

$$\frac{P(x)}{Q(x)} = \frac{G(x) + \frac{R(x)}{Q(x)}}{Q(x)}$$

$$y^{1/n} = x$$

$$S^2 = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}}$$

$$\sim \forall x \forall y [p(x, y) \equiv \exists x \exists y [\sim p(x, y)]]$$

$$(a^m)^n = a^{m \times n}$$



$$S^2 = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}}$$

$$\cos \frac{A}{2} = \pm \sqrt{\frac{1 + \cos A}{2}}$$



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA



Estadística en Acción, Educación y Movimiento

Una guía práctica con software libre y aplicado

Patricia Guevara Vallejo, Pablo Guevara Vallejo
Emma Guevara Vallejo

Estadística en Acción, Educación y Movimiento

Patricia Guevara Vallejo, Pablo Guevara Vallejo y Emma Guevara Vallejo

Primera edición electrónica: agosto, 2026

ISBN: 978-9942-652-53-9

Ingreso al proceso: 28/01/2026 • Aprobación de ingreso: 26/02/2026

Aprobación de pares: 23/03/2026 • Finalización de edición: 12/06/2026

Universidad de las Fuerzas Armadas ESPE

Oswaldo Mauricio González Mosquera, Crnl., Ph. D.

Rector

Publicación autorizada por:

Comisión Editorial de la Universidad de las Fuerzas Armadas-ESPE

Cpfg. César Antonio Portero Barrantes, Mgtr. - Vicerrector de Investigación,
Innovación y Transferencia de Tecnología (E) - Presidente

Corrección de estilo y diseño

Mtr. Xavier Chinga Mármol

Imagen de cubierta: Archivo de autores

Derechos reservados. Se prohíbe la reproducción de esta obra por cualquier medio impreso, reprográfico o electrónico. El contenido, uso de fotografía, gráficos, cuadros, tablas, y referencias es de exclusiva responsabilidad de los autores.

Universidad de las Fuerzas Armadas ESPE

Av. General Rumiñahui s/n, Sangolquí, Ecuador

www.espe.edu.ec

Los derechos de esta edición electrónica son de la Universidad de las Fuerzas Armadas ESPE, para consulta de profesores y estudiantes de la universidad e investigadores en

www.repositorio.espe.edu.ec.

Estadística en Acción: Educación y Movimiento

Una guía práctica con software libre y aplicado

Patricia Guevara Vallejo
Pablo Guevara Vallejo
Emma Guevara Vallejo

EDITORIAL



UNIVERSIDAD DE LAS FUERZAS ARMADAS - ESPE

AGRADECIMIENTOS

Expresamos nuestro sincero agradecimiento a la Universidad de las Fuerzas Armadas ESPE por hacer posible la elaboración de este libro, concebido como un aporte a la formación académica de los estudiantes.

Asimismo, expresamos nuestro reconocimiento al Departamento de Ciencias Exactas por la confianza depositada y el valioso respaldo brindado a esta propuesta.

De manera especial, agradecemos a los revisores de esta obra por su significativa contribución.

DEDICATORIA

A nuestros padres †, ejemplo de vida y fuente de inspiración.

Presentación

Estadística en Acción: Educación y Movimiento es una obra dirigida a estudiantes y profesionales de la Educación en la primera infancia y de la Actividad Física.

Estadística en Acción: Educación y Movimiento trasciende la mera exposición de fórmulas y procedimientos para promover la aplicación del conocimiento, fomenta el análisis y el pensamiento estadístico para transformar las preguntas en hipótesis y los datos en evidencia.

Estadística en Acción: Educación y Movimiento está diseñado para motivar el aprendizaje estadístico con aplicaciones educativas, que se complementan con los laboratorios de utilización de software especializado como jamovi y Microsoft Excel, proporcionando guías detalladas del procedimiento sin descuidar la interpretación de los resultados dentro de cada contexto. Este enfoque permite un balance entre la comprensión conceptual y las competencias instrumentales o tecnológicas.

Los autores.

Introducción

Estadística en Acción: Educación y Movimiento se desarrolla en seis capítulos: Estadística Descriptiva, Probabilidades, Variables Aleatorias, Estimación de parámetros, Pruebas de hipótesis para una muestra y Variables en la práctica. En los primeros cinco capítulos se presentan definiciones, propiedades, ejemplos tipo, aplicaciones con software estadístico y ejercicios propuestos; finalizando con un capítulo de propuestas de análisis estadístico.

El capítulo de Estadística Descriptiva presenta las definiciones y herramientas básicas para realizar un análisis exploratorio de datos mediante tablas, gráficos y medidas descriptivas.

