) Llevar a cabo un intervalo de confianza del 95% de confianza.
- h) Graficar todos los resultados.
- i) Interpretar los resultados.
- j) Se recomienda ir más allá de los puntos anteriores y desarrollar algunos argumentos lógicos basados en la evidencia recabada y otras lecturas o conocimientos que se tengan en cuanto a los temas abordados en cada uno de los problemas.

Nota: En los problemas del Capítulo 1, no se proveen sets de datos y se asume que las estadísticas provienen de puntajes con una distribución semejante a la normal (véase: Ponce, 2019, para más detalles). Cuando se trabaje con sets de datos, se recomienda hacer Test de Normalidad para conseguir evidencias que apoyen el uso de distribuciones normales (véase

esta página para los Test de Normalidad de Kolmogorov-Smirnov: <https://www.spss-tutorials.com/spss-kolmogorov-smirnov-test-for-normality/>).

1. Una investigadora evaluó a estudiantes de sexto grado de primaria en lectoescritura. Dada la similitud entre los datos demográficos de la muestra y la población, así como que tomó a la escuela primaria al azar de una población de 10 escuelas, ella asume que la muestra es representativa de la población de interés. Ahora se plantea la siguiente interrogante: ¿existe una diferencia estadísticamente significativa entre el promedio de la población y el de una muestra? La investigadora cuenta con los siguientes parámetros y estadísticas:

- ⊗ Promedio de la población (μ) = 8
- ⊗ Desviación estándar de la población (σ) = 1.50
- ⊗ Promedio de la muestra (\bar{x}) = 7.70
- ⊗ Tamaño de la muestra (n) = 45

Otros cuestionamientos más allá del contenido de esta obra serían: ¿qué implicaciones habría si se está por abajo, igual o por encima del promedio de la población?; ¿cuáles son los factores o variables independientes que se asocian con la lectoescritura?; y ¿qué teorías explican el fenómeno de la lectoescritura?

2. Un investigador tiene la tarea de comparar el promedio de matemáticas de una escuela secundaria privada con el resto de las escuelas privadas en una ciudad. Se supondría que el plantel en cuestión es parte de esa población de interés e, históricamente, ha sufrido rezago en esta materia. Entonces, se plantea: ¿existe una diferencia estadísticamente significativa entre el promedio de matemáticas de una escuela secundaria privada con el de la población? Esta pregunta también puede ser estructurada de la siguiente forma coloquial: estadísticamente hablando, ¿es el promedio de la población superior al de la muestra?

- ⊗ Promedio de la población (μ) = 8.2
- ⊗ Desviación estándar de la población (σ) = 1.10
- ⊗ Promedio de la muestra (\bar{x}) = 8.0
- ⊗ Tamaño de la muestra (n) = 550

Otras interrogantes que podrían surgir como un reto de dar solución a la lectora o el lector serían: ¿qué dice la literatura acerca del rezago en matemáticas?; ¿existe una relación estadísticamente significativa entre el aprendizaje de las matemáticas y otros factores?; ¿por qué se considera a las matemáticas una materia difícil de aprender?; ¿qué área de las matemáticas es más difícil de entender?

3. Se tiene la idea de que los varones tienen mayor capacidad para las matemáticas y las ciencias. Para poner a prueba esta idea, un grupo de estudiantes universitarios deciden definir a su escuela como la población de interés para su proyecto de investigación. Acuden a las oficinas administrativas de su universidad para conseguir los puntajes del examen de admisión a esa institución. Los administradores les proporcionan los puntajes de matemáticas y ciencias de las damas (150 puntos) y la desviación estándar (20 puntos). Sin embargo, por alguna razón desconocida no les otorgan los puntajes de los varones. Entonces, el grupo toma una muestra representativa, y al azar, de los varones ($n = 600$), a quienes se les pregunta por este puntaje y el promedio es de 153 puntos.

Más allá de la comparación estadística de los pasos antes descritos, se le cuestiona a la lectora o el lector: ¿existen diferencias naturales entre mujeres y hombres que ayuden a unas o a otros a entender mejor las matemáticas y las ciencias?; ¿qué dicen las investigaciones previas al respecto?; y ¿cuál es el argumento personal que se puede hacer acerca de este tema?

4. En una ciudad, las y los docentes gastan un promedio de \$500.00 con una desviación estándar de \$50.00 en materiales para sus clases. Una escuela dentro de esta ciudad gasta \$550.00 y cuenta con 65 docentes. El director desea saber si sus subordinados se están excediendo en estos gastos y decide llevar a cabo una comparación.

Además de esta comparación, el director se podría preguntar: ¿por qué existe una diferencia si es que la hay?; ¿qué implicaciones tiene lo que se encontró con la evidencia?; y ¿qué tan bueno podría ser, a la larga, que las y los docentes estén gastando más de la cuenta en materiales?

5. Se tiene la idea que el dar tutorías a las y los estudiantes ayuda en su proceso de aprendizaje. Se ha encontrado que 5 horas (300 minutos) en promedio semanal de tutorías son benéficas con una desviación estándar de una hora (60 minutos). Una escuela tiene como promedio 4 horas (240 minutos) a la semana de tutorías ($n = 150$). La directora de la institución desea saber si no está muy por debajo del promedio recomendado.

Otras posibilidades más allá del contenido de este capítulo son: muchas veces se miden las tutorías en tiempo, pero ¿qué hay acerca de la calidad de estas?; ¿qué entrenamiento necesita una docente o un docente para dar tutorías de buena calidad?; y ¿qué se sabe al respecto de las tutorías en forma teórica y empírica?

PREGUNTAS PARA REFLEXIONAR

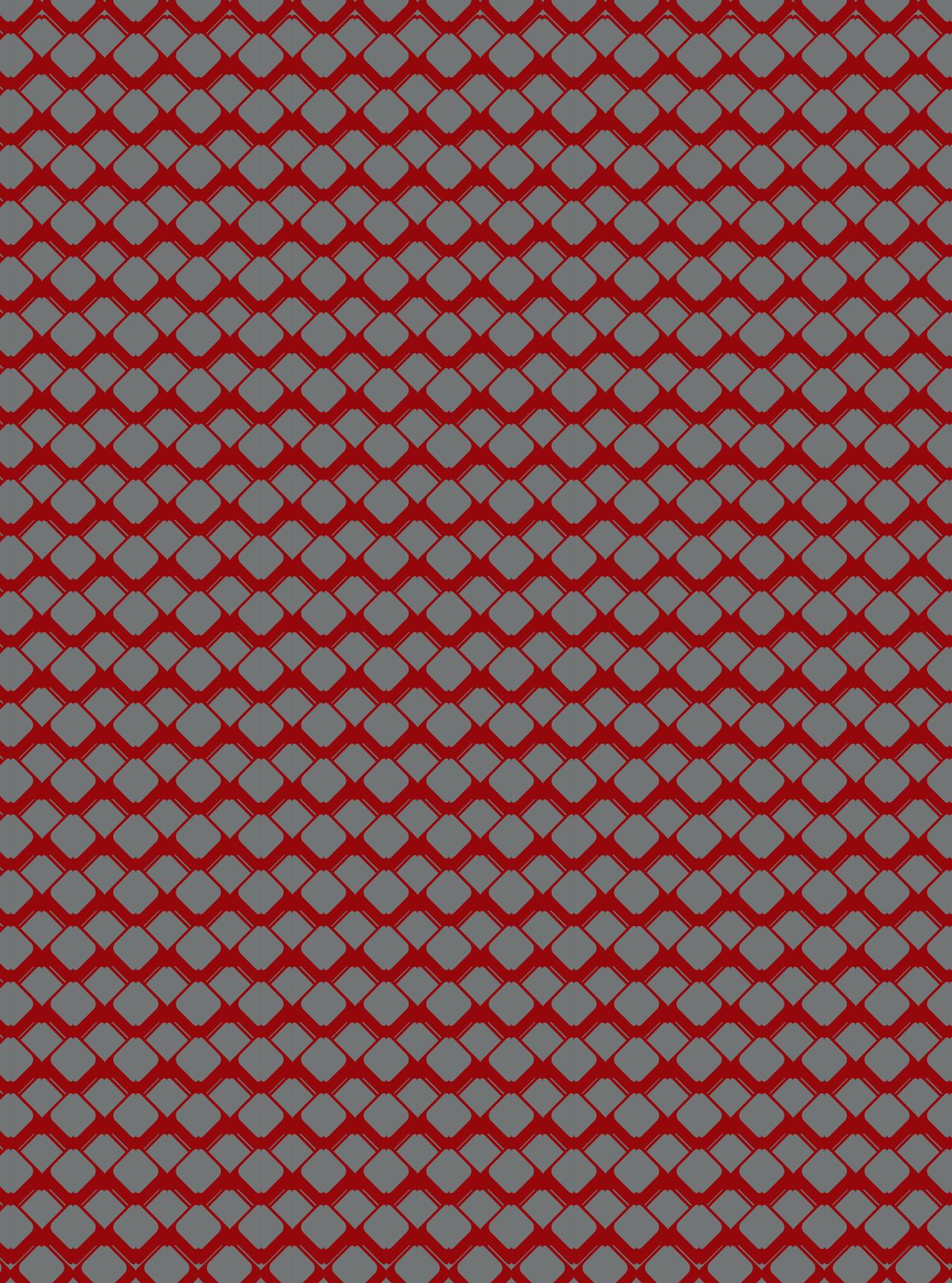
1. ¿Qué tan parecidos son los argumentos encontrados en artículos de investigación educativa cuantitativa respecto a un argumento con lógica matemática y lógica aplicada a la estadística?
2. ¿En qué situaciones no es apropiado hacer una comparación entre una muestra y una población?
3. ¿Qué habría que hacer si los resultados de un Test de Significancia y del intervalo de confianza se contradicen?
4. ¿Cuál error es más grave de cometer: Tipo I o II?
5. ¿Qué relación tiene el nivel de confianza de un intervalo de confianza y la longitud de sus límites?
6. ¿Qué valores hay que utilizar para estimar el error estándar cuando no se tienen los parámetros de una población?

OPINIÓN PERSONAL

Argumentar es fundamental en la investigación educativa y, para ello, la lógica matemática es vital. Por esta razón, se recomienda estudiar libros de texto que traten de lógica matemática, como el de Hurley (2014). Por otro lado, la advertencia aquí es que no se suele calcular un tamaño del efecto entre el promedio de una muestra y el de la población, según algunos investigadores; es decir, no se suele medir la magnitud de la diferencia entre esta estadística y el parámetro. Esto es porque el d de Cohen se utiliza cuando se tienen tamaños de grupos similares también como valores de desviación estándar similares. En el caso de los datos empíricos de una muestra, se puede calcular el tamaño exacto de este grupo en cuestión; sin embargo, el tamaño de la población es una historia diferente, porque suele ser muy grande. Esta diferencia en tamaños de los grupos puede hacer el cálculo del d de Cohen cuestionable. La recomendación sería buscar en la literatura el tema que se está abordando, para ver si se han comparado muestras con poblaciones. Si hay precedentes, estos se podrían usar como justificación para realizar esta comparación: siempre y cuando suenen razonables. Otra alternativa sería crear uno sus propios argumentos para llevar a cabo esta comparación. El Apéndice E muestra una serie de fórmulas del d de Cohen y de otros efectos similares para diferentes situaciones. Más al respecto, se recomienda el Manual de la APA (2020) para ver cómo se reportan otros efectos.

Asimismo, está el problema de interpretar este efecto antes mencionado. No encuentro cómo interpretarlo, porque dependería del contexto. Lo anterior, se retoma en el siguiente capítulo para analizar el tamaño del efecto. Por otro lado, las comparaciones entre el promedio de una muestra y el de una población pueden ser un primer paso para ver si se cumple con algún criterio escolar, que puede ser cierto tabulador que indique la obtención de becas o la atención a estudiantes con aptitudes sobresalientes o retraso académico, entre muchas otras posibilidades. De igual manera, los intervalos de confianza ayudan a estimar el valor verdadero, porque una estadística de punto no incluye el error como lo hacen estos primeros.

En contraparte, existen procedimientos de significancia estadística para comparar tamaños del efecto encontrados en las poblaciones contra los encontrados en las muestras. Estos se llaman *Test de Equivalencia* y de *No-Inferioridad*. Con los Test de Equivalencia, se puede poner a prueba si el tamaño de una muestra es equivalente al encontrado en una población o en la literatura, al establecer cierto rango en el cual, si el efecto observado cae, se le considera igual al que se está comparando (véase: Lakens, Scheel, & Isager, 2018; Wellek, 2010). También, están los Test de No-Inferioridad donde se pone a prueba si un efecto observado *no* es menor a cierto previamente encontrado. La ventaja de estos test es que están en el mismo paradigma de la Teoría de la Frecuencia; sin embargo, es necesario leer las fuentes antes mencionadas para tener una idea de cómo poner a prueba el tamaño de un efecto observado.



CAPÍTULO 2

TEST DE COMPARACIÓN DE GRUPOS CON LA DISTRIBUCIÓN T DE ESTUDIANTE

En algunas ocasiones, algunas y algunos educadores dicen que cierto método de enseñanza es mejor que el resto. Más allá de algún discurso para tratar de venderlo, se debió de hacer alguna comparación para llegar a esta conclusión, la cual, a su vez, debió de haber tenido un par de análisis para tener un sustento teórico y práctico: efecto estadístico y efecto práctico. Por lo menos, se esperaría que lo anterior pasara en las más de las ocasiones en la investigación educativa. Uno de los sustentos teóricos está en la *Teoría Clásica de la Probabilidad* aplicada a las estadísticas: efecto estadístico. Esta teoría tiene que ver con el azar y un ejemplo es con lanzamientos de una moneda o de un dado al aire, para ver en cuántas ocasiones sucede algún evento (cae cara o cruz; o algún número del 1 al 6). Al aplicarse a las estadísticas, estos eventos forman una distribución binomial que es análoga (similar) a la distribución normal que se utiliza como modelo para comparar grupos (véase: Ponce, 2019). Con este sustento teórico de la probabilidad y la estadística, se estimaría si el promedio

de una muestra fue diferente al de otra. Se examina si posiblemente la diferencia fue al azar o no: efecto estadístico. El otro sustento sería el efecto práctico, del cual muchos teóricos han promovido su uso en la investigación. De hecho, la *Asociación Americana de Psicología (APA)* (2020) ha insistido en que una investigación no está completa sin incluir algún efecto en la narrativa de un estudio. Este efecto práctico o tamaño del efecto indica la magnitud de la diferencia cuando se comparan promedios de muestras. Dependiendo del tamaño de la magnitud, se podría concluir que algún tratamiento tuvo éxito. Contrastando el efecto estadístico al práctico, puede pasar que se infiera que existe un efecto estadístico cuando no hay uno práctico y viceversa. Por ejemplo, un tratamiento con cierto medicamento podría tener un efecto estadístico en disminuir un dolor de cabeza; pero, por otro lado, esta disminución del dolor de cabeza no implica que la persona pueda llevar a cabo sus labores cotidianas. Entonces, el efecto práctico no sería grande. En resumen, lo anterior es una posibilidad en la investigación educativa cuando se encuentra un efecto estadístico, pero no un *efecto práctico*; y otro término es significancia práctica: “La medida en que el resultado de un estudio tiene aplicaciones significativas en entornos del mundo real” (VandenBos, 2015, p. 817). La recomendación es indagar en la literatura pertinente al tema de investigación acerca de la magnitud de los tamaños del efecto. Un *efecto estadístico* abre la posibilidad de encontrar un efecto práctico.

A. GENERALIDADES DE LA DISTRIBUCIÓN T

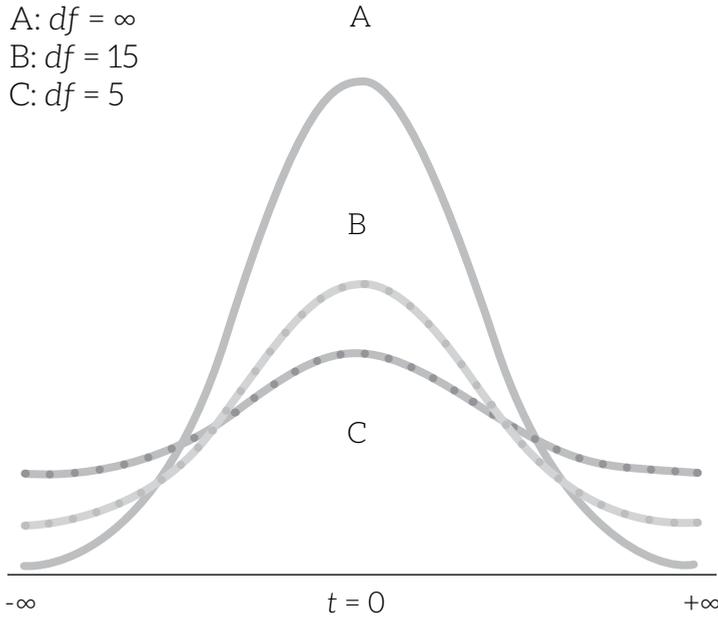
Las distribuciones *t* de estudiante son una *familia* de distribuciones con forma similar a una distribución normal (*i. e.*, simétricas, centradas en el promedio, en forma de campana, asintóticas) que dependen del tamaño de la muestra para su forma. Existen otras distribuciones *t* sin simetría que tienen aplicaciones diferentes a las expuestas en este libro (véase: Kotz, & Nadarajah, 2004). VandenBos (2015, p. 1069) definió la distribución *t* (*i. e.*, distribución *t* de estudiante: *Student's t distribution*):

Es una distribución teórica de probabilidad que juega un rol central en poner a prueba la hipótesis acerca de los promedios de las poblaciones, entre otros parámetros. Es la distribución de muestras de una estadística como el promedio: $(\bar{x} - \mu_0) / SD$. Donde μ_0 es el promedio de la población de la cual se sacó una muestra, \bar{x} es la estimación del promedio de la población obtenida por medio de una muestra de los datos, y SD es la desviación estándar del set de datos.

La forma de la distribución t cambia de acuerdo con el tamaño de la muestra n , más precisamente, los *grados de libertad* (df ; véase Figura 2.1). Según VandenBos (2015, p. 291), los df representan “El número de elementos que pueden variar en el cálculo estadístico o el número de escores menos el número de restricciones”. Por ejemplo, un estudiante tiene un promedio de 89 en una materia al final del semestre, el cual proviene de cuatro evaluaciones. Los df se calculan con el número de escores (*i. e.*, evaluaciones) menos las restricciones (*i. e.*, promedio): $4 - 1 = 3 df$. En otras palabras, si se conoce el valor de alguna de las evaluaciones (*e. g.*, la evaluación conocida = a), las otras tres pueden variar: *e. g.*, promedio es $89 = (a + x + y + z) / 4$. Para una explicación más profunda acerca de los df , se recomienda ver a Walker (1940).

La Figura 2.1 muestra tres diferentes distribuciones t con distintos df : 5, 15 e ∞ . La distribución t con menos df tiene colas que cubren más área. Conforme los df aumentan, las colas se vuelven más delgadas. Llega un punto en que la distribución t , se vuelve de la misma forma que la distribución normal: *e. g.*, $df = \infty$. De hecho, la distribución t se vuelve la distribución normal al nivel de 120 df , aproximadamente.

Figura 2.1 Distribuciones t con diferentes grados de libertad (df)



Al emplear la desviación estándar de una muestra (SD) en lugar de la desviación estándar de una población (σ), para estimar el error estándar (σ_x), se tiene un efecto en el tamaño de este último, especialmente para muestras pequeñas (Hinkle *et al.*, 2003, p. 190); es decir, entre más pequeño sea el tamaño de la muestra, mayor será el error estándar; *ceteris paribus*. Por esta razón del error estándar es más adecuado el uso de la distribución t que la distribución normal z para muestras pequeñas.

La Figura 2.2 muestra las áreas de rechazo y no-rechazo de la hipótesis nula para la distribución t de muestras. A diferencia de la distribución normal con los valores z fijos, la distribución t cambia su forma de acuerdo con los grados de libertad. Cuando se tienen menos df , hay más área en las colas y, por lo tanto, más probabilidad de que un valor $t_{\text{calculado}}$ aparezca en los extremos que cuando se tienen más df . En contraste, cuando se tienen más df , hay más área en el centro de la distribución t y, por lo tanto, es más probable que un valor $t_{\text{calculado}}$ aparezca en el centro de la distribución. Esto tiene la implicación de que

si se elige el tradicional Coeficiente Alfa de .05 y se cambian los df (e. g., al modificar el tamaño de la muestra), los valores críticos cambian también. De hecho, si se mantiene fijo un alfa y se incrementan los df , el valor crítico t disminuirá. Esta relación entre las variables, se puede expresar de la siguiente manera:

- ⊗ Δ = Valor crítico t ($\Delta df, \bar{\alpha}$)
- ⊗ Δ = La letra griega delta significa cambio
- ⊗ $\bar{\alpha}$ = Coeficiente Alfa fijo

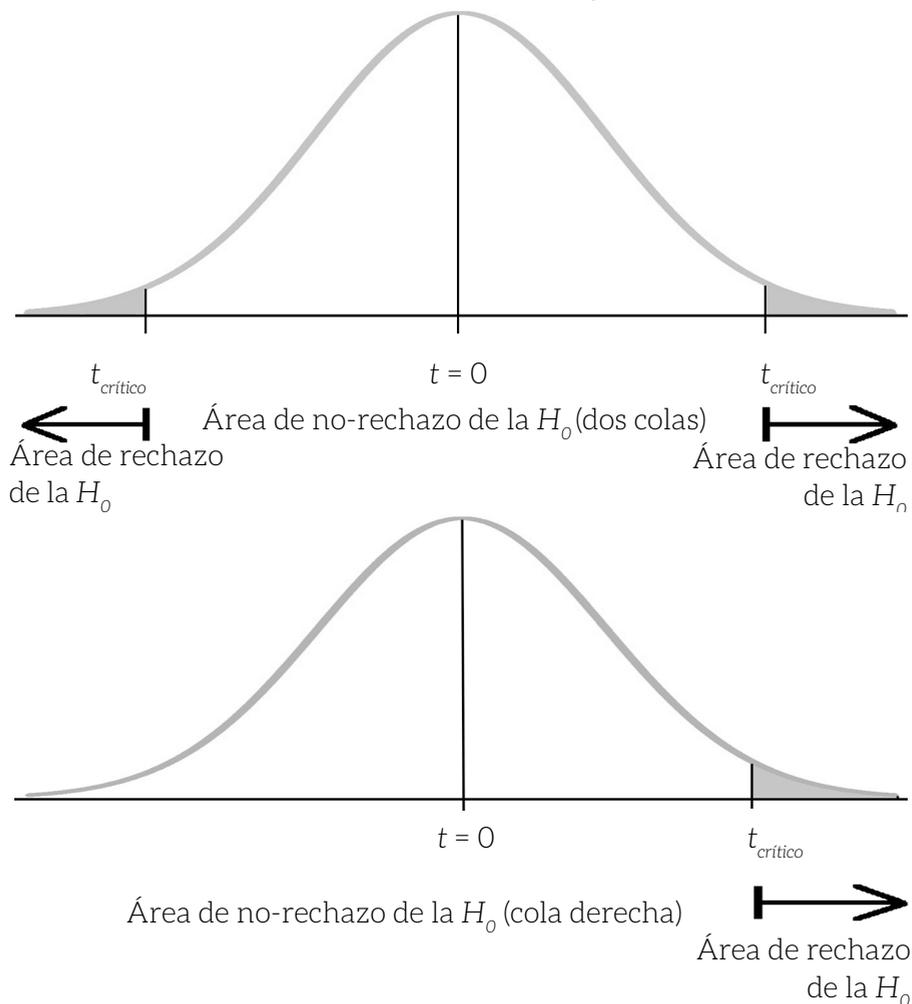
Esta fórmula indica que el valor $t_{\text{crítico}}$ está en función o depende del alfa que en este ejemplo está fijo (indicado por el guion encima del α) y el cambio en los df . Por ejemplo, el valor $t_{\text{crítico}}$ cambia (disminuye) cuando los df cambian (aumentan), manteniendo el alfa fijo. Asimismo, en la Figura 2.2 se indican los tres escenarios a los que se puede enfrentar una investigadora o un investigador educativo cuando pone a prueba una hipótesis nula ($H_0: \mu = \bar{x}$), y la hipótesis alternativa se ajusta a estos escenarios (H_A):

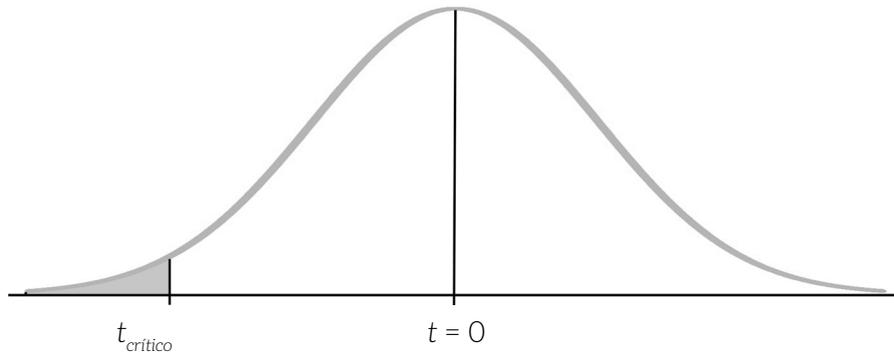
- ⊗ Emplear dos colas cuando no se espera que el promedio de una muestra (\bar{x}) sea más grande o más pequeño que el promedio de la población (μ): entonces, la hipótesis alternativa sería: $H_A: \mu \neq \bar{x}$. Lo anterior podría ser más apropiado cuando un estudio es exploratorio y no existe una expectativa del sentido de la diferencia entre los promedios de los grupos que se están comparando.
- ⊗ Usar la cola derecha cuando se espera que el promedio de una muestra sea más grande que el promedio de la población: $H_A: \mu < \bar{x}$. Esta expectativa puede haber surgido de la literatura o de un diseño experimental, entre otros.
- ⊗ Utilizar la cola izquierda cuando se espera que el promedio de una muestra sea más pequeño que el promedio de la población: $H_A: \mu > \bar{x}$.

Con estos tres escenarios, se puede decidir que: el test tendrá dirección hacia la derecha o la izquierda del promedio de la población; o

no tendrá dirección, empleando las dos colas. Una implicación de usar una cola es que se tiene mayor poder estadístico y mayores posibilidades de rechazar la hipótesis nula, porque el valor $t_{\text{crítico}}$ es más pequeño que el de dos colas; *ceteris paribus* (véase a Ponce, 2019, para más detalles). Al igual que con los otros Test de Significancia Estadística, se pueden cometer –ya mencionados previamente– los Errores Tipo I y II.