El capítulo de probabilidades estudia las propiedades y axiomas de la probabilidad que son de utilidad para comprender las variables aleatorias, tanto discretas como continuas, y sus principales distribuciones de probabilidad.

En los capítulos cuatro y cinco se estudian la Estimación de parámetros y Pruebas de hipótesis para una muestra, con métodos y procedimientos que permiten extrapolar los resultados de la muestra hacia la población. El capítulo Variables en la práctica es una propuesta de análisis estadístico a partir de un contexto, sin llegar a su ejecución.

Se ha elegido el software estadístico jamovi para el análisis estadístico por ser de código abierto y tener una interfaz amigable que facilita la generación de resultados, dejando al lector la interpretación de los mismos.

En la parte final del libro, se presentan los Apéndices con tablas de distribución de probabilidad, datos y respuestas de ejercicios seleccionados.

Patricia Guevara Vallejo
peguevara@espe.edu.ec, patyguevara.v@gmail.com



Máster en Bioinformática y Bioestadística por la Universitat Oberta de Catalunya y la Universidad de Barcelona; Máster en Investigación de Mercados por la Universidad Internacional de La Rioja; y Máster en Docencia Universitaria por la Universidad de las Fuerzas Armadas ESPE. Licenciada en Ciencias de la Educación, con especialidad en Físico-Matemáticas, por la Universidad Técnica Particular de Loja (UTPL), y Doctora en Matemática por la Escuela Superior Politécnica de Chimborazo. Docente de Estadística del Departamento de Ciencias Exactas. Ha cumplido varias funciones como Docente de apoyo en la Universidad de las Fuerzas Armadas ESPE. Fue docente de la Universidad Técnica Particular de Loja y de la Escuela Superior Politécnica del Chimborazo y otras instituciones de educación superior del país a nivel de grado y posgrado.

Pablo Guevara Vallejo

pvguevarav@ejercito.mil.ec



Máster en Entrenamiento Deportivo y Máster en Estrategia Militar Terrestre, otorgados por la Universidad de las Fuerzas Armadas ESPE. Licenciado en Ciencias Militares y en Actividad Física y Deportes por la misma institución. Ha desempeñado funciones como docente e instructor en la Escuela de Caballería Blindada, AGE y la Escuela de Formación de Soldados. Se desempeñó como Jefe de Cultura Física del Ejército entre 2020 y 2023.

Emma Guevara Vallejo

ecguevara@espe.edu.ec



Máster en Educación Inicial otorgado por la Universidad Central del Ecuador; Licenciada en Ciencias de la Educación, especialidad Educación Infantil por la Universidad de las Fuerzas Armadas ESPE; Diplomado en Docencia Universitaria e Investigación CIFE - México. Docente del departamento de Ciencias Humanas y Sociales de la Universidad de las Fuerzas Armadas ESPE. Fue Técnico docente de la Universidad Central del Ecuador y Docente de la carrera de Educación Inicial en la Universidad Indoamérica.

Índice

Capítulo 1 - Estadística Descriptiva	1
1.1 Introducción	3
1.1.1 La Estadística y sus áreas	3
1.1.2 Población y muestra	3
1.2 Variables	4
1.2.1 Niveles de medida	4
1.2.2 Valores y etiquetas de las variables	5
1.3 Fuentes de información	5
1.3.1 La Encuesta	6
1.4 Organización de la información	8
1.4.1 Matriz de datos	8
1.4.2 Software estadístico	8
1.4.3 Análisis de la información	8
1.5 Laboratorio 1. Introducción a Microsoft Excel	9
1.5.1 Creación de variables con Excel	9
1.6 Laboratorio 2. Introducción a jamovi	10
1.6.1 Sitio de descarga y registro	10
1.6.2 Actualización de la versión de jamovi	11
1.6.3 Menú de jamovi	11
1.6.4 Creación de variables con jamovi	12
1.7 Tabla de distribución de frecuencias	13
1.7.1 Tabla de frecuencias, caso 1	14
1.7.2 Tabla de frecuencias, caso 2	16
1.7.3 Tabla de contingencia	16
1.8 Gráficos estadísticos	17
1.9 Laboratorio 3. Tablas y Gráficos	21
1.9.1 Creación de variables en jamovi	21
1.9.2 Tablas y gráficos para variables cualitativas o discretas	22
1.9.3 Integración de R en jamovi	28
1.9.4 Tablas y gráficos para variables continuas con Excel	30