Figura 2.2 Distribución t de muestras con sus áreas de rechazo y no-rechazo de la H_0





← Área de no-rechazo de la H_0 (cola izquierda)
 Área de rechazo
 de la H_0

Para complementar lo anterior, se tiene la Tabla 2.1 que muestra diferentes grados de libertad y cómo estos cambian el tamaño de los valores $t_{críticos}$. En corto, al aumentar los df , los valores $t_{críticos}$ disminuyen; al disminuir estos últimos, la posibilidad de rechazar una hipótesis nula aumenta. Por ejemplo, se tiene una muestra con 5 df con un valor $t_{calculado} = 2.30$ y su correspondiente valor $t_{crítico} = 2.571$ para dos colas y un alfa = .05. Se hace la comparación: $t_{calculado} = 2.30 < t_{crítico} = 2.571$ y, por lo tanto, se falla en rechazar la hipótesis nula. En contraparte, si se tienen 15 df , el $t_{calculado} = 2.30 > t_{crítico} = 2.131$, por lo que sí se rechazaría la hipótesis nula. En resumen, el valor de $t_{calculado}$ está fijo para este ejemplo y al cambiar los df , se va desde no rechazar la hipótesis nula hasta rechazarla. De esto, se puede concluir que este Test t de Significancia Estadística es sensible a los df , los cuales dependen del tamaño de la muestra.

Tabla 2.1 Rechazar o no con el mismo valor $t_{calculado}$ y diferentes df

df	$t_{crítico}$	$t_{calculado}$	Rechazar la hipótesis nula
5	2.571	2.30	No
15	2.131	2.30	Sí
∞	1.960	2.30	Sí

Nota: Los valores $t_{críticos}$ corresponden a un alfa = .05 para dos colas.

Pasando al uso del *software* SPSS (IBM *software*), se muestran en el Apéndice A las generalidades del uso de este. Los otros apéndices (B-D) contienen ejemplos de los análisis con la distribución *t*. El objetivo es enseñar cómo se hacen estos análisis con el *software* SPSS de la versión 23.

B. EJEMPLO DE TEST T CON UNA SOLA MUESTRA

Tabla 2.2 Pasos para el Test *t* y el *d* de Cohen

1.º Establecimiento de la o las preguntas de investigación.
2.º Test de Levene.
3.º Derivar las hipótesis de la pregunta de investigación: una vs. dos colas.
4.º Establecer el criterio para rechazar la hipótesis nula: el nivel del alfa (α); (e. g., $\alpha = .01$; $\alpha = .05$; o $\alpha = .10$).
5.º Calcular el valor de la prueba estadística (Test Statistic); i. e., en este caso, el valor $t_{\text{calculado}}$ y su correspondiente probabilidad $_{\text{calculada}}$.
6.º Llevar a cabo un intervalo de confianza.
7.º Cálculo del <i>d</i> de Cohen.
8.º Interpretar los resultados.

Un ejemplo heurístico para el uso de la distribución *t* en un Test *t* es el siguiente (donde se siguen los pasos de la Tabla 2.2): un grupo de aspirantes a ingresar a una universidad pública realizan un examen de admisión para ver si son aceptados en diferentes pregrados. Se tiene una situación en la cual se quiere comparar un grupo de aspirantes de ingeniería ($\bar{x}_{\text{Ing.}} = 102$; $SD = 1.63$; $n = 16$) con el resto de la población ($\mu = 100$) en cuestión de los puntajes de admisión a una universidad (el Apéndice B contiene la solución matemática de este ejemplo). Entonces, la pregunta de investigación es:

1. *Establecimiento de la o las preguntas de investigación*: ¿existe una diferencia entre los puntajes promedio de los aspirantes de ingeniería y el resto de la población? Esta misma interrogante también puede ser expresada de la siguiente manera: ¿existe una diferencia estadísticamente significativa entre los puntajes promedio de los aspirantes de ingeniería

y el resto de la población? De esta pregunta se derivan las siguientes hipótesis:

2. *Derivar las hipótesis de la pregunta de investigación:* como no se tiene expectativa de si el promedio de la muestra estará por encima o por debajo del promedio de la población, se opta por utilizar un Test No-Direccional de dos colas:

$$H_0: \mu = \bar{x}_{Ing.}$$

$$H_A: \mu \neq \bar{x}_{Ing.}$$

Donde:

- ⊗ μ = Promedio de la población de aspirantes a una universidad
- ⊗ $\bar{x}_{Ing.}$ = Promedio de la muestra de aspirantes a ingeniería

Una vez establecidas las hipótesis, la Tabla 2.3 contiene los datos crudos de los aspirantes de ingeniería a ingresar a una universidad, así como las estadísticas que indican que el promedio = mediana = moda, así que estos resultados, junto con un Coeficiente de Curtosis menor a $|3|$ y simetría = 0, apoyan la idea de que el set de datos está distribuido normalmente (véase a Ponce, 2019, para más pruebas de normalidad). Más al respecto y como una recomendación, Cain, Zhang y Yuan (2017) abordaron los problemas de falta de normalidad. El set de datos es considerado pequeño ($n = 16$). Por esta razón, el Test t es más adecuado que el Test z . También, se ignora la desviación estándar de la población (σ), pero para el Test t no es necesario conocerla para el cálculo del error estándar ($\sigma_{\bar{x}}$), porque se tiene la SD de la muestra.

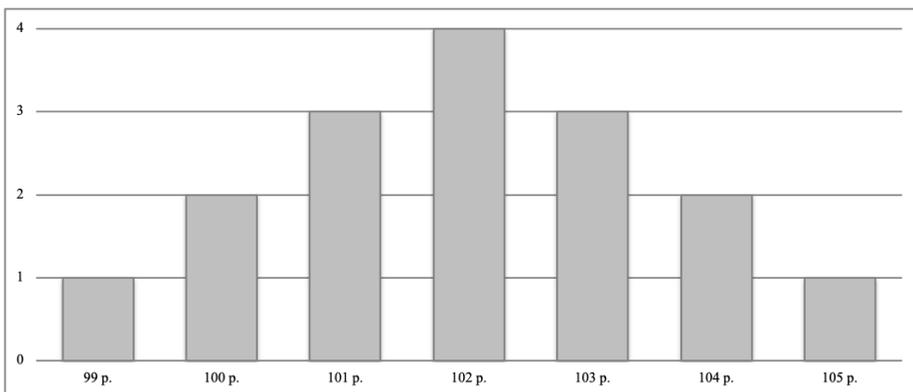
Tabla 2.3 Puntajes y estadísticas de los aspirantes a ingeniería

Aspirantes	Puntajes
1	99
2	100
3	100
4	101

Continúa...

Aspirantes	Puntajes
5	101
6	101
7	102
8	102
9	102
10	102
11	103
12	103
13	103
14	104
15	104
16	105
Promedio	102
Mediana	102
Moda	102
SD	1.63
Curtosis	-0.46
Simetría	0

Figura 2.3 Gráfica de barras de los datos de los aspirantes a ingeniería

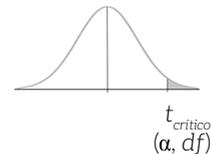


Asimismo, la Figura 2.3 indica que los datos parecen estar distribuidos normalmente.

3. Establecer el criterio para rechazar la hipótesis nula: el nivel del alfa (α): al igual que en el Test z, el Test t también ofrece diferentes criterios de significancia. De la misma forma, el Coeficiente Alfa de .05 es el tradicional. Cuando el Test t es de dos colas, el alfa se divide entre dos: $.05 / 2 = .025$ por cola. Empleando la Tabla 2.3 (un extracto de una Tabla de valores t), el alfa dividido entre dos es mostrado en la columna sombreada. Ahora, los grados de libertad son: $16 - 1 = 15$ df por tener una sola muestra (i. e., $n - 1$), que se encuentra en el renglón sombreado. Esta columna y renglón son las coordenadas para ubicar el valor $t_{\text{crítico}}$: 2.131 (Tabla 2.4).

Tabla 2.4 Algunos valores t y sus correspondientes probabilidades en la distribución t

Tabla A.2 Valores t y sus correspondientes Coeficientes Alfa



Grados de libertad (df)	Coeficientes Alfa						
	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.076	31.821	63.657	318.310	636.620
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073

Nota: Por tradición, la distribución t se utiliza cuando las muestras son menores a 30 observaciones, participantes u objetos.

Para estimar la probabilidad $p_{calculada}$, se usa la calculadora para el Test t en línea: <http://www.socscistatistics.com/tests/tsinglesample/Default2.aspx>

La Figura 2.5 muestra los campos que se tienen que llenar para calcular el valor t y su correspondiente p . Hay que proveer el promedio de la población, el set de datos de la muestra, seleccionar un alfa y escoger una o dos colas para el test.

Figura 2.5 Calculadora del valor t

Single Sample T-Test Calculator

The value of t is 4.898979.

Population mean (μ)

100

Sample X

99
100
100
101
101
101
102
102
102
102
103
103
103
104
104
105

Significance Level:

0.01

0.05

0.10

One-tailed or two-tailed hypothesis?:

One-tailed

Two-tailed

The t -value is 4.898979. The value of p is .000193. The result is significant at $p < .05$.

Calculate

Reset

Fuente: Social Science Statistics.

Esta calculadora indica para el set de datos con dos colas y un alfa = .05: $t_{cal.} = 4.89$ y $p_{cal.} = .000193$.

5. *Llevar a cabo un intervalo de confianza:* en este caso, hay que encontrar un valor $t_{crítico}$ en una Tabla de valores t , dado el alfa de .05 y su correspondiente nivel de confianza del 95%. Las siguientes fórmulas se emplean para calcular el IC para los valores t (Ecuación 1.3 y Ecuación 1.4; con la diferencia de que en estas ecuaciones, se utiliza el $z_{crítico}$ y en las siguientes, el $t_{crítico}$):

Límite bajo (*lower limit*): $\bar{x} - \text{Valor } t_{crítico} \times \text{Error estándar del promedio}$

Límite alto (*upper limit*): $\bar{x} + \text{Valor } t_{crítico} \times \text{Error estándar del promedio}$

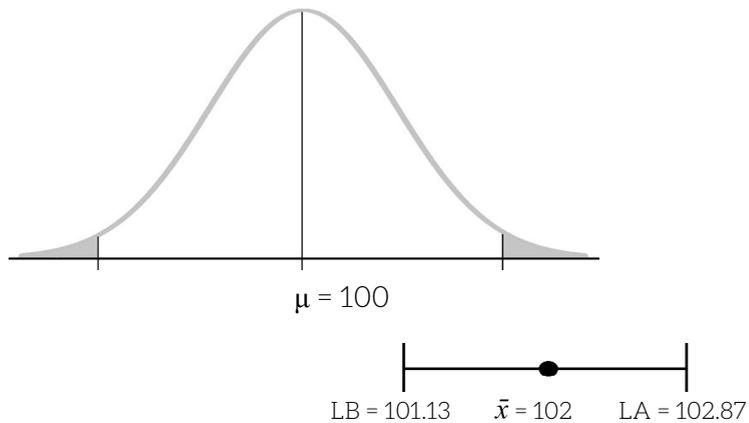
Dado que ya se había previamente encontrado el valor $t_{crítico}$ en la Tabla t , este resulta: 2.131. *Recordando:* a diferencia de los Test z donde los valores críticos son fijos, los valores críticos de los Test t tendrán diferentes valores críticos, dependiendo de los grados de libertad. También, se había encontrado el error estándar del promedio: $SE_{\bar{x}} = 0.41$. Solo se hace la sustitución de la fórmula:

$$\text{Límite bajo (lower limit): } 102 - 2.131 \times 0.41 = 101.13$$

$$\text{Límite alto (upper limit): } 102 + 2.131 \times 0.41 = 102.87$$

El $IC_{95\%} = [101.13, 102.87]$. Este intervalo de confianza aparece en la Figura 2.6.

Figura 2.6 Intervalos de confianza del 95%



6. Cálculo del d de Cohen: un tamaño de efecto como el d mide la magnitud de la diferencia expresada en unidades de desviación estándar entre dos promedios (en este caso, es entre el promedio de la muestra y el promedio de la población). Al respecto, Cohen (1988) dijo que depende del tamaño del efecto la posible existencia de un fenómeno; es decir, si el d es grande esto indica un efecto, el cual es una inferencia (se recomienda ver a Hurley, 2014). La fórmula de este efecto es la Ecuación 2.2:

$$d_1 = \frac{(\bar{x}_1 - \text{Promedio de otro grupo})}{SD_1} \quad (\text{Ecuación 2.2})$$

Donde:

- ⊗ d_1 = Tamaño del efecto: d de Cohen ante una estadística de los datos y un parámetro de una población o estadística encontrada en la literatura u otra fuente
- ⊗ \bar{x}_1 = Promedio calculado de los datos del grupo 1
- ⊗ Promedio de otro grupo = Promedio de una población o muestra que se haya encontrado en la literatura u otra fuente

- ⊗ SD_1 = Desviación estándar calculada de los datos del grupo (véase: Apéndice E para más información acerca de este efecto y cómo calcular la desviación estándar)

Sustituyendo en la fórmula:

$$d_1 = \frac{(102 - 100)}{1.63} = 1.23.$$

7. *Interpretar los resultados:* i. e., decidir con base en la evidencia si se rechaza o no la hipótesis nula. Los resultados del test indican que el $t_{\text{calculado}} = 4.88 > t_{\text{crítico}} = 2.131$. Usando la ya mencionada calculadora de <http://www.socscistatistics.com/pvalues/tdistribution.aspx>, se obtiene la $p_{\text{cal.}} = .000193 < \alpha = .05 / 2 = .025$ (en el Apéndice B [Figura B-4] se muestran algunos de estos resultados). Por lo tanto, se rechaza la hipótesis nula y los resultados del test apoyan la hipótesis alternativa: el promedio de la muestra es mayor que el promedio de la población. Se puede adicionar que la probabilidad de encontrar una muestra con un promedio de 102 puntos en el examen de admisión al azar sería de .000193 cuando la hipótesis nula es cierta: $H_0: \mu = \bar{x}$. Podría parecer que una diferencia entre los dos promedios de dos puntos es pequeña, pero al obtener el error estándar de la desviación estándar y el tamaño de la muestra resulta una diferencia estadísticamente significativa. En otras palabras, la distancia entre el promedio de los aspirantes de ingeniería y el resto de los aspirantes es de 4.88 errores estándar.

El IC del 95% [101.13, 102.87] indica que, dadas las muestras de 16 participantes, unas 95 de 100 de las muestras capturarían el valor verdadero de la población. Este $IC_{95\%}$ no captura al promedio de 100 entre sus límites (véase Figura 2.6).

C. TEST T PARA DOS MUESTRAS INDEPENDIENTES

Las comparaciones entre dos grupos diferentes (*i. e.*, independientes) son comunes en la biología y medicina, entre muchas otras disciplinas: cuando se compara el crecimiento de un grupo de plantas contra otro o el efecto de algún medicamento. De una manera similar, las y los investigadores educativos utilizan estas comparaciones. Por ejemplo, un experimento es el *epítome* de las comparaciones en la ciencia. En un experimento, se comparan los promedios de un grupo que recibió un tratamiento (*i. e.*, grupo tratamiento) contra los de otro grupo que posiblemente recibió solo un placebo (*i. e.*, grupo control). En la medida en que la diferencia entre ambos aparezca, se podría inferir un efecto del tratamiento. Si mediante el tratamiento algo cambió en el promedio del grupo tratamiento respecto al otro, se podría concluir que hubo un efecto positivo. Pudo suceder que nada ocurrió y los promedios resultaron iguales. Entonces, no hubo efecto de tratamiento (*i. e.*, neutral). Por otro lado, si el promedio del grupo tratamiento fue menor al del grupo control, hubo un efecto negativo.

Ejemplos de efectos positivo, neutral y negativo pueden acontecer en la investigación educativa en diseños experimentales (véase: Schneider *et al.*, 2007). Un ejemplo puede ser: Un grupo de estudiantes son seleccionadas y seleccionados para un experimento, en el cual se pone a prueba la efectividad de un tratamiento de matemáticas para mejorar los conocimientos en esta materia. De este grupo, se seleccionan al azar el grupo control y el grupo tratamiento. Se les mide antes de empezar el tratamiento para observar si son *grupos equivalentes* (*i. e.*, mismos niveles en una variable de interés: *e. g.*, conocimientos, inteligencia, etcétera) bajo la hipótesis:

$$\text{Hipótesis: } \bar{x}_{\text{control}} = \bar{x}_{\text{tratamiento}}$$

Donde:

- ⊗ \bar{x}_{control} = Promedio del grupo control antes del tratamiento
- ⊗ $\bar{x}_{\text{tratamiento}}$ = Promedio del grupo tratamiento antes del tratamiento

Entonces, se miden los conocimientos de matemáticas en ambos grupos y la evidencia apoya la hipótesis antes descrita (i. e., no hay diferencia entre sus promedios). Por lo tanto, se tiene lo que se llama en los experimentos: grupos equivalentes. Esto hace la comparación entre los grupos más válida. Dados estos resultados, se parte del mismo punto, así que se aplica el tratamiento al grupo que corresponde y un placebo al grupo control. Después de esto, se escribe la hipótesis alternativa:

$$\text{Hipótesis: } \bar{x}_{\text{control}} < \bar{x}_{\text{tratamiento}}$$

Siguiendo con el ejemplo heurístico, los resultados muestran una diferencia entre los grupos en cuestión al final del tratamiento: i. e., $\bar{x}_{\text{control}} < \bar{x}_{\text{tratamiento}}$. En otras palabras, los resultados apoyan esta hipótesis de que el tratamiento dio un resultado positivo.

A grandes rasgos, esta es una explicación conceptual del material que se va a cubrir en esta sección sobre poner a prueba una hipótesis nula con un Test t . Más al respecto: la estadística puesta a prueba es la posible existencia de una diferencia entre dos promedios de muestras independientes. Dados los resultados de esta prueba, puede pasar que se apoye o no la posible existencia de una diferencia entre los parámetros de dos poblaciones independientes de donde fueron tomadas las muestras en cuestión. Expresado de otra manera, lo anterior puede ser descrito:

Estadística de prueba (diferencia entre los promedios de dos muestras): $\bar{x}_1 - \bar{x}_2$

Donde:

- ⊗ \bar{x}_1 = Promedio de la muestra uno
- ⊗ \bar{x}_2 = Promedio de la muestra dos

Bajo el punto de partida, la hipótesis nula (la diferencia entre los parámetros de dos poblaciones) es cero: $\mu_1 - \mu_2 = 0$

Donde:

- ⊗ μ_1 = Promedio de la población uno
- ⊗ μ_2 = Promedio de la población dos

De las descripciones anteriores, se pueden inferir las hipótesis nula y alternativa, respectivamente:

$$H_0: \mu_1 = \mu_2$$
$$H_A: \mu_1 \neq \mu_2$$

Otra versión de la hipótesis nula y de la hipótesis alternativa es:

$$H_0: \mu_1 - \mu_2 = 0$$
$$H_A: \mu_1 - \mu_2 \neq 0$$

El poner a prueba la hipótesis nula permite estimar la diferencia entre la estadística de prueba y la diferencia entre los parámetros con el fin de tener suficiente evidencia para rechazar o no la hipótesis nula. Según Hinkle *et al.* (2003, p. 236), se usa la distribución de todas las *posibles diferencias* entre las muestras. Esto marca una conceptualización diferente con los test abordados anteriormente en la presente obra acerca de la significancia estadística. En estos test solamente se había empleado la distribución de muestras sin considerar una diferencia *per se*. Enfatizando, el nuevo concepto que se maneja aquí es:

La distribución de las diferencias de muestras (*sampling distribution of differences*) representa todas las posibles diferencias entre las muestras: $\bar{x}_1 - \bar{x}_2$.

En muchos casos, en la investigación educativa se desconoce la varianza de la población. En este caso, se utiliza la varianza de una muestra para estimar el error estándar de la distribución de muestras de las diferencias: $SE_{diferencias}$. Para poner a prueba las diferencias entre promedios, se emplea la ya antes mencionada distribución *t* que es análoga a la distribución de las diferencias de muestras. Hinkle *et al.* (2003, p. 239) dijeron que: “Conforme el tamaño de dos muestras de

sus respectivas poblaciones incrementa, la distribución de las diferencias de muestras contiene las siguientes características”:

- ⊗ La forma de la distribución de las diferencias de muestras, se aproxima a la curva normal
- ⊗ Esta distribución tiene una tendencia central: i. e., su promedio es igual al de la diferencia del promedio de dos poblaciones ($\mu_1 - \mu_2 = 0$)
- ⊗ La variabilidad de esta distribución tiene una desviación estándar que se llama error estándar de la diferencia entre promedios (Ecuación 2.3). Esto es:

$$\sigma_{\text{diferencia}} = \sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \quad (\text{Ecuación 2.3})$$

Donde:

- ⊗ $\sigma_{\text{diferencia}}$ = Error estándar de la diferencia entre promedios de las poblaciones
- ⊗ σ^2 = Parámetro que es igual a la varianza de la población uno y a la varianza de la población dos: $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Este es el Supuesto de *Homogeneidad* (i. e., las dos varianzas son iguales)
- ⊗ n_1 = Tamaño de la muestra uno
- ⊗ n_2 = Tamaño de la muestra dos

Nota: Las diferencias de las varianzas de las poblaciones ($\sigma_{\text{diferencia}}$) suelen ser desconocidas, así que en su defecto se utilizan las varianzas de las muestras en cuestión. Por el motivo anterior, la fórmula se ajusta a las varianzas de las diferencias entre las muestras de este modo (Ecuación 2.4):

$$s_{\text{diferencia}} = \sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \quad (\text{Ecuación 2.4})$$

Donde:

- ⊗ $s_{\text{diferencia}}$ = Error estándar de la diferencia entre promedios de las muestras
- ⊗ s^2 = Varianza combinada de ambas muestras

Nota: La varianza combinada de ambas muestras, se estima de la siguiente manera (Ecuación 2.5):

$$s^2 = \frac{\Sigma(x_{i1} - \bar{x}_1)^2 + \Sigma(x_{i2} - \bar{x}_2)^2}{n_1 + n_2 - 2} \quad (\text{Ecuación 2.5})$$

Donde:

- ⊗ x_{i1} = Cada uno de los valores del set de la muestra uno
- ⊗ \bar{x}_1 = Promedio de la muestra uno
- ⊗ x_{i2} = Cada uno de los valores del set de la muestra dos
- ⊗ \bar{x}_2 = Promedio de la muestra dos
- ⊗ n_1 = Tamaño de la muestra uno
- ⊗ n_2 = Tamaño de la muestra dos
- ⊗ 2 = Representa las dos muestras que se sustraen de la suma de las muestras para obtener los grados de libertad

D. SUPUESTOS DEL TEST T

Uno de estos supuestos es que las muestras fueron obtenidas de dos poblaciones independientes. Una parte de las muestras independientes es que hayan sido obtenidas al azar. Al obtener una muestra al azar, se puede hacer la generalización al resto de la población si esta tiene el tamaño adecuado y si es representativa de la población de interés; *ceteris paribus* (véase a Ponce, 2019, para el muestreo). Ahora, un texto especializado en muestreo es el de Sudman (1976). Además, la otra parte de la independencia es que las respuestas de una población a algún instrumento no tienen influencia en las de otra. Por ejemplo, un investigador desea comparar los puntajes de un examen de admisión con dos universidades. En este caso, las universidades son las dos poblaciones, de las cuales se van a obtener dos muestras al azar, respectivamente. Los puntajes de este examen de admisión no serían independientes si algunos estudiantes que presentaron el examen de una universidad tomaron fotografías de las preguntas del examen y las compartieron con los alumnos de la otra institución educativa. Esto sería asumiendo que los exámenes de admisión de ambas universidades sean iguales. En otras palabras, se admite que una población afecta

a la otra al compartir información. Lo anterior violaría el Supuesto de *Independencia*.

El segundo supuesto es que las varianzas de ambas poblaciones son iguales. Usualmente, la varianza de una población es desconocida, pero se puede estimar de la varianza de una muestra. Se tienen dos muestras que vienen de dos poblaciones con varianzas iguales. Por lo tanto, la varianza de estas muestras es igual también. Este supuesto se llama *Homogeneidad de la Varianza* y se puede resumir así:

$$\sigma_1^2 = \sigma_2^2$$

Donde:

- ⊗ σ_1^2 = Varianza de la población uno
- ⊗ σ_2^2 = Varianza de la población dos

Entonces,

$$s_1^2 = s_2^2$$

Donde:

- ⊗ s_1^2 = Varianza de la muestra uno
- ⊗ s_2^2 = Varianza de la muestra dos

Hinkle *et al.* (2003) explicaron que el efecto de violar este Supuesto de Homogeneidad depende de la diferencia que exista entre el tamaño de las muestras. Es decir, si el tamaño de las muestras es igual, entonces la violación al Supuesto de Homogeneidad no es seria. En contraparte, si las muestras no son del mismo tamaño ni tampoco las varianzas, se debe buscar otro procedimiento para poner a prueba la hipótesis nula. Lo anterior se resume de la siguiente manera:

Dado que se viola el Principio de Homogeneidad cuando las varianzas de dos muestras no son iguales:

$$s_1^2 \neq s_2^2.$$

Por otro lado, esta violación no es seria cuando se tienen muestras del mismo tamaño:

$$n_1 = n_2$$

Donde:

- ⊗ n_1 = Tamaño de la muestra uno
- ⊗ n_2 = Tamaño de la muestra dos

Aunque se asume que las varianzas de ambas poblaciones son fijas e iguales, se acepta que las varianzas no sean exactamente iguales en el sentido *estricto* de la palabra (e. g., $\sigma_1^2 \approx \sigma_2^2$: i. e., son aproximadamente lo mismo). Esta aceptación se refiere a que la diferencia entre las muestras, se debe a errores de muestreo. Debido a esa fluctuación del tamaño de las varianzas, se realiza un Test de Significancia Estadística llamado Test de Levene. Hinkle *et al.* (2003, p. 736) definieron el Test de Levene como un examen para poner a prueba la homogeneidad de las varianzas de las muestras; tanto el Test *t* como el análisis de la varianza de un factor (*one-way ANOVA*) y otros análisis para comparar promedios entre grupos (e. g., análisis multivariante de la varianza; véase: Hair *et al.*, 2014; Tabachnick, & Fidell, 2012). Este Test de Levene especifica una hipótesis nula y una alternativa. En el caso de un Test *t*, se tienen solo dos varianzas:

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_A: \sigma_1^2 \neq \sigma_2^2$$

Donde:

- ⊗ σ_1^2 = Varianza de la población uno
- ⊗ σ_2^2 = Varianza de la población dos

Otra forma de expresar las hipótesis anteriores es:

$$H_0: \sigma_1^2 / \sigma_2^2 = 1$$
$$H_A: \sigma_1^2 / \sigma_2^2 \neq 1$$

Como no se conocen estos parámetros de las poblaciones, se utilizan las varianzas de las muestras en cuestión:

$$H_0: s_1^2 = s_2^2$$
$$H_A: s_1^2 \neq s_2^2$$

Donde:

- ⊗ s_1^2 = Varianza de la muestra uno
- ⊗ s_2^2 = Varianza de la muestra dos

Asimismo, otra forma de expresar las hipótesis de las muestras es:

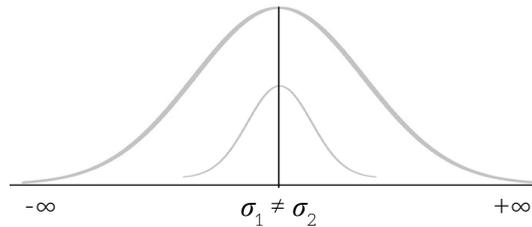
$$H_0: s_1^2 / s_2^2 = 1$$
$$H_A: s_1^2 / s_2^2 \neq 1.$$

Para continuar con el Test *t*, hay que pasar el Test de Levene, lo cual significa no rechazar la hipótesis nula. Al no tener evidencia para rechazarla, se asume que las varianzas de las muestras son homogéneas. El Test de Levene usa un Test *F* (para una discusión profunda acerca del tema, se recomienda ver a Kotz, 2006, p. 4132).