1.10	Medidas descriptivas	33
1.10.1	Medidas de tendencia central	33
1.10.2	Medidas de dispersión	33
1.10.3	Medidas de posición	34
1.10.4	Medidas de forma	35
1.11	Ejemplos de reforzamiento	35
1.12	Laboratorio de medidas descriptivas	39
1.12.1	Medidas descriptivas con jamovi	39
1.12.2	Medidas descriptivas con Excel	42
1.13	Ejercicios propuestos del capítulo 1	43
Capítulo 2 - Probabilidades		47
2.1	Introducción	49
2.2	Conceptos básicos	49
2.2.1	Axiomas de la probabilidad	49
2.3	Técnicas de conteo	53
2.4	Reglas de probabilidad	55
2.4.1	Regla de adición de eventos	55
2.4.2	Regla de probabilidad conjunta	58
2.4.3	Regla de probabilidad marginal	58
2.4.4	Tablas de contingencia y diagramas de árbol para probabi- lidades	61
2.4.5	Probabilidad condicional <i>a posteriori</i>	66
2.5	Laboratorio de probabilidades en Jamovi	70
2.6	Ejercicios propuestos del capítulo	75
Capítulo 3 - Variables aleatorias		79
3.1	Introducción	81
3.1.1	Tipos de variables aleatorias	81
3.1.2	Modelos de distribución de probabilidad	82
3.2	Variables aleatorias discretas	82
3.2.1	Función de probabilidad para una variable aleatoria discreta	83
3.2.2	Función de distribución para una variable aleatoria discreta	83
3.2.3	Media y varianza de una variable aleatoria discreta	84
3.3	Variables aleatorias continuas	88
3.3.1	Resumen de funciones para una variable aleatoria conti- nua X	88

3.3.2	Resumen elementos clave de las variables aleatorias continuas	89
3.4	Distribución binomial	90
3.4.1	Características del experimento aleatorio binomial	90
3.4.2	Medidas descriptivas de la variable aleatoria binomial	90
3.5	Distribución Hipergeométrica	92
3.5.1	Características del experimento aleatorio hipergeométrico	92
3.5.2	Medidas descriptivas para la variable aleatoria hipergeométrica	93
3.6	Distribución Poisson	95
3.6.1	Características del experimento aleatorio Poisson	95
3.6.2	Medidas descriptivas para la variable aleatoria Poisson	97
3.7	Distribución normal	99
3.7.1	Introducción	99
3.7.2	Funciones de densidad y distribución	99
3.7.3	Variable aleatoria normal estándar	100
3.7.4	Distribución normal estándar	100
3.7.5	Tabla de distribución acumulativa $F(z)$	101
3.8	Distribución t-Student	111
3.8.1	Introducción	111
3.8.2	Tabla de distribución t-Student	111
3.9	Laboratorio de variables aleatorias	113
3.9.1	Distribuciones de probabilidad discretas y continuas con Excel	113
3.10	Ejercicios del capítulo	121
Capítulo 4 - Estimación de parámetros		127
4.1	Introducción	129
4.2	Conceptos	129
4.3	Distribución es muestrales	131
4.3.1	Distribución muestral de las medias	132
4.3.2	Teorema del límite central	132
4.3.3	Distribución muestral de la proporción	133
4.4	Estimación por intervalo	135
4.5	Intervalo de confianza bilateral para la media poblacional μ	135
4.5.1	Intervalo de confianza bilateral para la media poblacional μ , con varianza poblacional σ^2 conocida.	136

4.5.2	Intervalo de confianza bilateral para la media poblacional μ , con varianza poblacional σ^2 desconocida	138
4.6	Intervalo de confianza bilateral para la proporción poblacional p	139
4.7	Tamaño de la muestra para poblaciones infinitas	141
4.7.1	Consideraciones generales sobre el tamaño de la muestra	142
4.7.2	Tamaño de la muestra para estimar la media	142
4.7.3	Tamaño de la muestra para estimar la proporción	144
4.8	Introducción a las técnicas de muestreo	146
4.8.1	Muestreo no probabilístico	147
4.8.2	Muestreo probabilístico	147
4.9	Laboratorio de Estimación	149
4.9.1	Intervalo de confianza para la media en Jamovi	149
4.9.2	Intervalo de confianza para la proporción en Jamovi	150
4.9.3	Muestreo en Jamovi	151
4.10	Ejercicio propuestos del capítulo	154
Capítulo 5 - Pruebas de hipótesis para una muestra		157
5.1	Definiciones básicas	159
5.1.1	Clasificación de las Pruebas de Hipótesis	159
5.1.2	Errores de Tipo I y tipo II	162
5.1.3	Pasos para realizar una prueba de hipótesis	163
5.2	Prueba de hipótesis sobre la media poblacional μ	164
5.2.1	Prueba de hipótesis sobre la media μ con varianza σ^2 conocida	164
5.2.2	Prueba de hipótesis sobre la media μ con varianza σ^2 desconocida	167
5.3	Uso del p-valor	169
5.4	Prueba de hipótesis sobre la proporción poblacional p	173
5.4.1	Laboratorio de Pruebas de hipótesis	177
5.5	Ejercicio propuestos del capítulo	182
Capítulo 6 - Variables en la práctica		185
6.1	Introducción	187
6.1.1	Esquema de análisis	187
6.2	Variables en el desarrollo infantil	187
6.2.1	Contexto motricidad	187
6.2.2	Contexto: Autonomía de niños	192