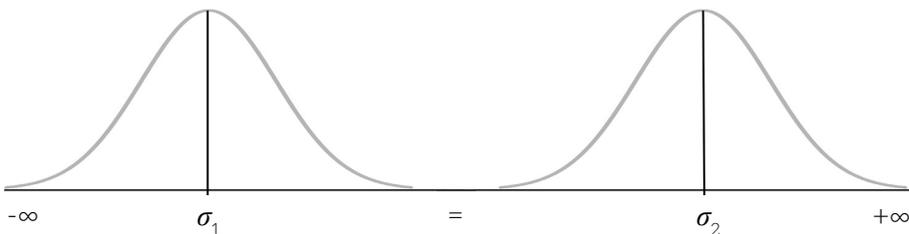
La Figura 2.7 señala algunas posibilidades cuando se está realizando el Test de Levene. Una posibilidad es el inciso A, donde dos muestras tienen diferentes varianzas (la extensión de ambas curvas normales es distinta). Como las varianzas son heterogéneas, se rechaza la hipótesis nula y no se puede llevar a cabo el Test *t*, porque no se cumplió con el supuesto de varianzas iguales. El caso contrario es el inciso B, el cual contiene dos muestras que tienen varianzas iguales; en este caso, no se rechaza la hipótesis nula y se puede proceder a efectuar un Test *t*. El inciso C contiene otro ejemplo con heterogeneidad de varianzas, donde se rechaza la hipótesis nula y no se puede

llevar a cabo el Test *t*. Existen correcciones a la violación del Supuesto de Homogeneidad que ofrecen paquetes, como *spss*, entre otros. Está más allá del propósito de este capítulo explicar en detalle las implicaciones de estas correcciones, pero se recomienda la guía de *spss* de Pallant (2016).

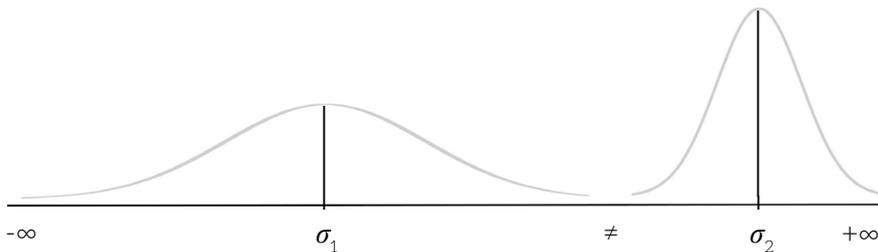
Figura 2.7 Muestras con homogeneidad y heterogeneidad de varianzas



a) Dos muestras con distribuciones normales y diferentes varianzas



b) Dos muestras con distribuciones normales y las mismas varianzas



c) Dos muestras con distribuciones normales y distintas varianzas

E. EJEMPLO DE TEST T PARA MUESTRAS INDEPENDIENTES Y EL *d* DE COHEN

Para realizar un Test *t* y del tamaño del efecto del *d* de Cohen, se pueden llevar a cabo los siguientes pasos (Tabla 2.2):

Un ejemplo heurístico para estos ocho pasos es que existen dos grupos de estudiantes que están llevando una clase de Biología con la misma profesora, quien tiene el propósito de comparar los dos grupos (alumnos de Medicina y de Enfermería), para observar si las diferencias en su promedio se deben probablemente al azar. Lo anterior sería el Test de Significancia con el Test *t* (también se le conoce como efecto estadístico). Asimismo, desea estimar la magnitud de la diferencia entre ambos promedios (también llamado efecto práctico [*d* de Cohen]).

- 1.º *Establecimiento de la o las preguntas de investigación:* ¿existe una diferencia estadísticamente significativa entre los promedios de Biología de los estudiantes de Medicina y de Enfermería? Esta pregunta sería para el Test de Significancia Estadística. Para el *d* de Cohen sería: ¿cuál es la magnitud de la diferencia entre los promedios de Biología de los alumnos de Medicina y de Enfermería?

La parte de la Tabla 2.5 sin sombreada muestra los datos de la clase, los promedios, la *SD*, el tamaño de las dos muestras (n_1 y n_2), la *SE* del promedio de cada muestra ($= SD_1 / \sqrt{n_1}$; $= SD_2 / \sqrt{n_2}$), la curtosis y la asimetría. Estos datos descriptivos proveen la oportunidad de observar los datos que se van a utilizar para las comparaciones de grupos y la posible distribución normal de los datos.

Tabla 2.5 Datos de la clase

Clase de Biología	Enfermería	Medicina	Enfermería $(x_{i1} - \bar{x}_1)^2$	Medicina $\Sigma (x_{i2} - \bar{x}_2)^2$
	9.0	8.0	$(9.0 - 9.5)^2 = .25$	$(8.0 - 8.5)^2 = .25$
	9.1	8.1	$(9.1 - 9.5)^2 = .16$	$(8.1 - 8.5)^2 = .16$
	9.2	8.2	$(9.2 - 9.5)^2 = .09$	$(8.2 - 8.5)^2 = .09$
	9.3	8.3	$(9.3 - 9.5)^2 = .04$	$(8.3 - 8.5)^2 = .04$
	9.4	8.4	$(9.4 - 9.5)^2 = .01$	$(8.4 - 8.5)^2 = .01$

Continúa...

Clase de Biología	Enfermería	Medicina		Enfermería $(x_{i1} - \bar{x}_1)^2$	Medicina $\Sigma (x_{i2} - \bar{x}_2)^2$
	9.5	8.5		$(9.5 - 9.5)^2 = 0$	$(8.5 - 8.5)^2 = 0$
	9.6	8.6		$(9.6 - 9.5)^2 = .01$	$(8.6 - 8.5)^2 = .01$
	9.7	8.7		$(9.7 - 9.5)^2 = .04$	$(8.7 - 8.5)^2 = .04$
	9.8	8.8		$(9.8 - 9.5)^2 = .09$	$(8.8 - 8.5)^2 = .09$
	9.9	8.9		$(9.9 - 9.5)^2 = .16$	$(8.9 - 8.5)^2 = .16$
	10	9.0		$(10 - 9.5)^2 = .25$	$(9.0 - 8.5)^2 = .25$
Promedio	9.5	8.5	Desarrollo de la Ecuación 2.4	Suma de la diferencia al cuadrado del grupo uno	Suma de la diferencia al cuadrado del grupo dos
Varianza: $s_{\text{de cada grupo}}^2$	0.11	0.11		$\Sigma (x_{i1} - \bar{x}_1)^2 = 1.1$	$\Sigma (x_{i2} - \bar{x}_2)^2 = 1.1$
				Suma de la diferencia al cuadrado de ambos grupos	
SD	0.33	0.33		$\Sigma (x_{i1} - \bar{x}_1)^2 + \Sigma (x_{i2} - \bar{x}_2)^2 = 2.2$	
SE	0.10	0.10		$s^2 = \Sigma (x_{i1} - \bar{x}_1)^2 + \Sigma (x_{i2} - \bar{x}_2)^2 / n_1 + n_2 - 2$	
Curtosis	-1.2	-1.2		Sustituyendo:	
Asimetría	0.66	0.66		$s^2 = 2.2 / 11 + 11 - 2$	
	$n_1 = 11$	$n_1 = 11$		$s^2 = 0.11$	

Nota: En este caso, la varianza a nivel de cada grupo ($s_{\text{de cada grupo}}^2 = 0.11$) coincidió con la varianza combinada de ambas muestras ($s^2 = 0.11$). Esto no implica que necesariamente al calcular una de las varianzas se esté calculando la otra. Habría que hacerlo por separado.

La parte sombreada de la Tabla 2.5 muestra cómo obtener la s^2 (la varianza combinada de ambas muestras) usando por pasos la siguiente fórmula:

$$s^2 = \Sigma (x_{i1} - \bar{x}_1)^2 + \Sigma (x_{i2} - \bar{x}_2)^2 / n_1 + n_2 - 2 \quad \text{(Ecuación 2.5)}$$

Luego, la s^2 se utiliza para calcular el $s_{\text{diferencia}}$ (error estándar de la diferencia entre promedios de las muestras). La ya mencionada Ecuación 2.4 muestra cómo se calculó este error:

$$s_{\text{diferencia}} = \sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \quad \text{(Ecuación 2.4)}$$

Se sustituyen los valores en la Ecuación 2.3:

$$s_{\text{diferencia}} = \sqrt{0.11 \left(\frac{1}{11} + \frac{1}{11} \right)}$$

$$s_{\text{diferencia}} = 0.14.$$

2.º *Test de Levene*: para este test también se puede adoptar un $\alpha = .05$ por ser el tradicional. Asimismo, se empleó el *software* SPSS para obtener evidencia para este test. La Tabla 2.6 muestra la hoja de resultados del Test de Levene. Las siguientes son las hipótesis nula y alternativa de este test:

$$H_0: s_1^2 = s_2^2$$

$$H_A: s_1^2 \neq s_2^2$$

Donde:

- ⊗ s_1^2 = Varianza del grupo uno
- ⊗ s_2^2 = Varianza del grupo dos

La Tabla 2.6 muestra los resultados del Test de Levene en SPSS. Se tiene un valor $F_{\text{calculado}} (< 0.0001)$ cercano a cero, que indica que las varianzas son similares. Sin embargo, habría que comparar con un valor $F_{\text{crítico}}$, pero está más allá del propósito de este capítulo explicar el Test F (para más información respecto a este test, se recomienda ver a Salkind, 2007, p. 33), así que se opta por comparar la $p_{\text{calculada}} (= 1)$ con el $\alpha (.05)$: i. e., $p_{\text{calculada}} (1) > \alpha (.05)$. Por lo tanto, no se rechaza la hipótesis nula y se asume que existe homogeneidad entre las varianzas de los dos grupos en cuestión.

Tabla 2.6 Hoja de resultados de un Test de Levene y Test t

Test de Levene de varianzas iguales		
	$F_{\text{calculado}}$	$p_{\text{calculada}}$
Supuesto de varianzas iguales	< .0001	1

- 3° Derivar las hipótesis de la pregunta de investigación (es de dos colas, porque no hay algún precedente que indique que alguno de los promedios es más grande que el otro).

Para las poblaciones:

$$\begin{aligned}H_0: \mu_{\text{enfermería}} &= \mu_{\text{medicina}} \\H_A: \mu_{\text{enfermería}} &\neq \mu_{\text{medicina}}\end{aligned}$$

Para las muestras:

$$\begin{aligned}H_0: \bar{x}_{\text{enfermería}} &= \bar{x}_{\text{medicina}} \\H_A: \bar{x}_{\text{enfermería}} &\neq \bar{x}_{\text{medicina}}\end{aligned}$$

- 4° Establecer el criterio para rechazar la hipótesis nula: el nivel del alfa (α). Se establece el tradicional $\alpha = .05$. Dados los grados de libertad ($df = n_1 + n_2 - 2$; i. e., $11 + 11 - 2 = 20$), dos colas y el alfa antes seleccionado ($\alpha / 2 = .025$), se encuentra en una tabla t que el $t_{\text{crítico}} = 2.086$. Sin embargo, se recomienda a las interesadas e interesados en la investigación educativa utilizar la literatura sobre el tema para tener un sustento para justificar cierto criterio.
- 5° Calcular el valor de la prueba estadística (Test Statistic); i. e., en este caso, el $t_{\text{calculado}}$ y su correspondiente p_{cal} . En general, la fórmula en forma conceptual es:

Valor de prueba estadística = Estadística – Parámetro / Error estándar de la estadística

Para el Test t , esta Ecuación 2.6 sería de esta forma:

$$t_{\text{calculado}} = [(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)] / s_{\text{diferencia}} \quad (\text{Ecuación 2.6})$$

Basado en la hipótesis nula, la diferencia entre los promedios de las poblaciones es igual a cero: $\mu_1 - \mu_2 = 0$.

Sustituyendo con los valores antes calculados en la fórmula:

$$t_{\text{calculado}} = [(9.5 - 8.5) - 0] / 0.14 = 7.14.$$

Empleando la calculadora en línea para valores t (<http://www.socscistatistics.com/pvalues/tdistribution.aspx>) para encontrar la $p_{\text{cal.}}$: $p < .0001$ (Figura 2.8).

Figura 2.8 Calculadora del p para el valor t

P Value from T Score Calculator

This should be self-explanatory, but just in case it's not: your T Score goes in the T Score box, you stick your degrees of freedom in the *DF* box ($N - 1$ for single sample and dependent pairs, $(N_1 - 1) + (N_2 - 1)$ for independent samples), select your significance level and whether you're testing a one or two-tailed hypothesis (if you're not sure, go with the defaults), then press the button!

If you need to derive a T Score from raw data, [then you can find t test calculators here.](#)

T Score:
DF:

Significance Level:

- .01
 .05
 .10

One-tailed or two-tailed hypothesis?:

- One-tailed
 Two-tailed

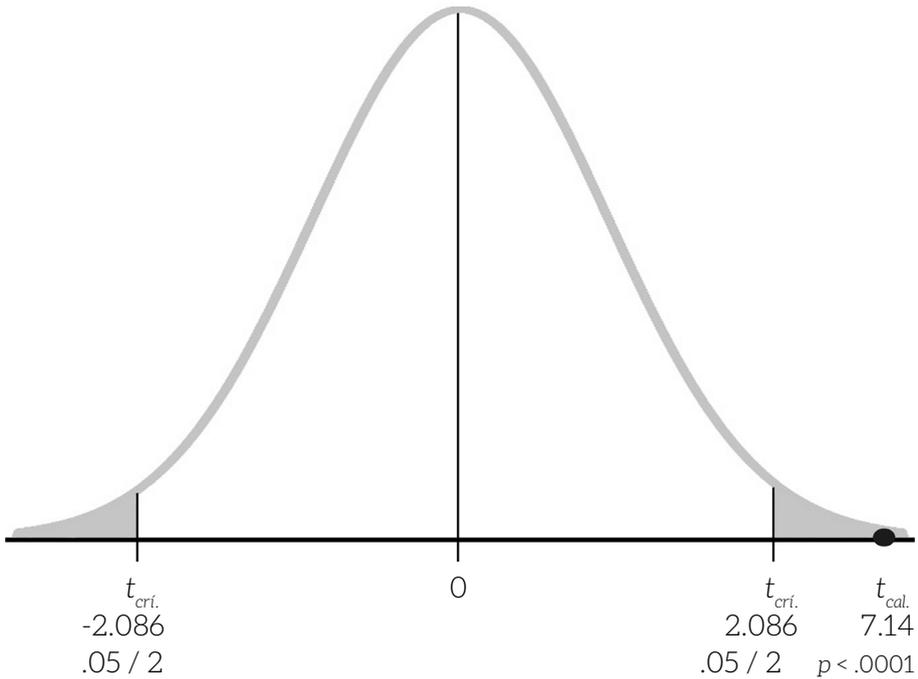
The p -value is $< .00001$.

The result is significant at $p < .05$.

Fuente: Social Science Statistics.

Asimismo, la Figura 2.9 muestra cómo quedarían en una distribución de muestras t los valores calculado y crítico, así como sus respectivas probabilidades, para ser comparados para el Test de Significancia Estadística.

Figura 2.9 Test de Significancia Estadística



6.º Llevar a cabo un intervalo de confianza: en este caso, el IC es para la diferencia entre promedios de las muestras (Ecuación 2.7), la cual es diferente a las fórmulas anteriores para estimar IC:

$$IC = (\bar{x}_1 - \bar{x}_2) \pm t_{crítico} (s_{diferencia}) \quad (\text{Ecuación 2.7})$$

Se adopta un IC tradicional del 95%, que corresponde al alfa del .05. También, se sustituyen los valores previamente calculados (i. e., $\bar{x}_1 = 9.5$; $\bar{x}_2 = 8.5$; $t_{crítico} = 2.086$; $s_{diferencia} = .14$):

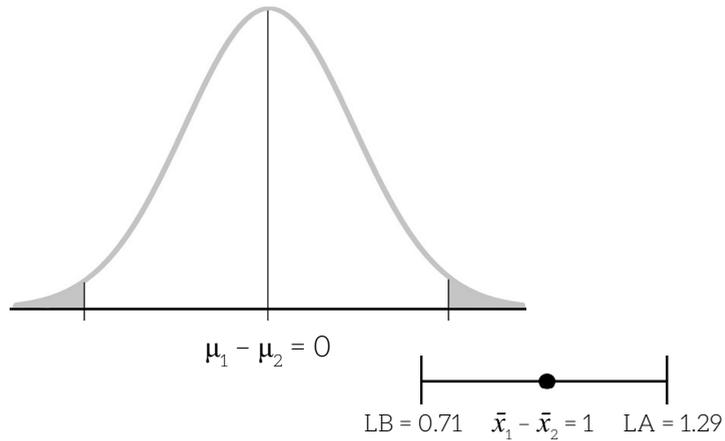
$$\text{Límite bajo: } IC_{95\%} = (9.5 - 8.5) - 2.086(.14) = 0.71$$

$$\text{Límite alto: } IC_{95\%} = (9.5 - 8.5) + 2.086(.14) = 1.29$$

El valor de la diferencia entre los promedios puede estar en este $IC_{95\%}$ [0.71, 1.29]. Este IC (Figura 2.10) no contiene al cero, que es hipo-

téticamente la diferencia entre los promedios de las dos poblaciones. Por no contener el IC al cero, esto apoya que la diferencia entre las poblaciones sea un número diferente de cero. Al no contener el cero y haber rechazado la hipótesis nula, esto explica la coherencia entre un Test de Significancia Estadística y un IC (Hinkle *et al.*, 2003, p. 246).

Figura 2.10 Intervalos de confianza del 95% para la diferencia entre promedios de las muestras



7° Cálculo del *d* de Cohen: *recordando*: un tamaño de efecto, como el *d*, mide la magnitud de la diferencia expresada en unidades de desviación estándar entre dos promedios. También, Cohen (1988) dijo que depende del tamaño del efecto la posible existencia de un fenómeno; es decir, si el *d* es grande esto indica un efecto, el cual es una inferencia (se recomienda ver a Hurley, 2014). La fórmula de este efecto es (Ecuación 2.8):

$$d = (\bar{x}_1 - \bar{x}_2) / SD^* \quad \text{(Ecuación 2.8)}$$

Donde:

- ⊗ *d* = Magnitud de la diferencia entre dos promedios expresada en un coeficiente estándar
- ⊗ \bar{x}_1 = Promedio del grupo uno

- ⊗ \bar{x}_2 = Promedio del grupo dos
- ⊗ SD^* = Desviación estándar del grupo uno y del grupo dos cuando son supuestamente iguales: *i. e.*, $SD^* = SD_1 = SD_2$.

La anterior igualdad se basa en que las varianzas de las poblaciones de donde se obtuvieron estas muestras son iguales; por lo tanto, sus coeficientes de desviación estándar respectivos son iguales también, porque se derivaron de las varianzas de las poblaciones: *i. e.*, una desviación estándar es igual a la raíz cuadrada de una varianza: $SD = \sqrt{s^2}$. Sin embargo, dos coeficientes de SD de dos muestras diferentes probablemente no sean iguales por un error de muestreo. Una corrección es sumar ambos coeficientes de SD al cuadrado, dividirlos entre dos y sacarles la raíz cuadrada: *i. e.*, $SD^* = \sqrt{(SD_1^2 + SD_2^2) / 2}$. Para más formulas sobre el d de Cohen y otros coeficientes que miden la magnitud de la diferencia entre dos promedios está el Apéndice E de este libro.

Se sustituyen los coeficientes del ejemplo en la Ecuación (2.8):

$$d = (9.5 - 8.5) / 0.33 = 3.3.$$

8.º *Interpretar los resultados:* los resultados (Tabla 2.7) muestran que el pregrado de Enfermería tuvo un promedio más alto que el de Medicina, por lo que existe una diferencia estadísticamente significativa entre ambos promedios ($t_{\text{calculado}} > t_{\text{crítico}}$; $p_{\text{cal}} < \alpha$); por lo tanto, se rechaza la hipótesis nula y la evidencia apoya la hipótesis alternativa. La probabilidad de que esta diferencia de un punto entre los promedios ocurriera al azar (bajo el supuesto de promedios iguales de ambas poblaciones: *i. e.*, una hipótesis nula cierta) es menor a .0001 cuando se asume que la hipótesis nula es cierta a nivel de las poblaciones de donde se tomaron las muestras. Esta es una probabilidad cercana a 0, así que se asume que no ocurrió por casualidad. Habría que repetir varias veces este análisis para ver si emerge un patrón, ya que los Test de Significancia Estadística fueron diseñados para el largo plazo.

El intervalo de confianza del 95% muestra que entre sus límites [0.71, 1.29] no está contemplado el cero (Figura 2.10). Esto apoya al Test

de Significancia Estadística que también indica que la diferencia entre ambos promedios no es probablemente cero.

El tamaño del efecto ($d = 3.3$) también apoya la idea de que la diferencia entre ambos promedios no es cero, porque es un efecto grande (Tabla 2.7): i. e., $d = 0.20$ = efecto chico; $d = 0.50$ = efecto mediano; y $d = 0.80$ = efecto grande (Cohen, 1988). Este efecto grande se podría interpretar como que los pregrados en los que están estos alumnos tienen algo que ver en sus puntajes, asumiendo para ambos grupos mismos niveles de motivación, horas de estudio, capacidad intelectual, entre muchas otras variables que tendrían alguna injerencia. En este sentido, el tamaño del efecto es una inferencia que le corresponde a la investigadora o investigador hacer y sustentar con evidencias (véase: Hurley, 2014, para crear inferencias lógicas).

Tabla 2.7 Resumen de resultados

\bar{x}_1	\bar{x}_2	SD^*	d de Cohen	n	$IC_{95\%}$ de la diferencia	
9.5	8.5	0.33	3.3	22		
$t_{calculado}$	df	$p_{calculada}$	Diferencia entre los promedios	$S_{diferencia}$	Límite bajo	Límite alto
7.071	20	< .0001	1	0.14	0.71	1.29
$t_{critico}$		Alfa para dos colas				
2.086		.05 / 2 = .025				

Una advertencia para las y los investigadores es que si se lleva a cabo una comparación de grupos con una muestra grande probablemente se obtendrán resultados estadísticamente significativos; es decir, los Test de Significancia Estadística son sensibles al tamaño de la muestra. Por otro lado, se podrían tener resultados de comparaciones que muestren no ser estadísticamente significativos, pero con un tamaño de efecto grande. Asimismo, puede haber significancia estadística y tener un tamaño de efecto pequeño. La recomendación es no solo considerar el Test de Significancia Estadística para hacer conclusiones. Habría que buscar en la literatura los tamaños de efecto para ver la magnitud de la diferencia y cuáles serían sus implicaciones.

F. TEST T PARA DOS MUESTRAS DEPENDIENTES

En la investigación educativa puede ser común el medir algún grupo de participantes en más de una ocasión: medidas repetidas. Un ejemplo de esto podría ser cuando se está estudiando la trayectoria académica de diferentes grupos de alumnos (variable independiente) a través del tiempo, utilizando calificaciones de materias como la variable dependiente. Esto sería para tratar de observar algún tipo de patrón; es decir, si cierto grupo consistentemente está arriba, otro estaría en medio y el resto, abajo. Otra forma de medir a un grupo sería cuando se emplea un pretest y un postest. Esto puede suceder cuando se realiza algún tipo de estudio experimental: se aplica algún tipo de instrumento para estimar dónde está un grupo en cuestión de algún conocimiento o percepción, entre otros; se emplea algún tratamiento para incrementar el conocimiento o cambiar alguna percepción; se vuelve a aplicar el instrumento para observar un posible cambio; y se analizan los resultados mediante un Test de Significancia Estadística y de un tamaño de efecto, como el d de Cohen.

En pocas palabras, el ejemplo anterior muestra un solo grupo medido en dos ocasiones. Se les llama muestras dependientes, porque los resultados de sus mediciones dependen del mismo grupo de participantes. Un ejemplo es cuando al mismo grupo, se le mide la presión sanguínea en dos ocasiones: *e. g.*, una antes de un tratamiento y la otra, después de este. Las mediciones dependerían de cada uno de los individuos participantes. Asimismo, como en la correlación, las mediciones están en pares: un par pertenece exclusivamente a un participante.

Se usa una sola población, porque se trata de una sola muestra medida en dos ocasiones para el Test t de muestras dependientes. El test para la hipótesis nula, se establece en forma similar al Test de Muestras Independientes: Promedio de la primera medición = Promedio de la segunda medición. Sin embargo, hay una pequeña diferencia al establecer la hipótesis nula y es que se adiciona la letra griega delta mayúscula (δ), que significa cambio. En el caso de la hipótesis nula esta letra significa *sin* cambio: *i. e.*, $\delta = 0$. La siguiente hipótesis nula establece que la diferencia entre los promedios de las mediciones es cero y, por lo tanto, el cambio entre ambas es cero. Con más precisión: la letra

delta (δ) significa el *promedio de las diferencias* que hay entre una y otra medición: e. g., pretest y posttest. La hipótesis nula se establece de la siguiente manera:

$$H_0: \delta = \mu_{\text{primera medición}} - \mu_{\text{segunda medición}} = 0$$

Donde:

- ⊗ δ = Promedio de las diferencias que hay entre la primera y la segunda medición de una población
- ⊗ $\mu_{\text{primera medición}}$ = Promedio de la primera medición: e. g., pretest
- ⊗ $\mu_{\text{segunda medición}}$ = Promedio de la segunda medición: e. g., posttest

La alternativa establece que el promedio de las diferencias y la resta entre los promedios de las mediciones de la población no son iguales a cero:

$$H_A: \delta = \mu_1 - \mu_2 \neq 0$$

A nivel de las muestras dependientes, las hipótesis se establecen de la siguiente manera:

$$H_0: \bar{d} = \bar{x}_1 - \bar{x}_2 = 0$$
$$H_A: \bar{d} = \bar{x}_1 - \bar{x}_2 \neq 0$$

Donde:

- ⊗ \bar{d} = Promedio de las diferencias que hay entre un pretest y un posttest para las muestras (no hay que confundirlo con el d de Cohen)
- ⊗ \bar{x}_1 = Promedio de la muestra uno
- ⊗ \bar{x}_2 = Promedio de la muestra dos

Tanto en el caso de las muestras independientes como en el de las dependientes, se utiliza una distribución que es análoga a la distribución de muestras: la *distribución de las diferencias entre los promedios*

de muestras dependientes (*distribution of differences between the means of dependent samples*). Al respecto de esta distribución, Hinkle *et al.* (2003, p. 265) explican que cuando el tamaño de las muestras dependientes se incrementa, su correspondiente distribución de muestras adquiere las siguientes propiedades:

- ⊗ La *forma* que las muestras toman, se aproxima a la distribución normal
- ⊗ La mayoría de las muestras tienden hacia el centro de la distribución y su promedio de diferencias es igual a cero: *i. e.*, $\delta = 0$
- ⊗ Para estimar la variabilidad de esta distribución, se utiliza el error estándar de la diferencia entre mediciones dependientes (Ecuación 2.9):

$$\sigma_{\bar{d}} = \sigma_d / \sqrt{n} \quad (\text{Ecuación 2.9})$$

Donde:

- ⊗ $\sigma_{\bar{d}}$ = Error estándar de la diferencia entre mediciones dependientes de una población
- ⊗ σ_d = Desviación estándar de la diferencia de mediciones dependientes de una población
- ⊗ n = Tamaño de la muestra

Nota: Sin embargo, como difícilmente se saben los valores de una población, se emplean los valores conocidos de las muestras (Ecuación 2.10):

$$SE_{\bar{d}} = SD_d / \sqrt{n} \quad (\text{Ecuación 2.10})$$

Donde:

- ⊗ $SE_{\bar{d}}$ = Error estándar de la diferencia entre mediciones dependientes de las muestras

- ⊗ SD_d = Desviación estándar de la diferencia de mediciones dependientes de las muestras
- ⊗ n = Tamaño de la muestra

Nota: La Ecuación 2.9 (error estándar de la diferencia entre mediciones dependientes de las muestras), se explica más adelante.

G. EJEMPLO DE MUESTRAS DEPENDIENTES USANDO EL TEST T Y EL d DE COHEN

El Test de Muestras Dependientes es similar al de las Muestras Independientes, como se verá a continuación. Un ejemplo heurístico para muestras dependientes es cuando se tiene un grupo de estudiantes seleccionados al azar de una población. Se les escogió para aplicarles un tratamiento que consiste en darles un par de horas diarias de tutorías en matemáticas. El propósito es comparar las calificaciones de este grupo antes (pretest) y después (postest) del tratamiento, para ver si hubo una diferencia. Se lleva a cabo un Test t de muestras dependientes para observar si las diferencias entre el pretest y el postest, se deben probablemente al azar (efecto estadístico). También, se desea estimar la magnitud de la diferencia entre ambos promedios, también llamado efecto práctico (d de Cohen), que, en este caso, es un coeficiente modificado (d^*), como se observa más adelante.

- 1.º *Establecimiento de la o las preguntas de investigación:* ¿existe una diferencia estadísticamente significativa entre el pretest y el postest? Esta interrogante sería para el Test de Significancia Estadística. Para el d^* modificado de Cohen sería: ¿cuál es la magnitud de la diferencia entre el pretest y el postest?

La Tabla 2.8 muestra los datos de la clase, las sumas de las columnas, los promedios del pretest y el postest, la SD , el tamaño de la muestra (n), la curtosis y la asimetría. Estos datos descriptivos proveen la oportunidad de observar los coeficientes que se van a utilizar para las comparaciones de grupos y su posible distribución normal.

Asimismo, la Tabla 2.8 enseña la diferencia entre los valores del pretest y el posttest en la columna: Diferencia: $x_2 - x_1$. Para calcular esta diferencia, las columnas fueron ordenadas de la siguiente manera: primero el posttest y, luego, el pretest, para que, al restarlos, resultara un set de números positivos para hacer el ejemplo más simple de explicar. También, la diferencia se eleva al cuadrado para ser utilizada en algunas de las siguientes fórmulas: Diferencia² = $(x_2 - x_1)^2$.

Tabla 2.8 Datos y estadísticas del pretest y posttest

Núm.	Postest	Pretest	Diferencia: $x_2 - x_1$	Diferencia ² : $(x_2 - x_1)^2$
1	8.20	8.00	0.20	0.04
2	8.30	8.10	0.20	0.04
3	8.25	8.20	0.05	0.0025
4	8.50	8.30	0.20	0.04
5	8.60	8.40	0.20	0.04
6	8.50	8.50	0	0
7	8.80	8.60	0.20	0.04
8	8.80	8.70	0.10	0.01
9	8.90	8.80	0.10	0.01
10	9.00	8.90	0.10	0.01
11	9.20	9.00	0.20	0.04
Suma	95.05	93.50	1.55	$\Sigma d^2 = 0.273$
Promedio	8.64	8.50	$\bar{d} = 1.55 / 11 = 0.14091$	
Varianza	.11	.11	$(\Sigma d)^2 = (1.55)^2 = 2.403$	
SD	.33	.33		
SE	.10	.10		
Curtosis	1	-1.2		
Asimetría	.20	0		
*r	.975			
	p < .001			

Nota: \bar{d} = Promedio de la diferencia entre el pretest y el posttest;

$$\bar{d} = \Sigma (x_2 - x_1) / n.$$

En este caso: $\bar{d} = 0.14091$ [los coeficientes Σd^2 y $(\Sigma d)^2$ son explicados en la siguiente sección]; *r = Correlación Producto-Momento de Pearson (véase: Ponce, 2020)

- 2.º** Derivar las hipótesis de la pregunta de investigación (se usan dos colas, porque puede ser que el tratamiento haya sido contraproducente). Por otro lado, en un modelo experimental donde se aplica algún tipo de tratamiento con ciertas expectativas, se esperaría que el promedio del postest fuera más alto que el del pretest en ciertos casos (e. g., tratamiento para incrementar el conocimiento en una materia: usar la cola derecha de la distribución t para el Test de Significancia Estadística). Además, cabe la posibilidad de que se espere que el promedio del postest sea menor al promedio del pretest: e. g., después de un tratamiento destinado para disminuir el ausentismo (usar la cola izquierda).

Para las poblaciones:

$$H_0: \delta = \mu_{pretest} - \mu_{postest} = 0$$
$$H_A: \delta = \mu_{pretest} - \mu_{postest} \neq 0$$

Para las muestras:

$$H_0: \delta = \bar{x}_{pretest} - \bar{x}_{postest} = 0$$
$$H_A: \delta = \bar{x}_{pretest} - \bar{x}_{postest} \neq 0.$$

- 3.º** Establecer el criterio para rechazar la hipótesis nula: el nivel del alfa (α). Se establece el tradicional $\alpha = .05$. Dados los grados de libertad ($n - 1$; i. e., $11 - 1 = 10$ df), dos colas y el alfa antes seleccionado ($\alpha / 2 = .025$), se encuentra en una Tabla t que el valor $t_{crítico} = 2.228$. Sin embargo, se recomienda a las interesadas e interesados en la investigación educativa utilizar la literatura sobre el tema para tener un sustento para justificar cierto criterio.
- 4.º** Calcular el valor de la prueba estadística (Test Statistic); i. e., en este caso, el $t_{calculado}$ y su correspondiente probabilidad $_{calculada}$. En general, la fórmula en forma conceptual es:

Valor de prueba estadística = Estadística - Parámetro / Error estándar de la estadística

Para el Test t , esta Ecuación sería de esta forma (2.11):

$$t_{\text{calculado}} = [(\bar{x}_{\text{medición 1}} - \bar{x}_{\text{medición 2}}) - (\mu_{\text{medición 1}} - \mu_{\text{medición 2}})] / SE_{\bar{a}} \text{ (Ecuación 2.11)}$$

Con base en la hipótesis nula, la diferencia entre los promedios de las mediciones de la población es igual a cero: $\delta = \mu_1 - \mu_2 = 0$.

Existe una serie de pasos para calcular el error estándar de la diferencia entre mediciones dependientes de las muestras ($SE_{\bar{a}}$). Se puede comenzar por estimar la varianza de la diferencia entre las mediciones uno y dos (i. e., pretest y postest; s_d^2 , Ecuación 2.12):

$$s_d^2 = \frac{[\Sigma d^2 - \frac{(\Sigma d)^2}{n}]}{n - 1} \text{ (Ecuación 2.12)}$$

Donde:

- ⊗ s_d^2 = Varianza de la diferencia entre el pretest y el postest
- ⊗ Σd^2 = Elevar al cuadrado las diferencias ($x_2 - x_1$), para después sumarlas. Para el presente ejemplo es: 0.273 (véase Tabla 2.8, donde se realizó su cálculo)
- ⊗ $(\Sigma d)^2$ = Sumar las diferencias, primero, para después elevarlas al cuadrado (véase Tabla 2.8, donde se efectuó su cálculo). Para el presente ejemplo es: $(\Sigma d)^2 = (1.55)^2 = 2.403$
- ⊗ n = Tamaño de la muestra

Se sustituyen las variables por los valores de la Tabla 2.8 en la Ecuación 2.12:

$$\begin{aligned} s_d^2 &= 0.273 - [(1.55)^2 / 11] / 11 - 1 \\ s_d^2 &= 0.273 - [2.403 / 11] / 10 \\ s_d^2 &= (0.273 - 0.22) / 10 \\ s_d^2 &= 0.005. \end{aligned}$$

Ya que se tiene la varianza de la diferencia entre el pretest y el postest (s_d^2), se puede obtener la desviación estándar de la diferencia con solo calcular la raíz cuadrada de la Ecuación 2.11 (Ecuación 2.13):

$$SD_d = \sqrt{\frac{[\sum d^2 - \frac{(\sum d)^2}{n}]}{n-1}} \quad (\text{Ecuación 2.13})$$

Donde:

⊗ SD_d = Desviación estándar de la diferencia

En síntesis, se le saca la raíz cuadrada de la varianza: i. e., $s_d^2 = .005$. Esto es:

$$SD_d = \sqrt{0.005} = 0.0707.$$

Finalmente, el error estándar de la diferencia se obtiene de la siguiente manera (Ecuación 2.10):

$$SE_{\bar{a}} = SD_d / \sqrt{n} \quad (\text{Ecuación 2.10})$$

Donde:

⊗ $SE_{\bar{a}}$ = Error estándar de la diferencia

⊗ n = Tamaño de la muestra

Se sustituyen las variables de la fórmula con los valores del ejemplo:

$$SE_{\bar{a}} = .0707 / \sqrt{11} = 0.0707 / 3.17 = 0.0223.$$

Para obtener la estadística de prueba, se emplea la siguiente Ecuación 2.14, que es una versión simplificada de la ya antes descrita (Ecuación 2.10):

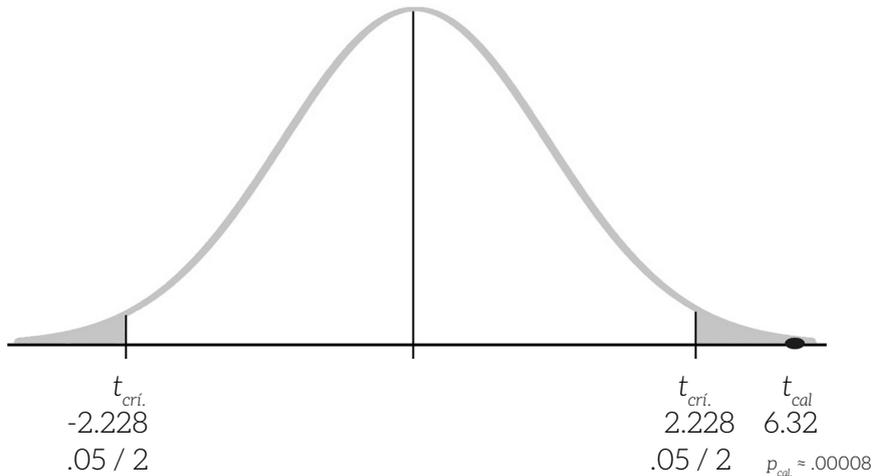
$$t_{\text{calculada}} = (\bar{d} - \delta) / SE_{\bar{a}} \quad (\text{Ecuación 2.14})$$

Se sustituye por los valores:

$$t_{\text{calculada}} = (0.14091 - 0) / 0.0223 = 6.32.$$

La Figura 2.11 muestra el valor de la prueba estadística junto con los valores correspondientes al criterio para rechazar la hipótesis nula. La Figura 2.12 ejemplifica cómo se calculó la p .

Figura 2.11 Test de Significancia Estadística



La Figura 2.12 muestra la forma en la cual se pueden calcular en línea con la *T-Test Calculator for 2 Dependent Means* las estadísticas estimadas manualmente en este capítulo: <https://www.socscistatistics.com/tests/ttestdependent/default2.aspx>

Figura 2.12 Calculadora del Test *t* para muestras dependientes

T-Test Calculator for 2 Dependent Means

The value of *t* is 6.35438.

Explanation of results

The output of this calculator is pretty straightforward. The values of *t* and *p* appear at the bottom of the page. If the text is blue, your result is significant; if it's red, it's not. The only thing that might catch you out is the way that we've rounded the data. The data you see in front of you, apart from the *t* and *p* values, has been rounded to 2 significant figures. However, we did not round when actually calculating the values of *t* and *p*. This means that if you try to calculate these values on the basis of the summary data provided here, you're likely going to end up with a slightly different - and less accurate - result.

Treatment 1	Treatment 2	Diff(T2 - T1)	Dev(Diff - M)	Sq. Dev
8.00	8.20	0.2	0.06	0
8.10	8.30	0.2	0.06	0
8.20	8.25	0.05	-0.09	0.01
8.30	8.50	0.2	0.06	0
8.40	8.60	0.2	0.06	0
8.50	8.50	0	-0.14	0.02
8.60	8.80	0.2	0.06	0
8.70	8.80	0.1	-0.04	0
8.80	8.90	0.1	-0.04	0
8.90	9.00	0.1	-0.04	0
9.00	9.20	0.2	0.06	0
		M: 0.14		S: 0.05

Significance Level:

- 0.01
- 0.05
- 0.10

One-tailed or two-tailed hypothesis?:

- One-tailed
- Two-tailed

Difference Scores Calculations

Mean: 0.14
 $\mu = 0$
 $S^2 = SS_{df} = 0.05/(11-1) = 0.01$
 $S^2_M = S^2/N = 0.01/11 = 0$
 $S_M = \sqrt{S^2_M} = \sqrt{0} = 0.02$

T-value Calculation

$t = (M - \mu)/S_M = (0.14 - 0)/0.02 = 6.35$

The value of *t* is 6.35438. The value of *p* is .00008. The result is significant at $p < .05$.

Calculate Reset

Fuente: Social Science Statistics.

Nota: Existe una diferencia entre el valor de la prueba estadística (calculada manualmente = 6.32 y en la calculadora en línea = 6.35), porque esta diferencia radica en el número de dígitos usados en los algoritmos.

5.º Calcular el valor del IC del 95% de confianza: la ecuación para este intervalo de confianza es (2.15):

$$IC_{95\%} = \bar{d} \pm t_{\text{crítico}} (SE_{\bar{d}}) \quad (\text{Ecuación 2.15})$$

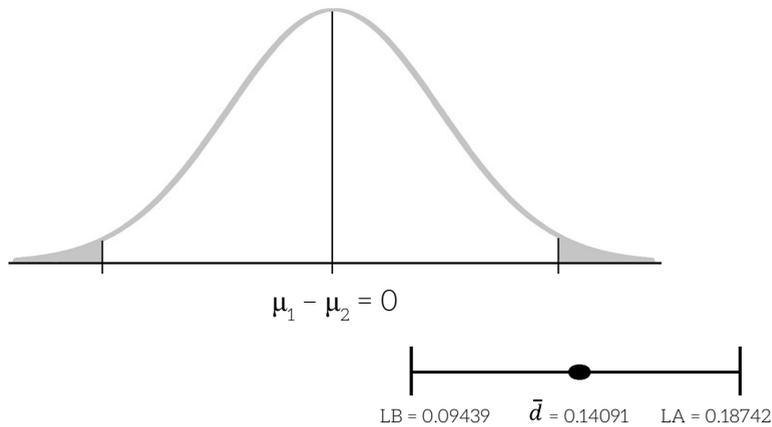
Sustituyendo:

$$\text{Límite bajo del IC}_{95\%}: 0.14091 - 2.086 (0.0223) = 0.09439$$

$$\text{Límite alto del IC}_{95\%}: 0.14091 + 2.086 (0.0223) = 0.18742$$

$$\text{IC}_{95\%}: [0.09439, 0.18742].$$

Figura 2.13 Intervalos de confianza del 95%



6.º Cálculo del d de Cohen: *recordando*: un tamaño de efecto, como el d , mide la magnitud de la diferencia expresada en unidades de desviación estándar entre dos promedios. Por otro lado, Hinkle *et al.* (2003) recomendaron una modificación al tradicional d de Cohen: $(\bar{x}_1 - \bar{x}_2) / SD$. Esta modificación es la siguiente (Ecuación 2.16):

$$d^* = \frac{\bar{d}}{SD_d} \quad (\text{Ecuación 2.16})$$

Donde:

- ⊗ d^* = d de Cohen modificado para calcular el tamaño del efecto en muestras dependientes
- ⊗ \bar{d} = Promedio de las diferencias entre los valores de un test (en este caso, el pretest) y los de otro (postest)
- ⊗ SD_d = Desviación estándar de la diferencia

$$d^* = 0.14091 / 0.0707 = 1.99.$$

Nota: El Apéndice E (Tabla E-2) muestra otras maneras de calcular el d de Cohen en medidas repetidas. El Apéndice F (Figura F-3) muestra cómo obtener un IC del 95% para el d de Cohen, así como una discusión acerca de los resultados de un par de calculadoras en línea.

7° *Interpretar los resultados:* la Tabla 2.9 muestra que el postest tuvo un promedio (8.64) más alto que el pretest (8.50) y que existe una diferencia estadísticamente significativa entre ambos test ($t_{\text{calculado}} > t_{\text{crítico}}$; $p < \alpha$); por lo tanto, se rechaza la hipótesis nula y la evidencia apoya la hipótesis alternativa. La probabilidad de que esta diferencia de 0.14091 ocurriera al azar (bajo el supuesto de que el pretest y postest son iguales en la población: *i. e.*, una hipótesis nula cierta) es aproximadamente .00008. Esta es una probabilidad cercana a cero, así que se asume que no ocurrió por casualidad: aquí puede haber un efecto verdadero y habría que seguir haciendo repeticiones del estudio.

El IC del 95% muestra que entre sus límites [0.09439, 0.18742] no está contemplado el cero (Tabla 2.9). Esto apoya al Test de Significancia Estadística que también indica que la diferencia entre ambos promedios no es probablemente cero.

El tamaño del efecto ($d = 1.99$) también apoya la idea de que la diferencia entre ambos promedios no es cero, porque es un efecto grande (Tabla 2.9): *i. e.*, $d = 0.20$ = efecto chico; $d = 0.50$ = efecto mediano; y $d = 0.80$ = efecto grande. Este efecto grande se podría interpretar como que el tratamiento de horas diarias de tutorías en matemáticas para pregrado funcionó en términos de incrementar las calificaciones en esta materia; *ceteris paribus*. En este sentido, el tamaño del efecto es una inferencia que le corresponde hacer a la investigadora o investigador y sustentarla con evidencias.

Tabla 2.9 Resumen de resultados

$\bar{x}_{pretest}$	$\bar{x}_{posttest}$	\bar{d}	SD_d	d^* de Cohen		
8.50	8.64	0.14091	0.0707	1.99		
$SD_{pretest}$	$SD_{posttest}$	r				
0.33	0.33	.975 $p < .001$				
$t_{calculado}$	df	$p_{cal.}$	\bar{d}	$SE_{\bar{d}}$	$IC_{95\%}$ de la diferencia	
6.67	10	$\approx .00008$	0.14091	0.0223	Límite bajo	Límite alto
$t_{critico}$	Alfa para dos colas					
2.228	.05 / 2 = .025					

PROBLEMAS PARA RESOLVER DEL CAPÍTULO 2

Siguiendo los pasos antes descritos en este segundo capítulo para resolver la siguiente serie de problemas, se recomienda desarrollarlos manualmente, con SPSS y con las calculadoras en línea:

- ⊗ Para estimar la probabilidad ^{calculada}, se usa la calculadora para el Test t en línea para Una Sola Muestra: <http://www.socscistatistics.com/tests/tsinglesample/Default2.aspx>
 - ⊗ Para estimar la probabilidad ^{calculada}, se utiliza la calculadora para el Test t en línea para Muestras Independientes y Dependientes: <http://www.socscistatistics.com/pvalues/tdistribution.aspx>
 - ⊗ Para estimar un Test de Muestras Dependientes: <https://www.socscistatistics.com/tests/ttestdependent/default2.aspx>
- a) Estadísticas descriptivas.
 - b) Establecimiento de la o las preguntas de investigación.
 - c) Derivar las hipótesis de la pregunta de investigación.
 - d) Establecer el criterio para rechazar la hipótesis nula: el nivel del alfa (α). Lo recomendable es el coeficiente de .05 por la tradición de su uso.
 - e) Calcular el valor de la prueba estadística (*Test Statistic*).
 - f) Llevar a cabo un intervalo de confianza.
 - g) Calcular el tamaño del efecto (el d de Cohen).
 - h) Interpretar los resultados.
 - i) Otras preguntas más allá del contenido de este segundo capítulo.

Nota: La idea de hacerlo en tres diferentes formas es para desarrollar las habilidades en estas tres modalidades, así como para observar hasta dónde los resultados de los tres métodos convergen. Una recomendación más sería graficar la distribución de los datos, obtener estadísticas de tendencia central (promedio, mediana y moda), así como de dispersión (rango, varianza y SD). Por último, ayudaría el graficar las distribuciones para poner a prueba la hipótesis nula, así como los intervalos de confianza.

1. Se contabilizaron los fonemas que pueden pronunciar un grupo de niñas y niños de 5 años del grupo A de un kínder. La Tabla Problema 1 del Capítulo 2 indica el número de aciertos que tuvieron. Es sabido que los infantes de esta edad tienen un promedio de aciertos de 7 palabras de 15 posibles. Algunas de las preguntas de investigación serían: a) ¿Cuáles son las estadísticas descriptivas de la muestra?; e) ¿Existe una diferencia estadísticamente significativa entre el promedio de la muestra y el promedio de la población ($\mu = 7$)?; y g) ¿Cuál es el tamaño del efecto? Se recomienda siempre obtener algunas de las estadísticas descriptivas de una muestra como la de este problema (véase el Apéndice A para el uso de SPSS en estos cálculos y el Apéndice B para este tipo de análisis).

Otras preguntas más allá del contenido de este segundo capítulo serían: ¿cuáles son los fonemas que los pupilos de kínder deberían saber?; ¿existe una regla universal de los fonemas que se deben de conocer a ciertas edades?; y ¿cómo mejorar los déficits en conocimientos fonológicos?

Tabla Problema 1 del Capítulo 2

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

2. Existe un segundo grupo (B) al que se le aplica la misma evaluación de los fonemas. Ahora, se desea comparar el grupo A con el grupo B. Se recomienda siempre obtener algunas de las estadísticas descriptivas de una muestra como la de este problema (véase el Apéndice A para el uso de SPSS en estos cálculos y el Apéndice C para este tipo de análisis).

Otras preguntas más allá del contenido de este segundo capítulo serían: en caso de existir una diferencia en los grupos, ¿cuál sería una posible solución a la disparidad?; ¿es apropiado sacar conclusiones con solo comparar dos grupos?; y ¿se debería de considerar algún parámetro de la población de interés para llevar a cabo este tipo de comparaciones?

Tabla Problema 2 del Capítulo 2

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

3. Se tienen 25 estudiantes de preparatoria, quienes realizaron un examen de matemáticas (la Tabla Problema 3 del Capítulo 2 muestra el pretest y el postest). Después de los resultados del pretest, la profesora decidió desarrollar una serie de videos para explicar los pasos para resolver problemas similares a los del examen. Luego de observar los videos por parte de los participantes (*i. e.*, tratamiento), se aplicó un examen equivalente para estimar el progreso de los alumnos. En otras palabras, se trató de medir el efecto estadístico y práctico que tendrían los videos en el aprendizaje de las matemáticas. Se recomienda siempre obtener algunas de las estadísticas descriptivas de una muestra como la de este problema

(véase el Apéndice A para el uso de spss en estos cálculos y el Apéndice D para este tipo de análisis).

Otras preguntas más allá del contenido de este segundo capítulo serían: ¿qué circunstancias afectan la inferencia de que funcionó o no el video?; ¿se ha estimado cómo los videos pueden tener un efecto en aprender matemáticas?; y ¿qué dicen las teorías respecto al asunto?

Tabla Problema 3 del Capítulo 2

Participante	Pretest	Postest
1	3	7
2	5	8
3	4	6
4	6	7
5	5	8
6	5	9
7	4	6
8	5	6
9	3	7
10	6	8
11	7	8
12	8	7
13	7	9
14	6	10
15	7	9
16	8	9
17	8	8
18	9	8
19	9	4
20	8	4
21	7	5
22	7	6
23	6	9
24	7	8
25	8	12

4. Una profesora tiene una clase de posgrado donde hay 5 estudiantes que trabajan y 6 que no trabajan. Desea asignarles un proyecto extra, pero quiere saber si hay una diferencia significativa y el tamaño del efecto entre los grupos, para ver si debe de utilizar dos criterios diferentes para calificarlo o hacerlo más fácil de completar para ambos grupos. Se recomienda siempre obtener algunas de las estadísticas descriptivas de una muestra como la de este problema (véase el Apéndice A para el empleo de SPSS en estos cálculos y el Apéndice C para este tipo de análisis).

Otras preguntas más allá del contenido de este segundo capítulo serían: ¿qué relación tiene el trabajo con el aprendizaje de los estudiantes universitarios?; ¿cuántas horas de trabajo a la semana son demasiadas para poder rendir en la universidad?; ¿cómo ayudar con sus estudios a las y los que trabajan?; y ¿qué se ha encontrado al respecto?

Tabla Problema 4 del Capítulo 2

Estudiantes del grupo de trabajadores	Horas disponibles para el proyecto del grupo de trabajadores	Estudiantes del grupo de no-trabajadores	Horas disponibles para el proyecto del grupo de no-trabajadores
1	7	1	12
2	8	2	14
3	10	3	18
4	4	4	13
5	6	5	11
		6	10

5. Un investigador desea saber si puede disminuir el número de ocasiones en las que sus estudiantes revisan su teléfono inteligente al día. Primero, lleva a cabo una encuesta para saber cuántas veces al día sus alumnos revisan el teléfono (Tabla Problema 5 del Capítulo 2). Luego, contrata a una profesional que les da un tratamiento para controlar sus impulsos, enfatizando el evitar revisar el teléfono a casi cada instante. Se recomienda siempre obtener algunas de las estadísticas descriptivas de una muestra como la de este pro-

blema (véase el Apéndice A para el uso de SPSS en estos cálculos y el Apéndice D para este tipo de análisis).

Otras preguntas más allá del contenido de este segundo capítulo serían: ¿hasta dónde la tecnología es un distractor para el aprendizaje?; ¿qué se puede hacer a nivel universitario para resolver este problema?; y ¿existen teorías modernas al respecto?

Tabla Problema 5 del Capítulo 2

Adolescente	Antes del tratamiento	Después del tratamiento
1	46	45
2	44	46
3	47	32
4	42	34
5	45	33
6	32	21
7	36	23
8	43	41
9	24	27
10	41	38
11	38	41
12	31	47
13	22	41
14	36	32
15	36	22
16	27	34
17	41	36
18	44	19
19	32	23
20	32	22

PREGUNTAS PARA REFLEXIONAR

1. ¿Qué información se puede obtener al comparar dos grupos?
2. ¿Cuándo la distribución t se vuelve distribución normal?
3. ¿De qué depende la forma de la distribución t ?
4. ¿Cuál es la fórmula conceptual para estimar una $t_{calculada}$?
5. ¿Qué significan los grados de libertad?
6. ¿Qué ventajas tiene la distribución t respecto a la distribución normal?
7. ¿Cuándo se puede utilizar una cola de la distribución t ?
8. ¿Qué relación tienen los grados de libertad y el $t_{critico}$?
9. ¿Es apropiado colocar un intervalo de confianza en la distribución de muestras que señale las áreas de rechazo?
10. ¿Qué pasa si se violan los supuestos del Test t ?
11. ¿Qué hacer si los resultados del Test de Significancia y el d de Cohen se contradicen?
12. ¿Cómo se podrían contradecir un Test de Significancia y un intervalo de confianza?

OPINIÓN PERSONAL

Las comparaciones de grupos nos permiten observar dos cuestiones: una estimación estadística de la diferencia; y la magnitud de la diferencia. Esta estimación estadística de la diferencia es el Test de Significancia Estadística. Con este, se puede estimar si la diferencia fue probablemente al azar o no; es decir, la diferencia se debió a un error de muestreo. La magnitud de la diferencia es qué tanta distancia existe entre los promedios medida en unidades de desviación estándar. Esta diferencia de magnitud explicada en este capítulo por el d de Cohen es también una estadística. Esto implica una diferencia entre una estadística y un parámetro. Está más allá del propósito de este capítulo calcular un intervalo de confianza para el d de Cohen, pero si se quiere más información, véase: Cohen (1988).

Puede suceder que el d de Cohen y el Test de Significancia Estadística, se contradigan. Esto puede pasar especialmente si se trabaja con una muestra pequeña. Si resulta que no hay evidencia sobre el efecto estadístico, pero se tiene aparentemente un tamaño de efecto grande. En ese caso, hay que prestar especial atención a la literatura sobre el tamaño del efecto, porque esta puede dar información acerca del tamaño del efecto. La APA (2020) indica que una investigación no estaría completa si *no* se reporta un tamaño de efecto. De este modo, habría que considerar esta indicación para tener más oportunidades de publicación. Además, este tamaño de efecto, también conocido como efecto práctico, puede presentar la magnitud de un cambio; por ejemplo, el cambio podría ser cómo el promedio de una clase por debajo de una calificación aprobatoria antes de un tratamiento, se volvió un promedio más alto ante este estándar. Otras cosas iguales: el tratamiento afectó el pasar de una categoría a otra.

También, las comparaciones de grupos nos permiten ver si la diferencia entre un pretest y un posttest fue al azar o no, así como la magnitud de la diferencia. Aunque en la práctica es complicado llevar a cabo un estudio con un diseño experimental, cabe la posibilidad de que se pueda hacer alguno de estos diseños. Este diseño experimental podría ser uno en el que no se haya seleccionado la muestra al azar y, por lo tanto, la posibilidad de generalizar los resultados se compli-

ca. Sin embargo, se podría aplicar un Test de Significancia Estadística cuando se compara el mismo grupo en dos tiempos diferentes. Estos tiempos pueden ser: uno antes de un tratamiento y otro después de este. Los promedios de ambas mediciones se pueden comparar estadísticamente, así como la magnitud, con un d de Cohen modificado.

SOLUCIONES CLAVE A LOS PROBLEMAS DEL CAPÍTULO 1

Todos estos resultados del Capítulo 1 (desde el problema 1 hasta el 5) fueron calculados manualmente, excepto el valor de p . Por esta razón, cabe la posibilidad de que no sean idénticos a otros calculados en la misma forma por usar menos o más dígitos en los algoritmos.

1. f) $\sigma_{\bar{x}} = 0.2236$; $z_{\text{calculado}} = 1.3416$; $p = .0898$.
g) $IC_{95\%} [7.2617, 8.1382]$

Nota: Para calcular los resultados anteriores, se tomaron los primeros cuatro decimales.

2. f) $\sigma_{\bar{x}} = 0.0469$; $z_{\text{calculado}} = -4.2643$; $p < .00001$.
g) $IC_{95\%} [-4.3562, -4.1724]$

Nota: Para calcular los resultados anteriores, se tomaron los primeros cuatro decimales.

3. f) $\sigma_{\bar{x}} = 2.0412$; $z_{\text{calculado}} = 1.4697$; $p = .0708$.
g) $IC_{95\%} [148.9993, 157.0007]$

Nota: Para calcular los resultados anteriores, se tomaron los primeros cuatro decimales.

4. f) $\sigma_{\bar{x}} = 6.2017$; $z_{\text{calculado}} = 8.0623$; $p < .0001$.
g) $IC_{95\%} [537.8447, 562.1553]$

Nota: Los cálculos fueron para un Test z de dos colas con un alfa = .05. Para calcular los resultados anteriores, se tomaron los primeros cuatro decimales.

5. f) $\sigma_{\bar{x}} = 4.8989$; $z_{\text{calculado}} = -12.2476$; $p < .00001$.
g) $IC_{95\%} [-249.6018, -230.3981]$

Nota: Los cálculos fueron para un Test z de dos colas con un alfa = .05. Ya sea que los cálculos se hayan hecho en horas o en minutos, el valor $z_{\text{calculado}}$ es casi el mismo. Para los cálculos de la presente solución parcial, se usaron los minutos. Para calcular los resultados anteriores, se tomaron los primeros cuatro decimales.

SOLUCIONES CLAVE A LOS PROBLEMAS DEL CAPÍTULO 2

- 1) a. Usando SPSS, estos son los resultados de las estadísticas descriptivas cuando solo se tiene una muestra (Figura 1. Capítulo 2, Problema 1):

Figura 1. Capítulo 2, Problema 1

	N	Rango	Mínimo	Máximo	Media	Desviación estándar	Asimetría	Curtosis			
	Estad.	Estad.	Estad.	Estad.	Estad.	Error estándar	Estad.	Error estándar			
Aciertos	30	14.00	1.00	15.00	5.43	.625	3.420	.926	.427	.732	.833
N válido (por lista)	30										

Nota: Estad. = Estadístico

Donde:

- ⊗ N = Tamaño de la muestra
- ⊗ Rango = Diferencia entre el valor máximo y el mínimo

- ⊗ Mínimo = Menor valor del set
 - ⊗ Máximo = Mayor valor del set
 - ⊗ Media = Promedio
 - ⊗ Error estándar = Error estándar del promedio = $SE_{\bar{x}}$
 - ⊗ Asimetría = Rango de aceptabilidad, que va de $| 2 |$ para muestras pequeñas a $| 3 |$ para muestras grandes
 - ⊗ Error estándar bajo asimetría = Error estándar de la asimetría
 - ⊗ Curtosis = Tiene los mismos rangos que la asimetría: error estándar de la curtosis
- e. Utilizando SPSS, estos son los resultados del Test t con una sola muestra (Figura 2. Capítulo 2, Problema 1):

Figura 2. Capítulo 2, Problema 1

Prueba de muestra única						
Valor de prueba = 7						
	<i>t</i>	<i>gl</i>	Sig. (bilateral)	Diferencia de medias	95% de intervalo de confianza	
					Inferior	Superior
Aciertos	-2.508	29	.018	-1.567	-2.844	-0.289

Estadísticas de muestra única				
	<i>N</i>	Media	Desviación estándar	Media de error estándar
Aciertos	30	5.433	3.421	0.625

Donde:

- ⊗ $t = t_{calculada}$
- ⊗ $gl = df =$ Grados de libertad
- ⊗ Sig. (bilateral) = Probabilidad_{calculada} = $p =$ Valor p
- ⊗ Diferencia de medias = Diferencia entre los promedios
- ⊗ 95% de intervalo de confianza = IC del 95% de confianza de la diferencia entre los promedios
- ⊗ Inferior = Límite bajo
- ⊗ Superior = Límite alto

- f. *Nota:* Esta tabla de resultados muestra lo siguiente: $IC_{95\%}$ [- 2.8440, - 0.2893]. Este es un IC para la diferencia entre los promedios que se están comparando (el promedio de la muestra y el parámetro de la población). Se obtiene al restar y sumar el producto del $t_{crítico} = 2.045$ ($df = 29$; $\alpha = .05 / 2$ colas) y del $SE_{\bar{x}} = 0.62456$ del promedio de la diferencia = -1.56667. En otras palabras, se calcula un IC de la misma manera en la que se ha hecho en la presente obra. La interpretación es la misma que se haría del promedio: i. e., de 100 muestras tomadas, se estima que 95 de ellas capturarían la verdadera diferencia entre la muestra y la población.

Por otro lado, lo que se ha estado calculando para este tipo de problemas es el $IC_{95\%}$ del promedio de la muestra; en este caso, 5.4333. Con un $t_{crítico}$ ($df = 29$; $\alpha = .05$ con dos colas) = 2.045, el $IC_{95\%}$ del promedio es: [4.1561, 6.7105].

- 2) a. Empleando SPSS, estos son los resultados de las estadísticas descriptivas de dos grupos independientes (Figura 1. Capítulo 2, Problema 2):

Figura 1. Capítulo 2, Problema 2

Estadísticos descriptivos											
	N	Rango	Mínimo	Máximo	Media	Desviación estándar	Asimetría	Curtosis			
	Estad.	Estad.	Estad.	Estad.	Estad.	Error estándar	Estad.	Estad.	Error estándar	Estad.	Error estándar
Grupo A	30	14.00	1.00	15.00	5.433	0.625	3.421	.926	.427	.732	.833
Grupo B	30	7.00	2.00	9.00	5.533	0.377	2.063	-.101	.427	-.908	.833
N válido (por lista)	30										

Nota: Estad. = Estadístico

- e. Usando spss, estos son los resultados del Test t con dos muestras independientes (Figura 2. Capítulo 2, Problema 2):

Figura 2. Capítulo 2, Problema 2

Estadísticas de grupo										
		N	Media	Desviación estándar	Error estándar de la media					
Aciertos	1.00	30	5.433	3.421	0.625					
	2.00	30	5.533	2.063	0.377					

Prueba de muestras independientes										
		Prueba de Levene de igualdad de varianzas			Prueba t para igualdad de medias					
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar	95% de intervalo de confianza de la diferencia	
								Inferior		Superior
Aciertos	Se asumen varianzas iguales	4.994	.029	-.137	58	.891	-.100	.729	-1.560	1.360
	No se asumen varianzas iguales			-.137	47.635	.892	-.100	.729	-1.567	1.367

Donde:

- ⊗ Prueba de Levene de igualdad de varianzas = Test de Levene (para ver si existe homogeneidad entre las varianzas de los dos grupos)
- ⊗ F = Valor F calculado y su correspondiente Sig. = p
- ⊗ $t = t_{calculada}$
- ⊗ $gl = df$ = Grados de libertad
- ⊗ Sig. (bilateral) = Probabilidad $_{calculada} = p$; bilateral = Usando las dos colas de la distribución
- ⊗ Diferencia de medias = Diferencia entre los promedios

- ⊗ Diferencia de error estándar = Error estándar de la diferencia = $SE_{\bar{a}}$
- ⊗ 95% de intervalo de confianza de la diferencia = IC del 95% de confianza de la diferencia entre los promedios
- ⊗ Inferior = Límite bajo
- ⊗ Superior = Límite alto

Nota: El Test de Levene no se pasó en este caso, porque la p (Sig.) = .029, que no es mayor a un alfa de .05; sin embargo, esta violación al Principio de Homogeneidad de la Varianza no es seria, porque las muestras son del mismo tamaño: 30 participantes en el Grupo A y 30 en el B. Además, SPSS ofrece la posibilidad de una corrección en caso de que las muestras no sean del mismo tamaño en el renglón: no se asumen varianzas iguales (*Equal variances not assumed*). En este caso, esta corrección no hace que la diferencia entre ambos grupos sea estadísticamente significativa, porque p [Sig. (bilateral)] = .892; es decir, los test sin corrección y con corrección arrojan una $p > .05$ y, por lo tanto, se falla en rechazar la hipótesis nula.

- f. En el caso del Capítulo 2, se habían estado calculando IC de la diferencia, así que el $IC_{95\%} = [-1.56001, 1.36001]$.
- g. Dada una $SD^* = 2.82 = [\sqrt{(SD_1^2 + SD_2^2) / 2}]$; d de Cohen = 0.035 (véase: Apéndice E, para más fórmulas del d de Cohen y otros efectos).

Desarrollo de la fórmula $\sqrt{(SD_1^2 + SD_2^2) / 2}$

$$\begin{aligned}
 &= (3.421^2 + 2.063^2) \\
 &= (11.70 + 4.26) \\
 &= 15.96;
 \end{aligned}$$

entre dos:

$$= 15.96 / 2 = 7.98;$$

raíz cuadrada:

$$\sqrt{7.98} = 2.82;$$

entonces, SD^*

$$= 2.82.$$

El d de Cohen = (Promedio del grupo 1 - Promedio del grupo 2) / SD^*

$$d = (5.533 - 5.433) / 2.82 = 0.035.$$

- 3) a. Utilizando SPSS, estos son los resultados de las estadísticas descriptivas de muestras dependientes (Figura 1. Capítulo 2, Problema 3):

Figura 1. Capítulo 2, Problema 3

Estadísticos descriptivos										
	N	Rango	Mínimo	Máximo	Media	Desviación estándar	Asimetría		Curtosis	
	Estad.	Estad.	Estad.	Estad.	Estad.	Estad.	Estad.	Error estándar	Estad.	Error estándar
Pretest	25	6.00	3.00	9.00	6.320	1.725	-.383	.464	-.652	.902
Posttest	25	8.00	4.00	12.00	7.520	1.829	.032	.464	.579	.902
N válido (por lista)	25									

Estadísticas de muestras emparejadas					
		Media	N	Desviación estándar	Error estándar de la media
Par 1	Pretest	6.3200	25	1.725	.345
	Posttest	7.5200	25	1.829	.366

Nota: Estad. = Estadístico

Donde:

- ⊗ N = Tamaño de la muestra
- ⊗ Error estándar de la media = Error estándar del promedio = SE_x

Correlación por defecto (Figura 2. Capítulo 2, Problema 3):

Figura 2. Capítulo 2, Problema 3

Correlación de muestras emparejadas				
		N	Correlación	Sig.
Par 1	Pretest y postest	25	.051	.810

Donde:

- ⊗ N = Tamaño de la muestra
- ⊗ Correlación = Correlación de producto-Momento de Pearson = r
- ⊗ Sig. = p

Nota: El Test de dos muestras dependientes en SPSS arroja una correlación por defecto. En este caso, es una correlación pequeña ($r = .051$), de acuerdo con los criterios de Cohen (1988): $r = .10$ (efecto pequeño); $r = .30$ (efecto mediano); y $r = .50$ (efecto grande). Además, la relación es no-estadística significativa ($p = .810 > \alpha = .05$), que significa que, posiblemente, es diferente de cero por mera casualidad. Para más información acerca de correlaciones de producto-Momento de Pearson, se recomienda ver a Ponce (2020).

- e. Test de Muestras Dependientes (Figura 3. Capítulo 2, Problema 3):

Figura 3. Capítulo 2, Problema 3

Prueba de muestras emparejadas									
		Diferencias emparejadas					t	gl	Sig. (bilateral)
Par		Media	Desviación estándar	Error estándar de la media	95% de intervalo de confianza de la diferencia				
					Inferior	Superior			
Par 1	Pretest y postest	-1.200	2.449	0.490	-2.211	-0.189	-2.449	24	.022

Donde:

- ⊗ Media = Diferencia entre los promedios
- ⊗ Desviación estándar = Desviación estándar de la diferencia
= SD_d
- ⊗ Error estándar de la media = Error estándar de la diferencia
= $SE_{\bar{d}}$
- ⊗ $t = t_{calculada}$
- ⊗ $gl = df =$ Grados de libertad
- ⊗ Sig. (bilateral) = Probabilidad $_{calculada} = p$

f. [- 2.21110, - 0.18890].

g. El tamaño del efecto es el d de Cohen modificado: $d^* = \frac{\bar{d}}{SD_d}$,
esto es: $0.490 / 2.449 = 0.20$.

4. a. Empleando SPSS, estas son las estadísticas descriptivas de
muestras independientes de diferente tamaño:

Figura 1. Capítulo 2, Problema 4

Estadísticos descriptivos											
	N	Rango	Mínimo	Máximo	Media	Desviación estándar	Asimetría	Curtosis			
	Estad.	Estad.	Estad.	Estad.	Estad.	Error estándar	Estad.	Estad.	Error estándar	Estad.	Error estándar
Trabajo	5	6.00	4.00	10.00	7.00	1.000	2.236	.000	.913	.200	2.000
Sin trabajo	6	8.00	10.00	18.00	13.00	1.155	2.828	1.193	.845	1.669	1.741
N válido (por lista)	5										

Estadísticas de grupo					
Grupo	N	Media	Desviación estándar	Error estándar de la media	
Horas	1.00	5	7.000	2.236	1.000
	2.00	6	13.000	2.828	1.155

Nota: Estad. = Estadístico

Donde:

- ⊗ Error estándar de la media = Error estándar del promedio = $SE_{\bar{x}}$
- e. Test t de Muestras Independientes (Figura 2. Capítulo 2, Problema 4):

Figura 2. Capítulo 2, Problema 4

Prueba de muestras independientes										
		Prueba de Levene de igualdad de varianzas		Prueba t para igualdad de medias						
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar	95% de intervalo de confianza de la diferencia	
								Inferior	Superior	
Horas	Se asumen varianzas iguales	.169	.690	-3.838	9	.004	-6.0	1.563	-9.537	-2.463
	No se asumen varianzas iguales			-3.928	8.991	.003	-6.0	1.528	-9.456	-2.544

Nota: El Test de Levene se *pasó* en este caso, porque la p (Sig.) = .69, que es mayor a un alfa de .05. Por lo tanto, no existe violación al Principio de Homogeneidad, a pesar de que son grupos de diferente tamaño: 5 de trabajadores y 6 de no trabajadores. En caso de no cumplir con el Supuesto

de Homogeneidad y tener grupos asimétricos, spss ofrece la posibilidad de una corrección en el renglón: se asumen varianzas iguales.

- f. En el caso del Capítulo 2, se habían estado calculando IC de la diferencia, así que el $IC_{95\%} = [-9.53682, -2.46318]$.
- g. Debido a que el tamaño de las muestras es diferente, al igual que las desviaciones estándar, se usa la siguiente fórmula del Apéndice E:

$$d_p = \frac{(\bar{x}_1 - \bar{x}_2)}{SD_p}$$

Donde:

- ⊗ d_p = Tamaño del efecto: d de Cohen entre grupos independientes con diferentes tamaños de muestras, así como de desviaciones estándar
- ⊗ \bar{x}_1 = Promedio calculado de los datos del grupo 1
- ⊗ \bar{x}_2 = Promedio calculado de los datos del grupo 2
- ⊗ SD_p = Desviación estándar compuesta de los datos del grupo 1 y grupo 2

$$SD_p = \sqrt{\frac{(n_1 - 1) SD_1^2 + (n_2 - 1) SD_2^2}{n_1 + n_2 - 2}}$$

Donde:

- ⊗ n_1 = Tamaño de la muestra 1
- ⊗ n_2 = Tamaño de la muestra 2
- ⊗ SD_1 = Desviación estándar del grupo 1
- ⊗ SD_2 = Desviación estándar del grupo 2

Sustituyendo la desviación estándar compuesta de los datos del grupo 1 y grupo 2 (los datos vienen de la Figura 1. Capítulo 2, Problema 4):

$$SD_p = \sqrt{\frac{(5 - 1)2.236^2 + (6 - 1)2.828^2}{5 + 6 - 2}}$$

$$SD_p = 2.582.$$

Sustituyendo en la fórmula del d_p de Cohen para muestras y desviaciones estándar diferentes (parte de los datos vienen de la Figura 1. Capítulo 2, Problema 4):

$$d_p = \frac{(13 - 7)}{2.582} = 2.32.$$

5. a. Usando SPSS, estos son los resultados de las estadísticas descriptivas de muestras dependientes:

Figura 1. Capítulo 2, Problema 5

Estadísticos descriptivos											
	N	Rango	Mínimo	Máximo	Media	Desviación estándar	Asimetría	Curtosis			
	Estad.	Estad.	Estad.	Estad.	Estad.	Error estándar	Estad.	Estad.	Error estándar	Estad.	Error estándar
Antes del tratamiento	20	25	22	47	36.950	1.658	7.416	-.518	.512	-.685	.992
Después del tratamiento	20	28	19	47	32.850	2.024	9.051	-.023	.512	-1.278	.992
N válido (por lista)	20										

Estadísticas de muestras emparejadas					
		Media	N	Desviación estándar	Error estándar de la media
Par 1	Antes del tratamiento	36.950	20	7.416	1.658
	Después del tratamiento	32.850	20	9.051	2.024

Nota: Estad. = Estadístico

Correlación por defecto (Figura 2. Capítulo 2, Problema 5):

Figura 2. Capítulo 2, Problema 5

Correlación de muestras emparejadas				
		N	Correlación	Sig.
Par 1	Antes del tratamiento y después del tratamiento	20	.184	.437

Nota: El Test de dos muestras dependientes en SPSS arroja una correlación por defecto. En este caso, es una correlación entre pequeña y mediana ($r = .184$), de acuerdo con los criterios de Cohen (1988): $.10 =$ correlación pequeña y $.30 =$ correlación mediana. Además, la relación es no-estadística significativa: posiblemente es diferente de cero por mera casualidad. Para más información acerca de correlaciones de producto-Momento de Pearson, se recomienda ver a Ponce (2020).

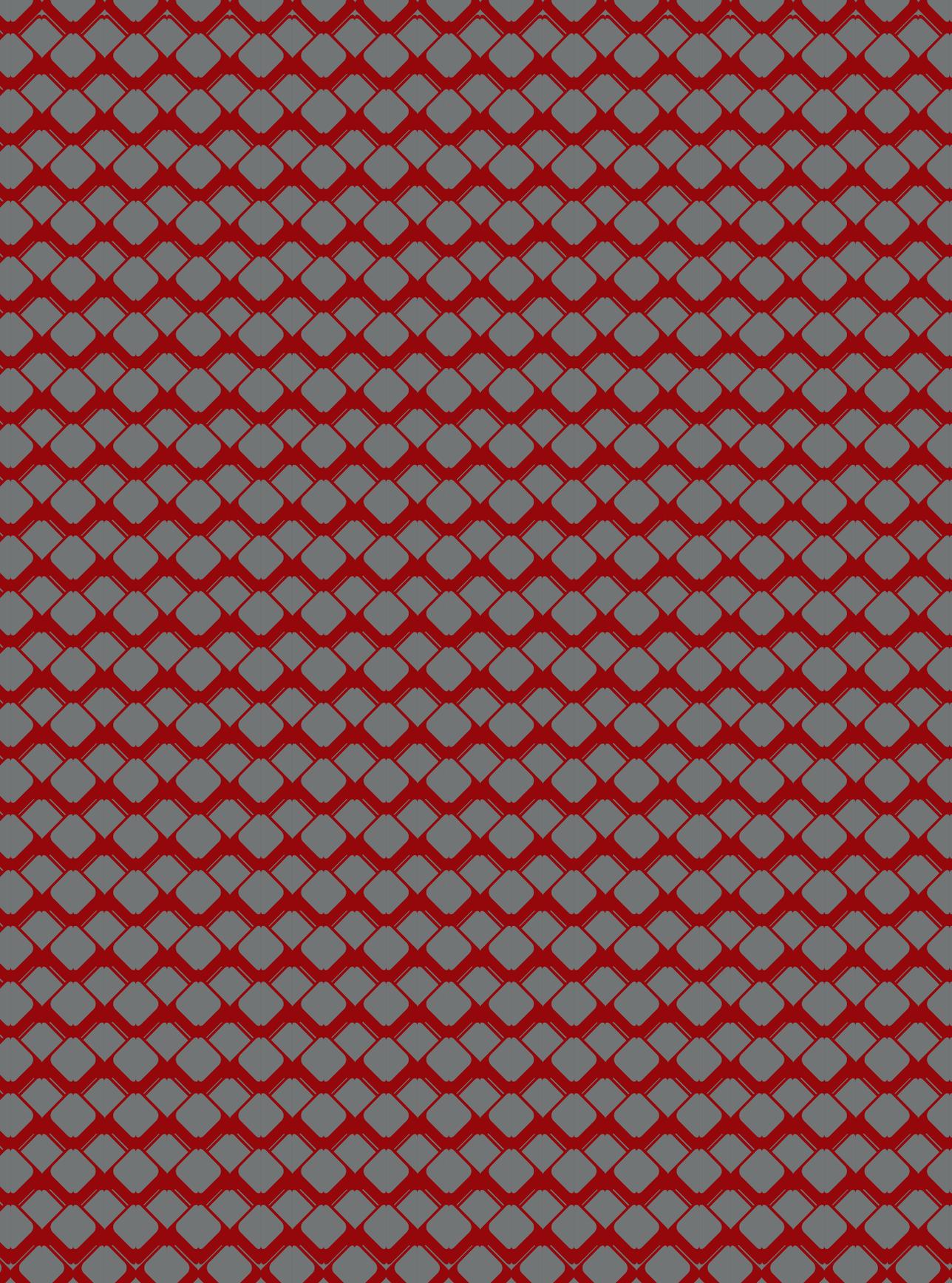
- e. Test t de Muestras Dependientes (Figura 3. Capítulo 2, Problema 5):

Figura 3. Capítulo 2, Problema 5

Prueba de muestras emparejadas									
		Diferencias emparejadas					t	gl	Sig. (bilateral)
Par		Media	Desviación estándar	Error estándar de la media	95% de intervalo de confianza de la diferencia				
					Inferior	Superior			
Par 1	Antes del tratamiento y después del tratamiento	4.100	10.592	2.369	-0.857	9.057	1.731	19	.100

Donde:

- ⊗ Media = Diferencia entre los promedios
 - ⊗ Desviación estándar = Desviación estándar de la diferencia = SD_d
 - ⊗ Error estándar de la media = Error estándar de la diferencia = $SE_{\bar{d}}$
- f. [- 0.857, 9.057].
- g. El tamaño del efecto es el d de Cohen modificado: $d^* = \frac{\bar{d}}{SD_d}$; esto es: $2.369 / 10.592 = 0.22$.

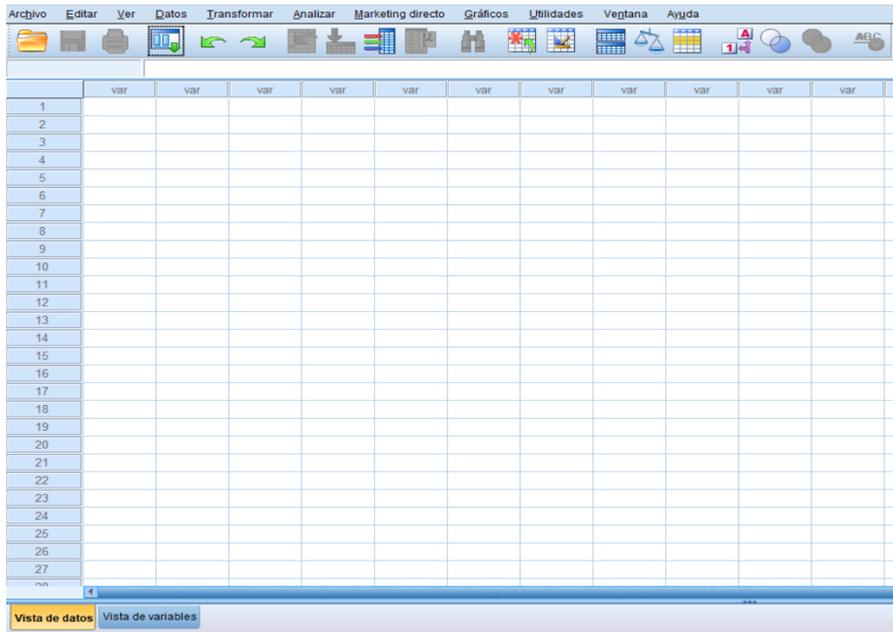


APÉNDICE A

GENERALIDADES DEL USO DE SPSS PARA LAS COMPARACIONES DE GRUPOS

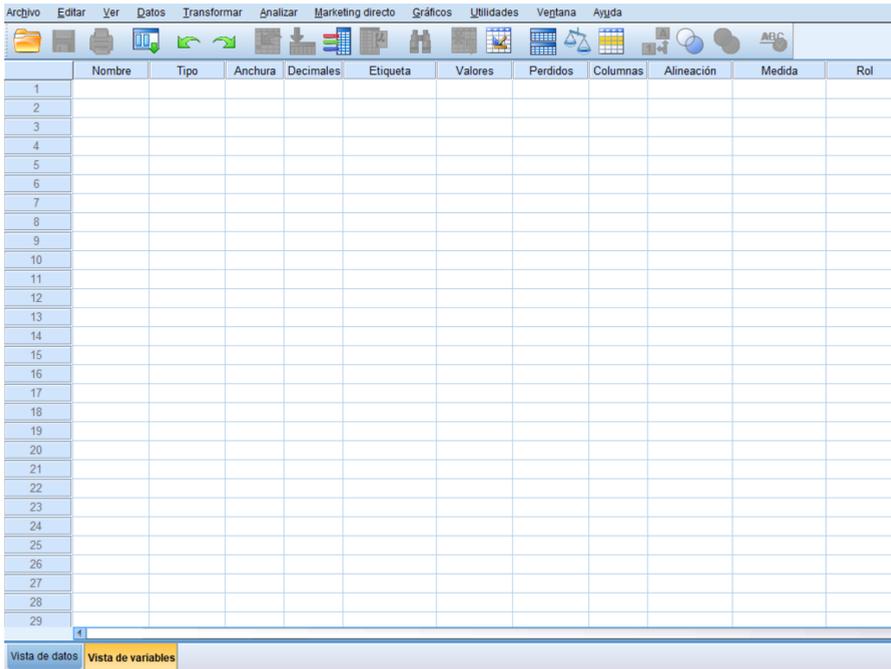
El *software spss* tiene dos vistas: *Vista de los datos* y *Vista de las variables* (Figura A-1). Asimismo, se produce una tercera hoja con resultados cuando se hace algún cálculo (véase: Apéndice B, para una hoja de resultados: *output*). Si se tienen los datos en forma numérica en Excel, se puede realizar un simple copiar y pegar (*copy-paste*) en *Vista de los datos*. Esto es posible porque este *software* reconocerá que se trata de números. Por otro lado, se puede especificar en *Vista de las variables*, si se trabaja con variables numéricas o no-numéricas, como nombres de personas, lugares u objetos.

Figura A-1. Vista de los datos y de las variables en spss



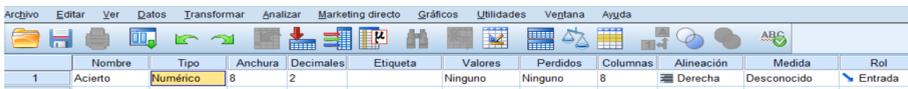
Al oprimir *Ver*, se va a esta vista que ofrece una serie de opciones para definir las variables (Figura A-2). La primera es el nombre: no puede empezar con un número, solo acepta nombres sin espacios (e. g., Juan_López: si se va a emplear un guion tiene que ser uno bajo), y la recomendación es utilizar algo breve y descriptivo de la variable (e. g., ítem_1; variable_dependiente; calificación_1; etcétera).

Figura A-2. Vista de la variable en spss



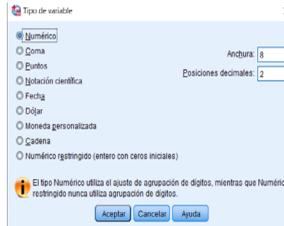
Al introducir un nombre a la variable, se activan las demás opciones para definir esta (Figura A-3).

Figura A-3. Opciones en la vista de la variable



La primera es el tipo de variable (Tipo). Aquí se tienen las opciones de Numérica (*Numérico*)... y Cadena, entre otros tipos. Estos dos tipos son los más usados y el Numérico, como su nombre lo indica, es para los números. Esto es muy importante si se define como número cuando se van a hacer test con la distribución t para las variables dependientes, porque de otra forma no se podrían realizar los análisis correspondientes. El tipo de Cadena es para los nombres de personas,

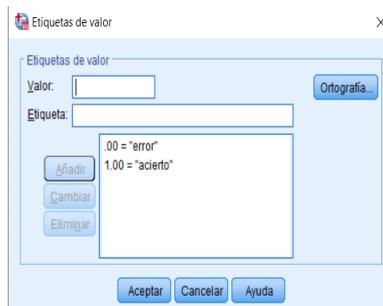
lugares u objetos. También, se puede especificar el número de dígitos que se necesitan del punto hacia la izquierda (*anchura*) y el número de dígitos para los decimales (*Posiciones de decimales*).



Otro nombre se puede colocar en la etiqueta (*Etiqueta*). Aquí sí se puede poner una pregunta completa o números sin necesidad de usar guiones bajos.



En donde dice *Valores*, se coloca en el campo de Valor (e. g., 0), y en *Etiqueta* lo que significa (e. g., error). Se oprime *Añadir* para agregar el valor con su correspondiente etiqueta. También, se podría complementar lo anterior cuando se trata de un ítem que tiene dos opciones: 1 = Acierto y 0 = Error.



Otra de las opciones importantes para los ejemplos que se muestran en el presente libro es la medición (*Medida*), donde hay tres opciones: *Escala*, *Ordinal* y *Nominal*. Si se seleccionó Numérica correspondería *Escala* y si se optó por Cadena sería *Nominal*.



Ejemplo 1 del Capítulo 2

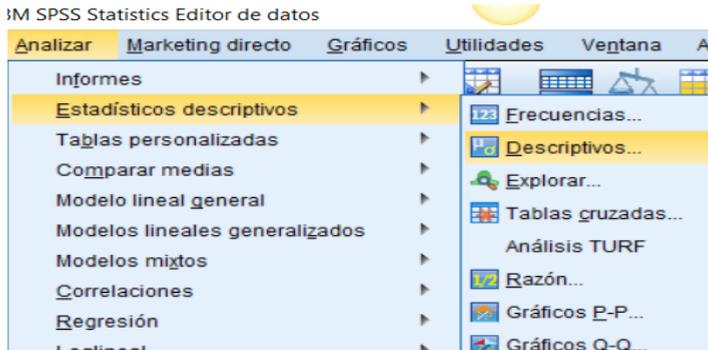
Inserción de algunos de los datos (Figura A-4).

Figura A-4. Algunos de los datos de la tabla del Problema 1 del Capítulo 2

	Aciertos
1	7.00
2	3.00
3	3.00
4	2.00
5	3.00
6	8.00
7	8.00
8	5.00
9	8.00
10	5.00
11	5.00
12	4.00
13	6.00
14	10.00
15	10.00
16	5.00
17	1.00
18	1.00
19	4.00
20	3.00
21	5.00
22	7.00
23	1.00
24	9.00
25	2.00
26	5.00
27	2.00

Para obtener algunas de las estadísticas descriptivas, se presiona: *Analizar > Estadísticos descriptivos > Descriptivos* (Figura A-5).

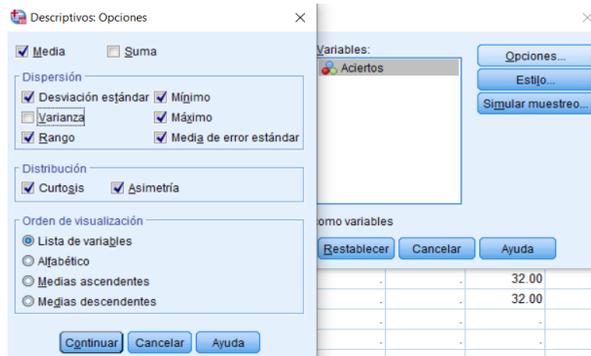
Figura A-5. Cálculo de estadísticas



Enseguida, se abre una ventana donde se pasa la variable de interés del lado izquierdo hacia el derecho.



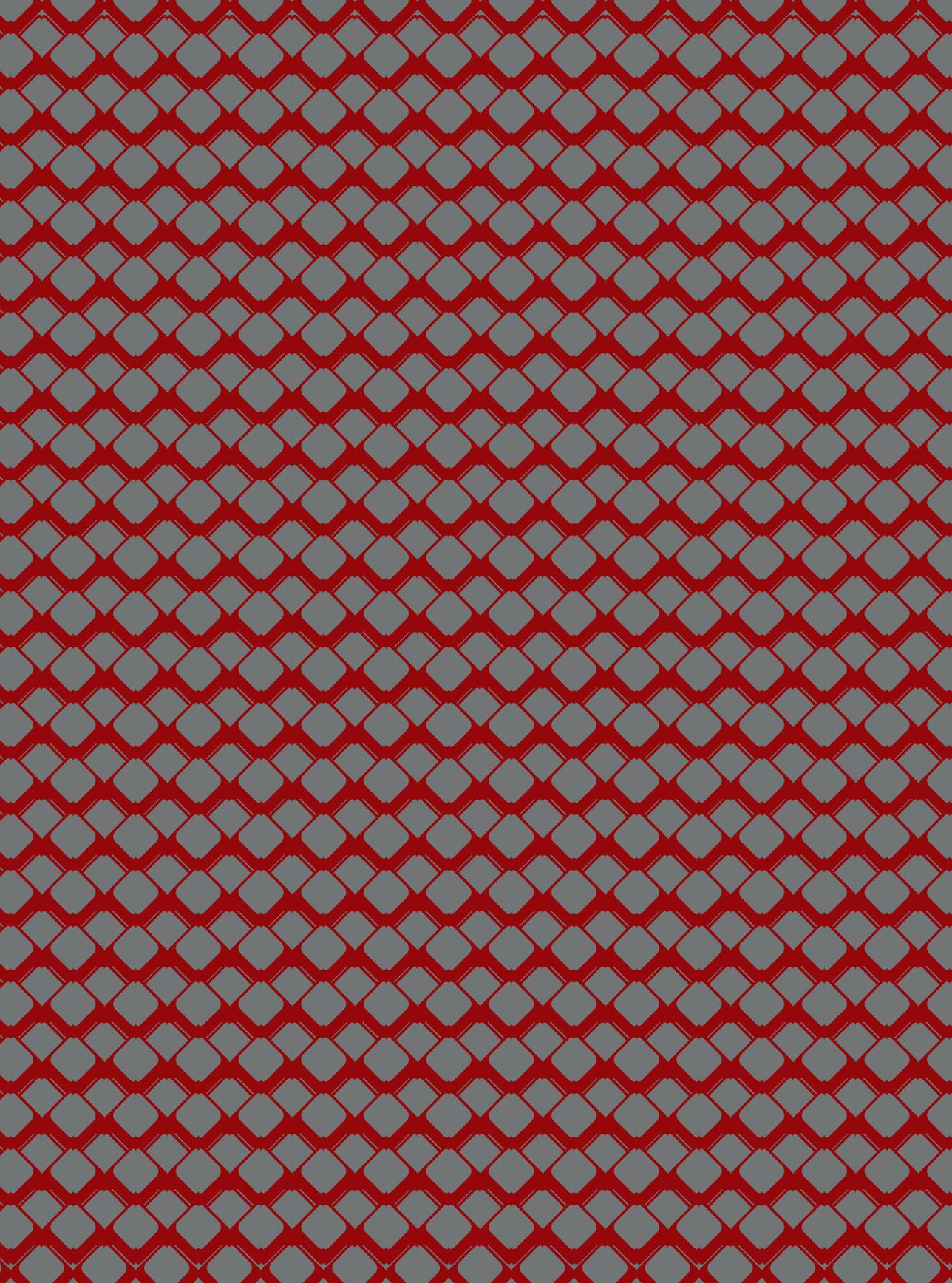
Luego, se presiona el botón de *Opciones* y se seleccionan las estadísticas señaladas en la ventana.



Por último, se presionan los botones de *Continuar* (primero) y *Aceptar* (segundo), y aparecerá la hoja con los resultados de las estadísticas seleccionadas.

Figura A-6. Tabla de resultados

Estadísticos descriptivos											
	N	Rango	Mínimo	Máximo	Media	Desviación estándar	Asimetría	Curtosis			
	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico	Error estándar	Estadístico	Estadístico	Error estándar	Estadístico	Error estándar
Aciertos	30	14.00	1.00	15.00	5.43	.625	3.420	.926	.427	.732	.833
N válido (por lista)	30										

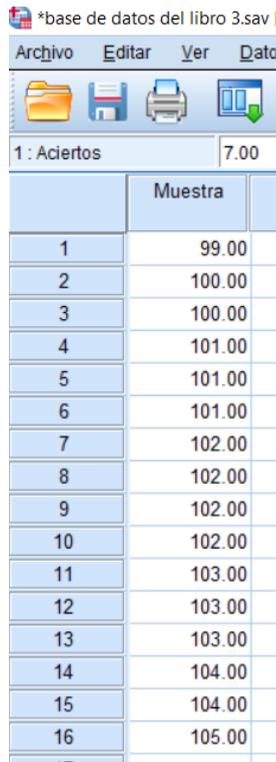


APÉNDICE B

COMPARACIONES CON UNA SOLA MUESTRA VS. UNA POBLACIÓN

Se pueden colocar los números directamente y SPSS reconocerá que se trata de números, y, por defecto, los reconocerá como una variable numérica. Otra opción sería utilizar el Apéndice A de la presente obra para definir la variable.

Figura B-1. Una sola muestra para el Test t



*base de datos del libro 3.sav |

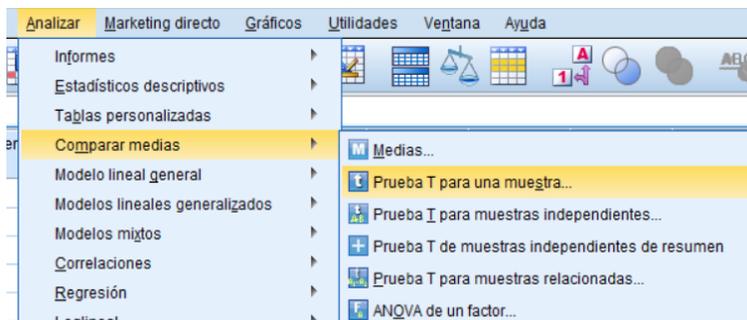
Archivo Editar Ver Dato

1 : Aciertos 7.00

	Muestra
1	99.00
2	100.00
3	100.00
4	101.00
5	101.00
6	101.00
7	102.00
8	102.00
9	102.00
10	102.00
11	103.00
12	103.00
13	103.00
14	104.00
15	104.00
16	105.00

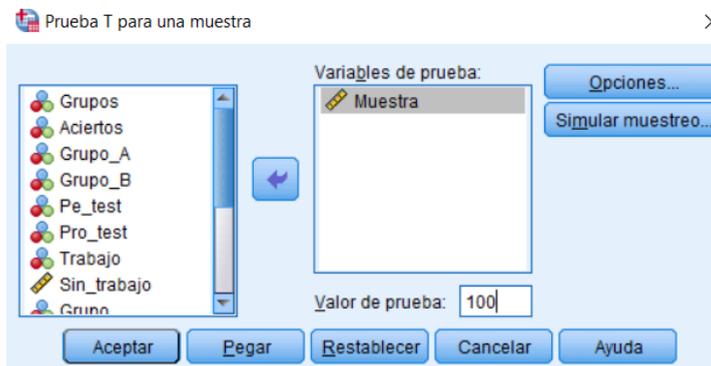
Se presiona en donde dice *Analizar > Comparar medias > Prueba T para una muestra*. Lo anterior es para comenzar el análisis (Figura B-2).

Figura B-2. Análisis de los datos de una sola muestra



Entonces, aparecerá una ventana donde se cambiará a la variable *Muestra* con un clic doble de izquierda a derecha y se especificará que el promedio de la población es 100 en el campo llamado *Valor de Prueba* (Figura B-3). Es fundamental dar el promedio de la población, porque si no se hace o se deja el cero que aparece por defecto, no se podrá efectuar el Test de una sola muestra o no se calculará apropiadamente el $t_{calculado}$. Por último, se presiona *OK*.

Figura B-3. Entrando la variable y promedio de la población



Después de presionar *Aceptar*, aparece la hoja de resultados (Figura B-4). En el primer cuadro de la hoja de resultados, se podrá observar el tamaño de la muestra ($n = 16$), el promedio (*Media*) = 102, la desviación estándar (*SD*) = 1.63 y el error estándar (*Media de error estándar*) = .408. En el segundo cuadro, aparecerá el $t_{calculado} = 4.899$, los grados de libertad (*gl*) = 15, la diferencia entre el promedio de la muestra y el de la población ($102 - 100 = 2$) y el IC del 95% de confianza [1.1298, 2.8702]. Este último indicaría que de 100 muestras que se tomaran, 95 de ellas capturarían esta diferencia (véase a Cumming, 2013, para más detalles). Aunque esta hoja de resultados no ofrece una $t_{critica}$, sí da una p (Sig. bilateral) que, en este caso, es menor a .001 (i. e., $p < .001$). Con esta p , se puede comparar con el alfa seleccionada que, en este caso, fue $\alpha = .05$. Por lo tanto, $p < \alpha$, así que existe una diferencia estadísticamente significativa y se rechaza la hipótesis nula.

Figura B-4. Hoja de resultados

Estadísticas de muestra única				
	N	Media	Desviación estándar	Media de error estándar
Muestra	16	102.00	1.633	0.408

Prueba de muestra única						
Valor de prueba = 100						
	t	gl	Sig. (bilateral)	Diferencia de medias	95% de intervalo de confianza de la diferencia	
					Inferior	Superior
Muestra	4.899	15	.000	2.00	1.130	2.870

APÉNDICE C

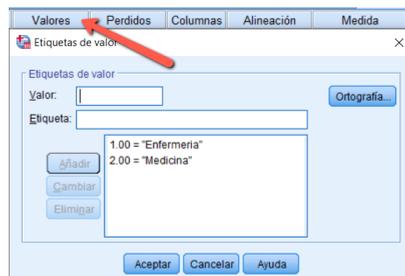
TEST T DE DOS MUESTRAS INDEPENDIENTES

Para comparar dos grupos independientes, hay que crear dos variables numéricas (véase: Apéndice A). En una de las columnas van los datos de la variable dependiente (Figura C-1). En este caso, son las calificaciones de ambos grupos. En la otra columna, se definen los grupos.

Figura C-1. Test t para grupos independientes

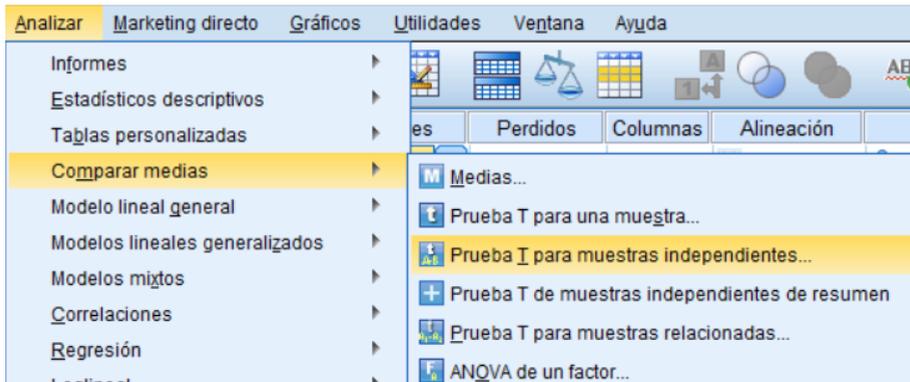
	Carrera	Variable dependiente
1	1.00	9.00
2	1.00	9.10
3	1.00	9.20
4	1.00	9.30
5	1.00	9.40
6	1.00	9.50
7	1.00	9.60
8	1.00	9.70
9	1.00	9.80
10	1.00	9.90
11	1.00	10.00
12	2.00	8.00
13	2.00	8.10
14	2.00	8.20
15	2.00	8.30
16	2.00	8.40
17	2.00	8.50
18	2.00	8.60
19	2.00	8.70
20	2.00	8.80
21	2.00	8.90
22	2.00	9.00

Con más precisión, la Vista de las variables ofrece la opción de *Valores*, donde se hace clic y se definen los diferentes grupos. En este caso, se colocó el número 1 en *Valor* y *Enfermería*, y se presionó el botón de *Añadir*. Lo mismo se hizo para el otro grupo: 2 > *Medicina* > *Añadir* > *Aceptar*.



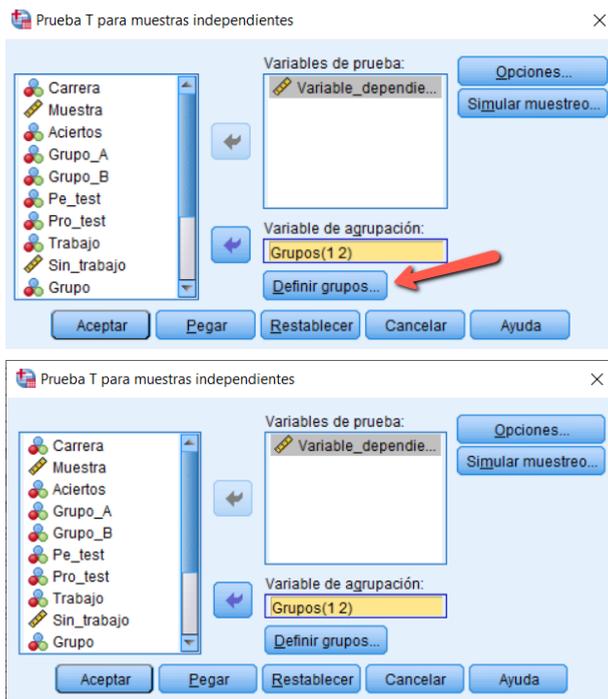
Una vez definidas las variables, se lleva a cabo el análisis: *Anali-zar > Compare medias > Prueba T para muestras independientes*.

Figura C-2. Análisis del Test t para grupos independientes

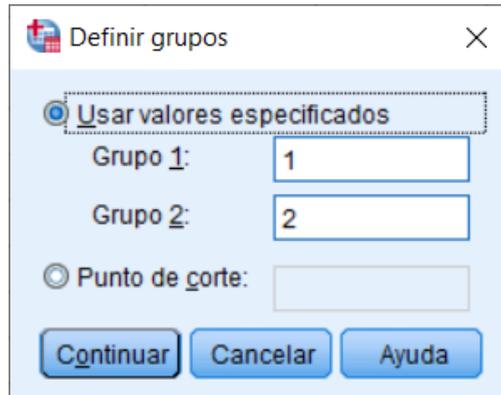


Ahora habría que colocar la Variable dependiente en donde dice: *Test variable(s)*, y Carrera (variable independiente) en *Definir grupos*.

Figura C-3. Detalles del análisis del Test t para grupos independientes



Después de lo anterior, se presiona donde dice: *Definir grupos...* Aquí surgirá otra ventana donde se definirán los dos grupos, de acuerdo con la información que se había llenado previamente en la Figura C-1 en la parte de *Valores; recordando: 1 = Enfermería y 2 = Medicina.* En esta ventana de *Definir grupos* solo es necesario colocar los números como se ven.

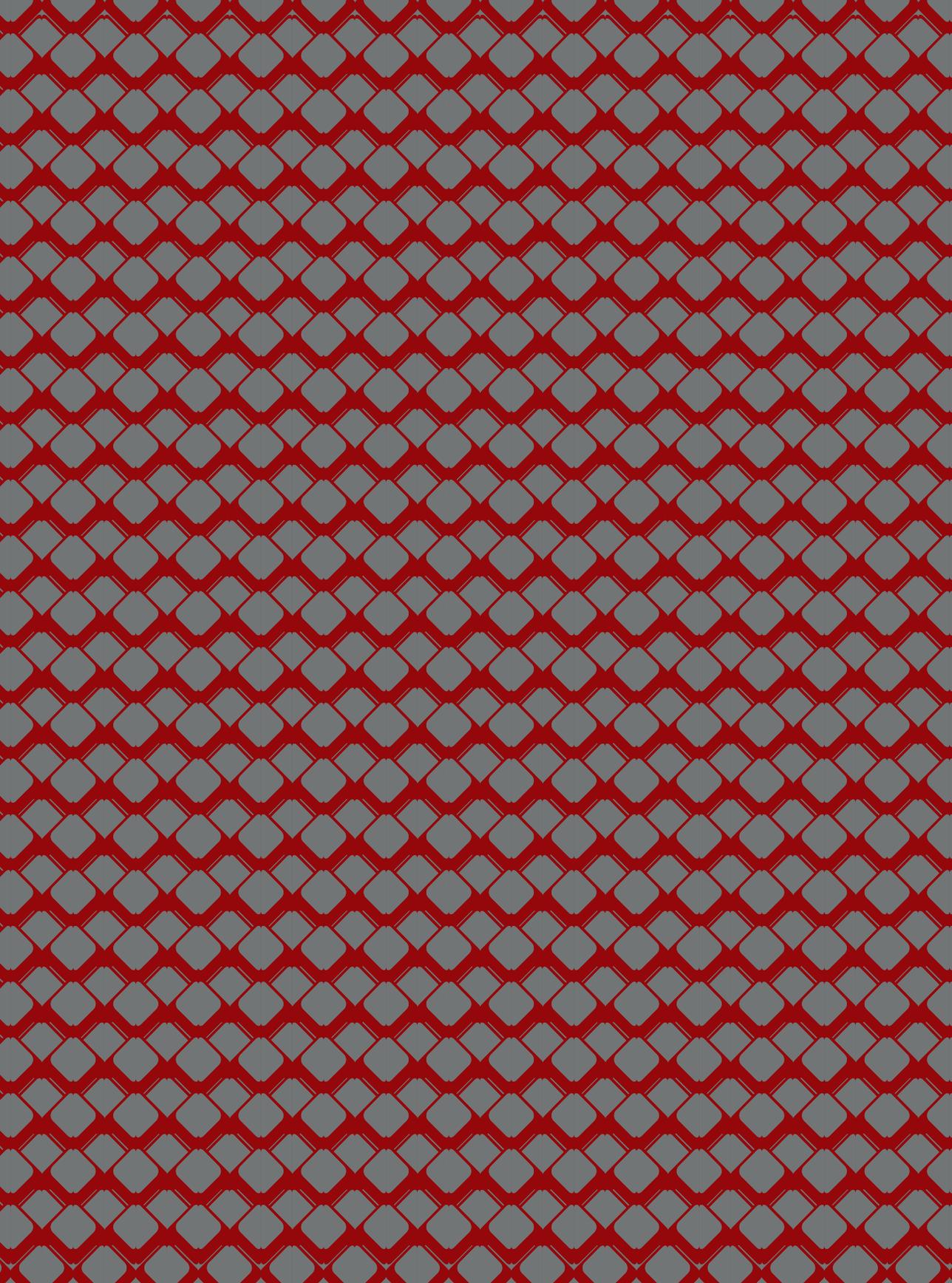


Después de presionar *Continuar* y, luego, *Aceptar*, aparece la hoja de resultados donde la primera de las tablas indica el tamaño de cada grupo (N), promedio (*Media*), desviación estándar (SD) y error estándar (*error estándar de la media*; Figura C-4). La segunda tabla indica, primero, el Test de Homogeneidad de Levene (*Prueba de Levene de igualdad de varianzas*), donde no se rechaza o se falla en rechazar la hipótesis nula (véase el Capítulo 2 para más detalles). Esta segunda tabla también provee un $t_{\text{calculado}}$ en caso de que no se cumpla con el supuesto de homogeneidad de las varianzas. En este caso, las estadísticas son idénticas en cuanto a df , p (Sig. [bilateral]), diferencia entre los promedios (*Diferencia de medias*), error estándar de la diferencia (*Diferencia de error estándar*) y el intervalo de 95% de confianza (*95% de intervalo de confianza de la diferencia*).

Figura C-4. Hoja de resultados del Test t para muestras independientes

Prueba de muestras independientes										
		Prueba de Levene de igualdad de varianzas		Prueba t para igualdad de medias						
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar	95% de intervalo de confianza de la diferencia	
								Inferior	Superior	
Variable dependiente	Se asumen varianzas iguales	0.00	1.000	7.071	20	.000	1.00	.141	.705	1.295
	No se asumen varianzas iguales			7.071	20	.000	1.00	.141	.705	1.295

Estadísticas de grupo					
Carrera		N	Media	Desviación estándar	Error estándar de la media
Variable dependiente	Enfermería	11	9.50	.332	.10
	Medicina	11	8.50	.332	.10

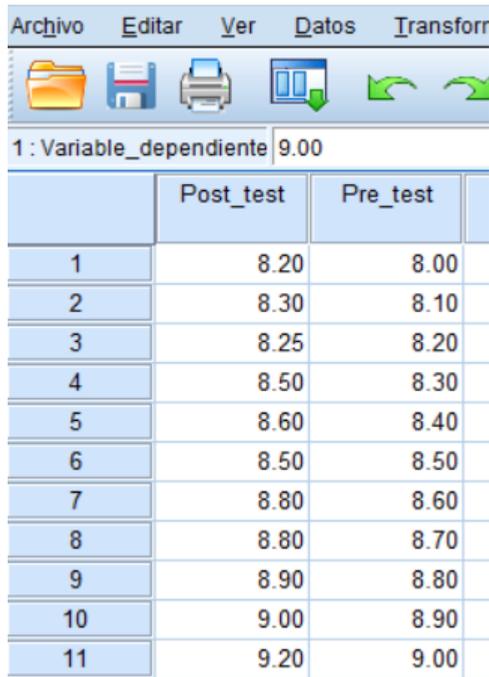


APÉNDICE D

TEST T DE DOS MUESTRAS DEPENDIENTES

Para comparar dos muestras dependientes, hay que crear dos variables numéricas (véase: Apéndice A). En una de las columnas van los datos del postest y en la otra, los del pretest: la variable dependiente son las calificaciones (Figura D-1). Por otro lado, la variable independiente son las ocasiones: la evaluación del mismo grupo en dos oportunidades. En este caso, se colocó primero el postest que el pretest, para obtener un $t_{\text{calculado}}$ positivo para una interpretación más sencilla de los resultados.

Figura D-1. Entrada de los datos del Test t para muestras dependientes

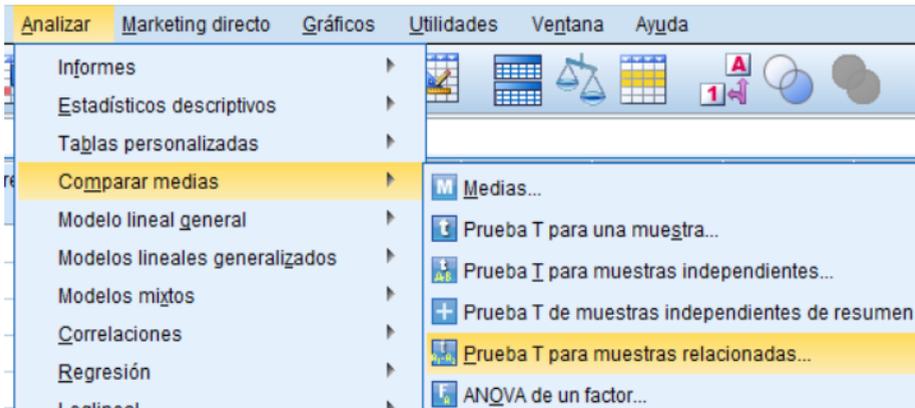


The image shows a screenshot of the SPSS data entry window. The menu bar includes 'Archivo', 'Editar', 'Ver', 'Datos', and 'Transformar'. Below the menu bar are icons for file operations (folder, save, print, refresh) and navigation (back, forward). The window title is '1 : Variable_dependiente 9.00'. The data is entered in a table with two columns: 'Post_test' and 'Pre_test'. The rows are numbered 1 through 11.

	Post_test	Pre_test
1	8.20	8.00
2	8.30	8.10
3	8.25	8.20
4	8.50	8.30
5	8.60	8.40
6	8.50	8.50
7	8.80	8.60
8	8.80	8.70
9	8.90	8.80
10	9.00	8.90
11	9.20	9.00

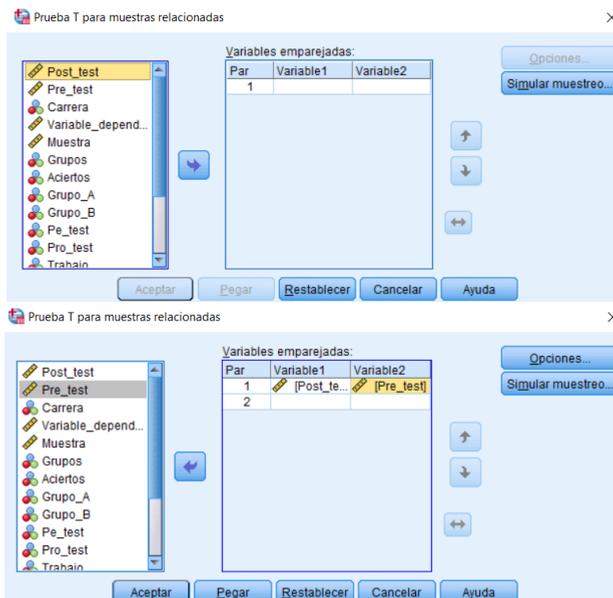
El análisis de los datos comienza con: *Analizar > Comparar medias > Prueba T para muestras relacionadas.*

Figura D-2. Análisis del Test t para muestras dependientes



Después, se pasa de las columnas del lado izquierdo hacia el lado derecho para el análisis. Aquí también se puede cambiar el orden de las columnas (Figura D-3). En este caso, se mantuvieron en la misma forma en la que fueron creadas: posttest y luego pretest. Al final, se oprime: *Aceptar*.

Figura D-3. Entrada de las columnas



La hoja de resultados muestra en la primera tabla (*Estadísticas de muestras emparejadas*) el promedio (*Media*), tamaño de la muestra (*N*), desviación estándar (*SD*) y media del error estándar (*SE*) (Figura D-4). La segunda tabla (*Correlación de muestras independientes*) expone la correlación de producto-Momento de Pearson (mejor conocida como la *r* de Pearson; véase: Ponce, 2020, para más información sobre este coeficiente). La tercera tabla (*Prueba de muestras emparejadas*) muestra las estadísticas del test y si la diferencia entre las ocasiones fue estadísticamente significativa. En este caso, $p < .001$, así que se rechaza la hipótesis nula.

Figura D-4. Hoja de resultados de los Test *t* para muestras dependientes

Estadísticas de muestras emparejadas									
		Media	N	Desviación estándar	Media del error estándar				
Par 1	Postest	8.641	11	.326	.098				
	Pretest	8.500	11	.332	.100				
Correlación de muestras emparejadas									
		N		Correlación	Sig.				
Par 1	Postest y pretest	11		.975	.000				
Prueba de muestras emparejadas									
Diferencias emparejadas									
		Media	Desviación estándar	Media del error estándar	95% de intervalo de confianza		t	gl	Sig. (bilateral)
					Inferior	Superior			
Par 1	Postest y pretest	.141	.074	.022	.091	.190	6.354	10	.000

APÉNDICE E

FÓRMULAS PARA ESTIMAR LA MAGNITUD DEL TAMAÑO DEL EFECTO

Tabla E-1. Fórmulas del d de Cohen para un solo grupo o grupos independientes

Notación y ecuación	Divisor
Diseños con un solo grupo:	
$d_1 = \frac{(\bar{x}_1 - \text{Promedio de otro grupo})}{SD_1}$	$SD_1 = \sqrt{\frac{1}{n_1-1} \sum (x_i - \bar{x}_1)^2}$
<p>Donde:</p> <ul style="list-style-type: none"> • d_1 = Tamaño del efecto: d de Cohen ante una estadística de los datos y un parámetro de una población o estadística encontrada en la literatura u otra fuente • \bar{x}_1 = Promedio calculado de los datos del grupo 1 • Promedio de otro grupo = Promedio de una población o muestra que se haya encontrado en la literatura u otra fuente • SD_1 = Desviación estándar calculada de los datos del grupo 1 	<p>Donde:</p> <ul style="list-style-type: none"> • x_i = Valor individual de cada número de los datos • n_1 = Tamaño de la muestra 1. Esto ocasiona que los grados de libertad sean: $n - 1$
<p>Supuestos: La distribución de la población es normal.</p>	

Continúa...

Notación y ecuación **Divisor**

Diseños de dos grupos independientes:

Con muestras y desviaciones estándar diferentes

$$d_p = \frac{(\bar{x}_1 - \bar{x}_2)}{SD_p}$$

Donde:

- d_p = Tamaño del efecto: d de Cohen entre grupos independientes con diferentes tamaños de muestras, así como de SD
- \bar{x}_1 = Promedio calculado de los datos del grupo 1
- \bar{x}_2 = Promedio calculado de los datos del grupo 2
- SD_p = Desviación estándar compuesta de los datos del grupo 1 y grupo 2

$$SD_p = \sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 - 2}}$$

Donde:

- n_1 = Tamaño de la muestra 1
- n_2 = Tamaño de la muestra 2
- SD_1 = Desviación estándar del grupo 1
- SD_2 = Desviación estándar del grupo 2

Con muestras y desviaciones estándar iguales

$$d_{\text{entre grupos}} = \frac{(\bar{x}_1 - \bar{x}_2)}{SD_{1 \text{ o } 2}}$$

Donde:

- $d_{\text{entre grupos}}$ = Tamaño del efecto: d de Cohen entre grupos independientes con muestras y SD iguales
- \bar{x}_1 = Promedio calculado de los datos del grupo 1
- \bar{x}_2 = Promedio calculado de los datos del grupo 2
- $SD_{1 \text{ o } 2}$ = Desviación estándar del grupo 1 o grupo 2, porque son iguales

Donde:

- SD_1 = Desviación estándar del grupo 1
- SD_2 = Desviación estándar del grupo 2

$$SD_1 = SD_2$$

Con muestras iguales, pero desviaciones estándar desiguales

$$d_{\text{con SD desiguales}} = \frac{(\bar{x}_1 - \bar{x}_2)}{SD_{\text{combinada}}}$$

Donde:

- $d_{\text{entre grupos}}$ = Tamaño del efecto: d de Cohen entre grupos independientes con muestras iguales y SD desiguales
- \bar{x}_1 = Promedio calculado de los datos del grupo 1
- \bar{x}_2 = Promedio calculado de los datos del grupo 2
- $SD_{\text{combinada}}$ = Suma de las dos desviaciones estándar al cuadrado promediadas y obtención de la raíz cuadrada

$$SD_{\text{combinada}} = \sqrt{\frac{SD_1^2 + SD_2^2}{2}}$$

Donde:

- SD_1 = Desviación estándar del grupo 1
- SD_2 = Desviación estándar del grupo 2

Supuestos: Las distribuciones de las poblaciones son normales y existe homogeneidad en las varianzas de los grupos.

Una fuente parcial: Goulet-Pelletier y Cousineau (2018).

Tabla E-2. Fórmulas del d de Cohen para medidas repetidas

Notación y ecuación	Divisor
---------------------	---------

Diseños de medidas repetidas:

d_p es igual que el anterior.

SD_p es igual que la anterior.

Con desviaciones estándar iguales

$$d_{\text{dentro de un grupo}} = \frac{(\bar{x}_{\text{pre-test}} - \bar{x}_{\text{post-test}})}{SD_{\text{pre-test}} = SD_{\text{post-test}}}$$

$$SD_{\text{pretest}} = SD_{\text{posttest}}$$

Donde:

- $d_{\text{dentro del mismo grupo}}$ = Tamaño del efecto: d de Cohen dentro de un grupo (también, es conocido como grupos dependientes) y SD iguales
- \bar{x}_{pretest} = Promedio calculado de los datos del pretest
- $\bar{x}_{\text{posttest}}$ = Promedio calculado de los datos del posttest

Donde:

- SD_{pretest} = Desviación estándar del pretest
- SD_{posttest} = Desviación estándar del posttest

Con desviaciones estándar desiguales

$$d_{\text{dentro de un grupo}} = \frac{(\bar{x}_{\text{pre-test}} - \bar{x}_{\text{post-test}})}{SD_{\text{promedio}}}$$

$$SD_{\text{promedio}} = \sqrt{\frac{SD_{\text{pre-test}}^2 + SD_{\text{post-test}}^2}{2}}$$

Donde:

- $d_{\text{dentro del mismo grupo}}$ = Tamaño del efecto: d de Cohen dentro de un grupo (también, es conocido como grupos dependientes) con muestras iguales y SD desiguales
- \bar{x}_{pretest} = Promedio calculado de los datos del pretest
- $\bar{x}_{\text{posttest}}$ = Promedio calculado de los datos del posttest
- SD_{promedio} = Suma de las dos desviaciones estándar al cuadrado promediadas y obtención de la raíz cuadrada

Donde:

- SD_1 = Desviación estándar del pretest
- SD_2 = Desviación estándar del posttest

$$d_{\text{correlación}} = \frac{(\bar{x}_{1\text{pre-test}} - \bar{x}_{\text{post-test}})}{SD_{\text{correlación}}}$$

$$SD_{\text{correlación}} =$$

$$\sqrt{SD_{\text{pre-test}}^2 + SD_{\text{post-test}}^2 - 2rSD_{\text{pre-test}}SD_{\text{post-test}}}$$

Donde:

- $d_{\text{correlación}}$ = Tamaño del efecto: d de Cohen dentro de un grupo (también, es conocido como grupos dependientes) con muestras iguales y SD desiguales
- \bar{x}_{pretest} = Promedio calculado de los datos del pretest
- $\bar{x}_{\text{posttest}}$ = Promedio calculado de los datos del posttest
- $SD_{\text{correlación}}$ = Suma de las dos desviaciones estándar al cuadrado restadas por el producto de la constante (2), por la correlación de producto-Momento de Pearson, por la desviación estándar del pretest y del posttest, y obtención de la raíz cuadrada

Donde:

- SD_1 = Desviación estándar del pretest
- SD_2 = Desviación estándar del posttest
- r = Correlación de producto-Momento de Pearson (véase: Ponce, 2020).

Continúa...

Otra fórmula con correlación:

$$d_{\text{correlación } 2} = d_{\text{correlación}} / \sqrt{2(1-r)}$$

Supuestos: Las distribuciones de las poblaciones son normales y existe homogeneidad en las varianzas de los grupos.

Una fuente parcial: Goulet-Pelletier y Cousineau (2018).

Tabla E-3. Efectos alternativos al d de Cohen

Notación y ecuación	Divisor
<p>Comparación para un diseño con un grupo base y grupos independientes:</p> $\Delta = \frac{(\bar{x}_1 - \bar{x}_2)}{SD_{base}}$ <p>Donde:</p> <ul style="list-style-type: none"> • Δ = Se le conoce en el lenguaje inglés como Glass' Δ o Glass' Delta (la traducción al español es el delta de Glass o el Δ de Glass) • \bar{x}_1 = Promedio calculado de los datos del grupo 1 • \bar{x}_2 = Promedio calculado de los datos del grupo 2 • SD_{base} = Puede ser la desviación estándar del grupo control cuando se tiene un diseño experimental y se están comparando los promedios del grupo control contra el grupo tratamiento 	<p>$SD_{base} = SD_1$ o SD_2, dependiendo de cuál grupo se toma como base.</p>
<p>Comparación para un diseño con un grupo base con grupos independientes:</p> $g = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\text{Promedio de cuadrados dentro de los grupos}}}$ <p>Donde:</p> <ul style="list-style-type: none"> • g = Tamaño de efecto de Hedge (Hedge's g). Este coeficiente se recomienda con muestras menores a 20 • \bar{x}_1 = Promedio calculado de los datos del grupo 1 • \bar{x}_2 = Promedio calculado de los datos del grupo 2 • Promedio de cuadrados dentro de los grupos = Promedio de una suma de las varianzas dentro de cada grupo (véase: Hinkle <i>et al.</i>, 2013, bajo el tema del análisis de la varianza de un factor: ANOVA factorial) 	<p>Obtención del Promedio de cuadrados dentro de los grupos</p> <p>Primer paso: Obtener la suma de cuadrados dentro de los grupos (SCDG):</p> $SCDG = \sum_{j=1}^c \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2$ <p>Donde:</p> <ul style="list-style-type: none"> • $\sum_{j=1}^c$ = Suma de todos los grupos, donde c = Número de grupos • $\sum_{i=1}^{n_j}$ = Suma de las diferencias al cuadrado por cada grupo • x_{ij} = Representa cada uno de los valores en un grupo que se van a restar del promedio, para luego ser elevados al cuadrado y sumandos por grupo. Luego, todos los grupos son sumandos • \bar{x}_j = Promedio de uno de los grupos
<p>Una fórmula alternativa a la anterior es:</p> $g = d / \sqrt{\frac{d}{n/df}}$ $d = \frac{(\bar{x}_1 - \bar{x}_2)}{SD^*}$ <p>Donde:</p> <ul style="list-style-type: none"> • SD^* = Desviación estándar que puede variar (véase: Tabla E-1 y Tabla E-2 para ver cuál sería la manera más adecuada de calcularla) • n = Tamaño de la muestra • df = Grados de libertad 	<p>Segundo paso: Obtener el promedio de la suma de cuadrados dentro de los grupos:</p> $SCDG / n - c$ <p>Donde:</p> <ul style="list-style-type: none"> • n = Tamaño total de la muestra • c = Número de grupos

Continúa...

Nota: Efectos alternativos al d de Cohen pueden ser más precisos y, especialmente, con muestras pequeñas. Para más información sobre el delta de Glass (1976) y el g de Hedges, véase: Hedges (1981) y Hedges y Olkin (1985). Estos efectos pueden usar la siguiente calculadora: *Social Science Statistics* (Calculadora en línea: efectos de d , delta y g).

Supuestos: Las distribuciones de las poblaciones son normales.

Una fuente parcial: Goulet-Pelletier y Cousineau (2018).

Tabla E-4. Conversión del valor t en el d de Cohen

Un solo grupo:

$$d = \frac{t}{\sqrt{n}}$$

Donde:

- d = Tamaño del efecto para un solo grupo comparado con el promedio de una población
- t = Valor del valor $t_{\text{calculado}}$
- n = Tamaño de la muestra

Despejando t :

$$t = d \sqrt{n}$$

Grupos independientes:

$$d = \frac{2t}{\sqrt{df}}$$

Donde:

- d = Tamaño del efecto para grupos independientes
- 2 = Valor fijo dos
- t = Valor del valor $t_{\text{calculado}}$
- df = Grados de libertad - 2

Despejando t :

$$t = \frac{d \sqrt{df}}{2}$$

Grupos dependientes:

$$d = t \sqrt{\frac{2(1-r)}{n}}$$

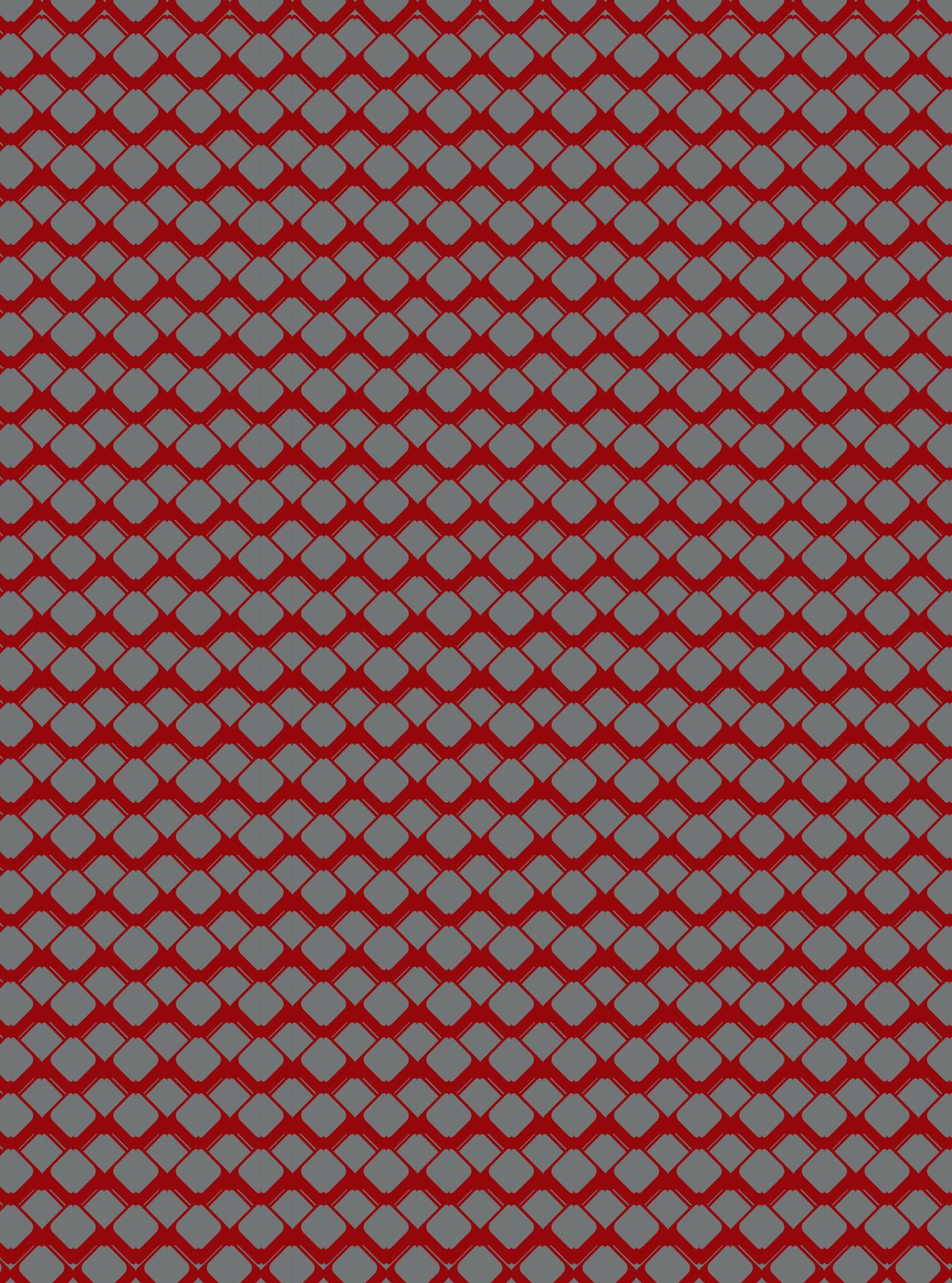
Donde:

- t = Valor del valor $t_{\text{calculado}}$
- d = Tamaño del efecto para grupos dependientes/medidas repetidas
- 2 = Valor fijo dos
- 1 = Valor fijo uno
- r = Correlación de producto-Momento de Pearson
- n = Tamaño de la muestra (un solo grupo)

Despejando t :

$$t = \frac{d}{\sqrt{\frac{2(1-r)}{n}}}$$

Fuente: Dunlap, Cortina, Vaslow y Burke (1996); Lenhard y Lenhard (2016).



APÉNDICE F

INTERVALO DE CONFIANZA PARA EL d DE COHEN

Un intervalo de confianza (*IC*) para un efecto como el d de Cohen ayuda a apreciar con qué tanta precisión se estimó este coeficiente. Por ejemplo, si un intervalo de confianza tiene un 95% de nivel de confianza quiere decir que, de 100 muestras tomadas, donde se haya calculado el d , 95 de ellas contendrán el valor verdadero de la población (δ). Hablando de precisión con el *IC*, si se tiene un $d = 0.20$ con un $[0.19, 0.21]$, se podría decir que este se calculó con mayor exactitud que otro coeficiente de $d = 0.20$ $[0, 0.40]$. En otras palabras, el primero de los d se estimó con más precisión que el segundo, el cual reporta un 0, que indica la posibilidad de sin efecto. Para mejorar la estimación del *IC*, se incrementa el tamaño de la muestra.

El cálculo de un *IC* para el d suele ser un reto, porque puede implicar distribuciones no-simétricas, como la t (véase: Kotz, & Nadarajah, 2004). Está más allá de los objetivos del presente libro esa estimación, pero, por esta razón, se recomiendan las siguientes fuentes para calcular un *IC*:

- ⊗ *Centre for Evaluation & Monitoring* (para calcular efectos como el d de Cohen, entre otros): <https://www.cem.org/effect-size-calculator>
- ⊗ Lenhard, W., & Lenhard, A. (2016). Calculation of Effect Sizes. *Psychometrica*: https://www.psychometrica.de/effect_size.html
- ⊗ *Real Statistics using Excel*. *ci Functions for Effect Size d* (sirve para calcular un intervalo de confianza para el d de Cohen): <https://www.real-statistics.com/students-t-distribution/confidence-interval-effect-size-power/ci-functions-effect-size-d/>
- ⊗ Uanhoro, J. O. (2017). Effect Size Calculators (para calcular intervalos de confianza para varios efectos, como el d de Cohen): <https://effect-size-calculator.herokuapp.com/>
- ⊗ Wuensch, K. L. (2012). *Using SPSS to Obtain a Confidence Interval for Cohen's d* : <http://core.ecu.edu/psyc/wuenschk/SPSS/CI-d-SPSS.pdf>

Advertencia: Las diversas calculadoras utilizan diferentes algoritmos y, por lo tanto, se obtienen diferentes resultados. La recomendación es usar un par o más para ver si los resultados coinciden. En el caso de la presente obra, los resultados de la calculadora de Uanhoro (2017) y de Lenhard y Lenhard (2016) no coincidieron. Se optó por la calculadora de Lenhard y Lenhard (2016), porque su IC capturaba el d de Cohen calculado a mano (véase Figura F-3).

Una sola muestra con el Test t

La Figura F-1 muestra cómo calcular un IC para el d de Cohen (véase la calculadora en línea de Uanhoro, 2017) de la sección b del Capítulo 2 de este libro llamada: Ejemplo de Test t con una sola muestra. En este caso, se llenan los campos de *Sample mean* (promedio de la muestra = 102); *Population mean* (promedio de la población = 100); *Sample standard deviation* ($SD = 1.63$); *Sample size* ($n = 16$); y el *Confidence Interval*, que puede variar, pero se selecciona el 95% por ser el más común. Los resultados son: $d = 1.23 [0.56, 1.87]$. Este IC indica que si se toman 100

muestras al azar, 95 de ellas capturarán el d de la población (i. e., δ). Otra interpretación fue dada por Ellis (2010): "... una manera mas segura de interpretar un intervalo de confianza del 95 es decir que tenemos un 95% de *confianza* qué el parámetro reside dentro los limites bajo y alto del intervalo estimado" (p. 18). La magnitud del *IC* indica la precisión con la que se midió el parámetro. Entre más precisión se desee más se va a ampliar el *IC*; otras cosas siendo iguales. Asimismo, si se aumenta el tamaño de la muestra, la amplitud del *IC* disminuye; *ceteris paribus*.

Figura F-1. Intervalo de confianza del d de Cohen para una sola muestra

One-sample t-test

Inputs

Sample mean: 102 Population mean: 100

Sample standard deviation: 1.63 Sample size: 16

Confidence Interval: 95 %

Calculate Clear

```
Entered values: {
":sample_mean": 102,
":pop_mean": 100,
":sample_sd": 1.63,
":n": 16,
":conf_int": 95
}
```

Results (CI using noncentral t distribution)

Cohen's d : 1.2269939

Lower limit on d : 0.560325 Upper limit on d : 1.87085

Fuente: Calculadora de Lenhard y Lenhard (2016).

Dos muestras independientes

La Figura F-2 muestra cómo calcular un *IC* para el d de Cohen (véase la calculadora en línea de Uanhoro, 2017) de la sección e del Capítulo 2 de esta obra llamada: Ejemplo de Test t para muestras independientes y el d de Cohen. En este caso, se llenan los campos de *Sample 1 Mean* (promedio de la muestra 1 = 9.5); *Sample 2 Mean* (promedio de la muestra 2 = 8.5); *Standard deviation* (muestra 1 = .33); *Standard deviation* (muestra

$d = .33$); *Sample size* (muestra 1; $n = 11$); *Sample size* (muestra 2; $n = 11$); y el *Confidence Interval*, que puede variar, pero se selecciona el 95% por ser el más común. Los resultados son: [1.76, 4.27]. También, esta calculadora produce el tamaño del efecto llamado g de Hedge (*Hedge's g* = 2.92): véase el Apéndice E (Tabla E-3) para más detalles. Asimismo, la calculadora convierte el g de Hedge en un Coeficiente $r = .84$ (véase: Ponce, 2020, para más información acerca de correlaciones).

En realidad, un Coeficiente de Correlación como el r no tiene mucho sentido cuando no se tienen medidas repetidas (e. g., una misma persona tomando un mismo examen en dos ocasiones). En contraparte, si el r es elevado al cuadrado, el r^2 indica cuánta varianza explica la variable independiente en la dependiente. Está más allá de los propósitos del presente libro el ahondar en temas de varianza explicada, así que se recomienda a Ponce (2020) para ello.

Figura F-2. Intervalo de confianza del d de Cohen para muestras independientes

Independent-samples t-test

Inputs

Sample 1	Sample 2
Mean: <input type="text" value="9.5"/>	Mean: <input type="text" value="8.5"/>
Standard deviation: <input type="text" value=".33"/>	Standard deviation: <input type="text" value=".33"/>
Sample size: <input type="text" value="11"/>	Sample size: <input type="text" value="11"/>
Confidence Interval: <input type="text" value="95"/> %	
<input type="button" value="Calculate"/> <input type="button" value="Clear"/>	
<p>Entered values: { ":mean1": 9.5, ":mean2": 8.5, ":sd_1": 0.33, ":sd_2": 0.33, ":n_1": 11, ":n_2": 11, ":conf_int": 95 }</p>	
Results (CI using noncentral t distribution)	
<u>Hedges' g (Unbiased):</u> <input type="text" value="2.9152282"/>	Lower limit on d : <input type="text" value="1.7609526"/>
<u>Conversion from g to r:</u> <input type="text" value="0.8368625"/>	Upper limit on d : <input type="text" value="4.2653362"/>

Fuente: Calculadora de Lenhard y Lenhard (2016).

Muestras dependientes

La Figura F-3 muestra cómo calcular un IC para el d de Cohen (véase la calculadora en línea de Uanhoro, 2017) de la sección g del Capítulo 2 de esta obra llamada: Ejemplo de muestras dependientes usando el Test t y el d de Cohen. En este caso, se llenan los campos de *Sample 1 Mean* (promedio de la muestra 1 = 8.64); *Sample 2 Mean* (promedio de la muestra 2 = 8.5); *Standard deviation* (muestra 1 = .33); *Standard deviation* (muestra 2 = .33); Coeficiente de Correlación $r = .975$; y *Number of pairs* ($n = 11$). Usando la calculadora de Uanhoro (2017), el *Confidence Interval* puede variar en el nivel de confianza, pero se selecciona el 95% por ser el más común. Los resultados son: [0.20, 0.65] y se recomienda usar el g de Hedge = 0.39 (en la Tabla E-3 del Apéndice E se tienen más detalles sobre este coeficiente).

El d calculado manualmente resultó en 1.99 (Ecuación 2.16). Este $d = 1.99$ no cabe en el IC de [0.20, 0.65]; asimismo, como este IC es contradictorio, se utiliza la calculadora de Lenhard y Lenhard (2016). La siguiente fórmula es empleada en el Capítulo 2; así también se sustituyen los valores para obtener el d :

$$d^* = \frac{\bar{d}}{SD_d} \quad (\text{Ecuación 2.16})$$

Donde:

- ⊗ $d^* = d$ de Cohen modificado para calcular el tamaño del efecto en muestras dependientes
- ⊗ \bar{d} = Promedio de las diferencias entre los valores de un test (en este caso, el pretest) y los de otro (postest)
- ⊗ SD_d = Desviación estándar de la diferencia
- ⊗ $d^* = 0.14091 / 0.0707 = 1.99$

A esta última se le introdujeron los datos antes mencionados. Da como resultado tres diferentes efectos: *Effect Size* $d_{\text{Repeated Measures}}$ (Tamaño del efecto de medidas repetidas; $d = 1.897$); *Effect Size* $d_{\text{Repeated Measures, pooled}}$ (Tamaño del efecto de medidas repetidas con desviación estándar

compuesta; $d = 1.897$); y *Effect Size* $d_{\text{Individual groups}}$ (Tamaño del efecto de grupos individuales = 0.424). Lenhard y Lenhard (2016) no especificaron en la página de la calculadora las ventajas o desventajas de estos diferentes efectos. Está más allá de los propósitos de este libro evaluar cuál efecto es más adecuado. Por ello, se recomienda consultar textos especializados que den efectos, como Ellis (2010), Goulet-Pelletier y Cousineau (2018), y Sakai (2018), entre otros. El IC = [0.891, 2.904] si captura al $d = 1.99$, que discrepa por 9 centésimas, aproximadamente, de los calculados por la calculadora de Lenhard y Lenhard (2016). Esto se pudo deber a emplear una fórmula diferente o menos dígitos al hacer la operación.

Figura F-3. Intervalo de confianza del d de Cohen para muestras dependientes: comparación entre calculadoras

Calculadora de Uanhoro

Paired-samples t-test

Inputs

Sample 1	Sample 2
Mean: <input type="text" value="8.64"/>	Mean: <input type="text" value="8.5"/>
Standard deviation: <input type="text" value=".33"/>	Standard deviation: <input type="text" value=".33"/>
Number of pairs: <input type="text" value="11"/>	r: <input type="text" value=".975"/>
Confidence Interval: <input type="text" value="95"/> %	
<input type="button" value="Calculate"/> <input type="button" value="Clear"/>	

Average	Repeated measures
Entered values: { ":mean1": 8.64, ":mean2": 8.5, ":sd_1": 0.33, ":sd_2": 0.33, ":n_pairs": 11 }	Entered values: { ":mean1": 8.64, ":mean2": 8.5, ":sd_1": 0.33, ":sd_2": 0.33, ":r": 0.975, ":n_pairs": 11, ":conf_int": 95 }

Results (CI using noncentral t distribution)

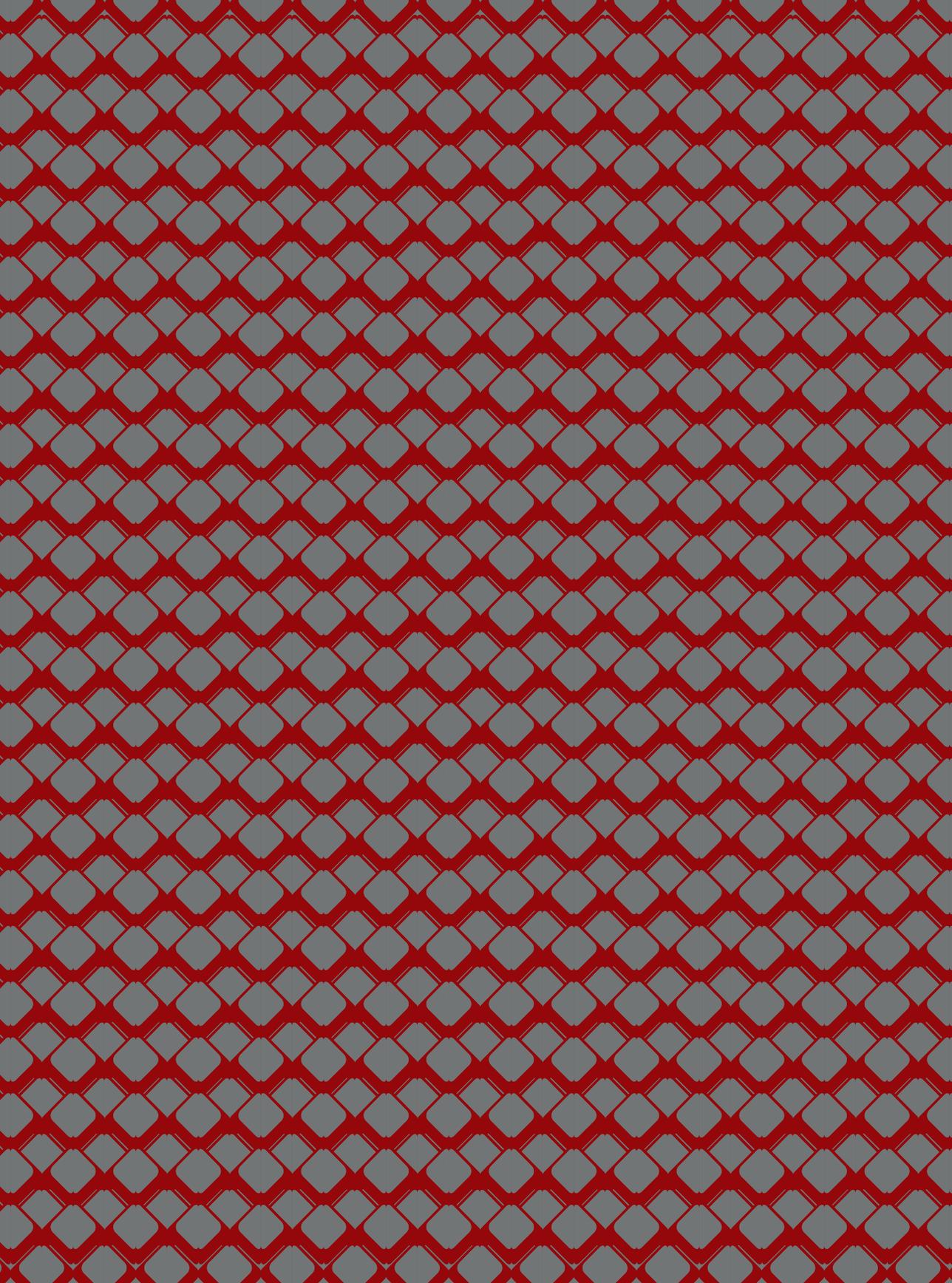
Hedges's g - average (recommended): <input type="text" value="0.3916084"/>	Lower limit on d : <input type="text" value="0.1949314"/>
Hedges's g - repeated measures: <input type="text" value="0.3916084"/>	Upper limit on d : <input type="text" value="0.6466318"/>

Fuente: Uanhoro (2017).

Calculadora de Lenhard y Lenhard

	Group 1	Group 2
Mean	8.5	8.64
Standard Deviation	.33	.33
Correlation	.975	
Effect Size $d_{\text{Repeated Measures}}$	1.897	
Effect Size $d_{\text{Repeated Measures, pooled}}$	1.897	
Effect Size $d_{\text{Individual Groups}}$	0.424	
N	11	
Confidence Coefficient	95%	
Confidence Interval for d_{RM}	0.891 - 2.904	

Fuente: Lenhard y Lenhard (2016).



REFERENCIAS

- American Psychological Association (APA) (2020). *Publication Manual of the American Psychological Association* (7th Ed.). American Psychological Association.
- Berenson, M. L., Levine, D. M., & Krehbiel, T. C. (2012). *Basic Business Statistics: Concepts and Applications* (12th Ed.). Pearson.
- Byrne, B. M. (2009). *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming*. Psychology Press.
- Cain, M. K., Zhang, Z., & Yuan, K.-H. (2017). Univariate and Multivariate Skewness and Kurtosis for Measuring Nonnormality: Prevalence, Influence and Estimation. *Behavior Research Methods*, 49(5), 1716-1735.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Ed.). Psychology Press.
- Cumming, G. (2013). *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-analysis*. Routledge.
- Devlin, K. (2012). *Introduction to Mathematical Thinking*. Keith Devlin.
- Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of Experiments with Matched Groups or Repeated Measures Designs. *Psychological Methods*, 1(2), 170-177. <https://doi.org/10.1037/1082-989X.1.2.170>

- Ellis, P. D. (2010). *The Essential Guide to Effect Sizes: Statistical Power, Meta-analysis, and the Interpretation of Research Results*. Cambridge University Press.
- Ercikan, K., & Roth, W.-M. (2006). What Good is Polarizing Research into Qualitative and Quantitative? *Educational Researcher*, 35(5), 14-23.
- Glass, G. V. (1976). Primary, Secondary, and Meta-analysis of Research. *Educational Researcher*, 5(10), 3-8. <https://doi.org/10.3102/0013189X005010003>
- Goulet-Pelletier, J. C., & Cousineau, D. (2018). A Review of Effect Sizes and their Confidence Intervals, Part I: The Cohen's *d* Family. *The Quantitative Methods for Psychology*, 14(4), 242-265. <https://doi.org/10.20982/tqmp.14.4.p242>
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2014). *Multivariate Data Analysis* (7th Ed.). Pearson New International Edition.
- Hedges, L. V. (1981). Distribution Theory for Glass's Estimator of Effect Size and Related Estimators. *Journal of Educational Statistics*, 6(2), 107-128.
- Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-analysis*. Academic Press.
- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied Statistics for the Behavioral Sciences* (5th Ed.). Houghton Mifflin.
- Hurley, P. J. (2014). *A Concise Introduction to Logic* (12th Ed.). Cengage Learning.
- Hurley, P. J., & Watson, L. (2018). *A Concise Introduction to Logic* (13th Ed.). Cengage Learning.
- Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's Reference Guide*. Scientific Software International.
- Kotz, S. (Ed.) (2006). *Encyclopedia of Statistical Sciences* (16 Vols.; 2nd Ed.). Wiley-Interscience.
- Kotz, S., & Nadarajah, S. (2004). *Multivariate *t* Distributions and their Applications*. Cambridge University Press.
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, 1(2), 259-269.

- LeMire, S. D. (2010). An Argument Framework for the Application of Null Hypothesis Statistical Testing in Support of Research. *Journal of Statistical Education*, 18(2), 1-23.
- Morris, D. (2016). *Bayes' Theorem: A Visual Introduction for Beginners*. Blue Windmill Media.
- Pallant, J. (2016). *SPSS Survival Manual* (6th Ed.). McGraw-Hill.
- Platt, J. R. (1964). Strong Inference. Certain Systematic Methods of Scientific Thinking May Produce Much More Rapid Progress than Others. *Science*, 146(3642), 347-353.
- Ponce, H. F. (2019). *Conceptos básicos de estadísticas inferenciales aplicadas a la investigación educativa*. Universidad Autónoma de Ciudad Juárez.
- (2020). *Estadística elemental para investigación educativa: probabilidad, distribuciones y correlación*. Universidad Autónoma de Ciudad Juárez.
- Ponce, H. F., Domínguez, C. T., & Arriaga, M. (2016). La importancia de la investigación en la educación especial. *Nósis*, 25(50), 216-242.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd Ed.). Sage.
- Rossi, R. J. (2018). *Mathematical Statistics: An Introduction to Likelihood Based Inference*. Wiley.
- Sakai, T. (2018). *Laboratory Experiments in Information Retrieval: Sample Sizes, Effect Sizes, and Statistical Power*. Springer.
- Salkind, N. J. (2007). *Encyclopedia of Measurement and Statistics* (Vols. 1, 2, 3). Sage.
- (2011). *Statistics for People Who (Think They) Hate Statistics* (4th Ed.). Sage.
- Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating Causal Effects using Experimental and Observational Designs*. American Educational Research Association.
- Schumacker, R., & Tomek, S. (2013). *Understanding Statistics using R*. Springer.
- Sudman, S. (1976). *Applied Sampling*. Academic Press.
- Tabachnick, B. G., & Fidell, L. S. (2012). *Using Multivariate Statistics* (6th Ed.). Pearson.

- VandenBos, G. R. (Coord.) (2015). *APA Dictionary of Psychology* (2nd Ed.). American Psychological Association.
- Walker, H. M. (1940). Degrees of Freedom. *Journal of Educational Psychology*, 31(4), 253-269.
- Wechsler, D. (1997). *The Wechsler Adult Intelligence Scale* (3rd Ed.). The Psychological Corporation.
- Wellek, S. (2010). *Testing Statistical Hypotheses of Equivalence and Non-inferiority* (2nd Ed.). CRC Press.
- Wuensch, K. L. (2012). *Using SPSS to Obtain a Confidence Interval for Cohen's d*. <http://core.ecu.edu/psyc/wuenschk/SPSS/CI-d-SPSS.pdf>

Libros recientes para usar SPSS

- Cronk, B. C. (2020). *How to Use SPSS®: A Step-by-step to Analysis and Interpretation* (11th Ed.). Routledge.
- Denis, D. J. (2019). *SPSS Data Analysis for Univariate, Bivariate, and Multivariate Statistics*. John Wiley & Sons.
- George, D., & Mallery, P. (2020). *IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference* (16th Ed.). Routledge.
- Howitt, D., & Cramer, D. (2017). *Understanding Statistics in Psychology with SPSS* (7th Ed.). Pearson.
- Salcedo, J., & McCormick, K. (2020). *SPSS Statistics for Dummies* (4th Ed.). John Wiley & Sons.
- Sarstedt, M., & Mooi, E. (2019). *A Concise Guide to Market Research: The Process, Data and Methods using IBM SPSS Statistics* (3rd Ed.). Springer.
- Stockemer, D. (2019). *Quantitative Methods for the Social Science: A Practical Introduction with Examples in SPSS and Stata*. Springer.

Recursos gratuitos de internet

- Centre for Evaluation & Monitoring. Effect Size Calculator (para calcular efectos como el *d* de Cohen, entre otros) (<https://www.cem.org/effect-size-calculator>).
- G*Power (Allgemeine Psychologie und Arbeitspsychologie) (<http://www.gpower.hhu.de/>).

- Lenhard, W., & Lenhard, A. (2016). Calculation of Effect Sizes. *Psychometrica* (https://www.psychometrica.de/effect_size.html).
- Mathcracker, Solvers Statistics* (Critical Z-Values: para valores z críticos) (https://mathcracker.com/z_critical_values.php).
- Minitab* (<http://www.minitab.com/es-mx/>).
- R* (*The R Project for Statistical Computing*) (<https://www.r-project.org/>).
- Real Statistics using Excel*. CI Functions for Effect Size d (sirve para calcular un intervalo de confianza para el d de Cohen) (<https://www.real-statistics.com/students-t-distribution/confidence-interval-effect-size-power/ci-functions-effect-size-d/>).
- SAS* (Analytics Platform) (https://www.sas.com/en_us/home.html).
- Social Science Statistics* (calculadora en línea: efectos de *d*, *delta* y *g*) (<https://www.socscistatistics.com/effectsize/default3.aspx>).
- (calculadora en línea: valores *p* de los puntajes *t*) (<https://www.socscistatistics.com/pvalues/tdistribution.aspx>).
- (calculadora en línea: valores *p* de los puntajes *z*) (<http://www.socscistatistics.com/pvalues/normaldistribution.aspx>).
- (calculadora en línea: Test *t* para muestras dependientes) (<https://www.socscistatistics.com/tests/ttestdependent/default2.aspx>).
- SPSS* (IBM Software) (<https://www.ibm.com/analytics/mx/es/technology/spss/index.html>).
- Uanhoro, J. O. (2017). Effect Size Calculators (para calcular intervalos de confianza para varios efectos, como el *d* de Cohen) (<https://effect-size-calculator.herokuapp.com/>).
- Z Score Table* (tabla para calcular los valores *z*, así como calculadora para este mismo propósito y explicaciones de cómo funciona esta) (<http://www.z-table.com/>).
- Σ + *SPSS Tutorials* (Test de Normalidad de Kolmogorov-Smirnov) (<https://www.spss-tutorials.com/spss-kolmogorov-smirnov-test-for-normality/>).

Recomendaciones de bibliografía clásica y moderna para la estadística en investigación educativa

- Bologna, E. (2013). *Estadística para psicología y educación* (3ª ed.). Brujas.
- Cohen, J. (1973). Eta-squared and Partial Eta-squared in Fixed Factor Anova Designs. *Educational and Psychological Measurement*, 33(1), 107-112. <https://doi.org/10.1177/001316447303300111>
- (1992). A Power Primer. *Psychological Bulletin*, 112(1), 155-159. <https://doi.org/10.1037/0033-2909.112.1.155>
- (1992). Statistical Power Analysis. *Current Directions in Psychological Science*, 1(3), 98-101. <https://doi.org/10.1111/1467-8721.ep10768783>
- (1992). Things I Have Learned (so Far). In A. E. Kazdin (Ed.), *Methodological Issues & Strategies in Clinical Research* (pp. 315-333). American Psychological Association. <https://doi.org/10.1037/10109-028>
- (1994). The Earth is Round ($p < .05$). *American Psychologist*, 49(12), 997-1003. <https://doi.org/10.1037/0003-066X.49.12.997>
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd Ed.). Lawrence Earlbaum.
- Cumming, G., & Finch, S. (2001). A Primer on the Understanding, Use, and Calculation of Confidence Intervals that are Based on Central and Noncentral Distributions. *Educational and Psychological Measurement*, 61(4), 532-574. <https://doi.org/10.1177/0013164401614002>
- (2005). Inference by Eye: Confidence Intervals and How to Read Pictures of Data. *American Psychologist*, 60(2), 170-180. <https://doi.org/10.1037/0003-066X.60.2.170>
- Estrellado, R. A., Freer, E. A., Mostipak, J., Rosenberg, J. M., & Velásquez, I. C. (2021). *Data Science in Education Using R*. Routledge.
- Fischer, H. (2010). *A History of the Central Limit Theorem: From Classical to Modern Probability Theory*. Springer.
- Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver & Boyd.
- Gaciu, N. (2021). *Understanding Quantitative Data in Educational Research* (1st Ed.). Sage.

- Gall, M. D., Borg, W. R., & Gall, J. P. (2006). *Educational Research: An Introduction* (8th Ed.). Allyn and Bacon.
- Illukkumbura, A. (2020). *Introduction to Hypothesis Testing* (1st Ed.). Easy Statistics Series.
- Lakens, D. (2013). Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for *t*-Test and ANOVAS. *Frontiers in Psychology*, 4(863), 1-12.
- Ross, S. M. (1997). *Introduction to Probability Models* (6th Ed.). Academic Press.
- Sawilowsky, S. S. (2009). New Effect Size Rules of Thumb. *Journal of Modern Applied Statistical Methods*, 8(2), 597-599.
- Schumacker, R., & Tomek, S. (2013). *Understanding Statistics using R*. Springer.
- Siegel, S., & Castellan, N. J. (1995). *Estadística no paramétrica aplicada a las ciencias de la conducta*. Trillas.
- Thalheimer, W., & Cook, S. (2002). *How to Calculate Effect Sizes from Published Research: A Simplified Methodology*. Retrieved September 4th, 2018, from www.bwgriffin.com/gsu/courses/edur9131/content/Effect_Sizes_pdf5.pdf
- Thompson, B. (2008). Computing and Interpreting Effect Sizes, Confidence Intervals, and Confidence Intervals for Effect Sizes. In J. W. Osborne (Ed.), *Best Practices in Quantitative Methods* (pp. 246-262). Sage.
- Yu, C. H. (2006). *Philosophical Foundations of Quantitative Research Methodology*. University Press of America.