6.3	Variables en la actividad física	194
6.3.1	Contexto Medidas antropométricas	194
6.3.2	Cotexto: Somatotipo	196
6.3.3	Contexto: Ciclismo de ruta	198
6.3.4	Contexto: Lesiones deportivas	201
6.4	Ejercicios del capítulo	203
Apéndices		207
Apéndice A. Datos		209
A.1	Datos de pesos	209
A.2	Corpus del habla de niños bilingües	210
Apéndice B. Tablas de probabilidad		211
B.1	Distribución normal	211
B.2	Distribución t-Student	213
Apéndice C.		
	Respuestas de ejercicios seleccionados	215



Fuente Magnific AI

CAPÍTULO 1

Estadística Descriptiva

1.1 Introducción

La estadística se aplica en múltiples campos del conocimiento. Por ejemplo, en Educación es frecuente aplicar la estadística para evaluar el rendimiento académico; en Medicina, para determinar la efectividad de tratamientos médicos; en Sociología, para estudiar fenómenos sociales; y en la actividad física, para analizar el rendimiento de un deportista en función de las diferentes tablas de entrenamiento.

1.1.1 La Estadística y sus áreas

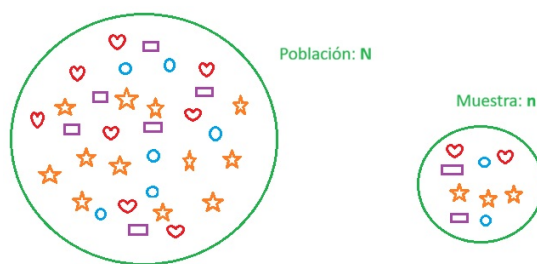
La estadística permite concluir, inferir y predecir variables o eventos mediante el análisis de patrones, relaciones y tendencias a partir de datos.

La estadística tiene dos grandes campos: la Estadística Descriptiva y la Estadística Inferencial. La estadística descriptiva permite explorar datos mediante su representación en tablas, gráficos y medidas descriptivas, sin hacer proyecciones sobre la población. La estadística inferencial permite realizar inferencias sobre la población, apoyándose en varias herramientas de la teoría de la estimación, modelos de regresión, entre otras.

1.1.2 Población y muestra

Figura 1.1

Población y muestra



La **población** o universo de estudio es el conjunto de elementos o unidades de observación que tienen características comunes y que son objeto de estudio. El estudio completo de la población es un censo. El número de elementos de la población se designa por N .

La **muestra** es una parte representativa de la población. El estudio de la muestra se denomina muestreo. El número de elementos de la muestra se designa por n , como se observa en la figura 1.1.

Ejemplo 1-1. Si la población está formada por estudiantes de bachillerato matriculados en un año lectivo 2024-2025 en instituciones de educación públicas del sector urbano en la ciudad de Quito, cuál sería la muestra probabilística?

La muestra de estudiantes se elige de forma aleatoria o sistemática a partir de la nómina de estudiantes.

Ejemplo 1-2. Si la población de interés son los deportistas de alto rendimiento de la provincia de la provincia de Pichincha, cuál sería la muestra no probabilística?

La muestra no probabilística sería el conjunto de deportistas de alto rendimiento de la provincia de Pichincha que acudieron a una competencia provincial.

1.2 Variables

En general, las variables se clasifican en dos tipos: cuantitativas y cualitativas. Las variables cuantitativas son susceptibles de medición, mientras que las cualitativas solo toman atributos o cualidades. Las variables se identifican con letras mayúsculas.

1.2.1 Niveles de medida

A cada tipo de variable le corresponde un nivel de medida. Las variables cuantitativas tienen dos niveles de medida: razón e intervalo, cuya diferencia se describe en la Tabla 1.1. Las variables cualitativas tienen los niveles de medida ordinal (interesa el orden de los atributos) y nominal (no interesa el orden de los atributos).

Tabla 1.1

Diferencias entre los niveles de medida intervalo y razón

Característica	Intervalo	Razón
Cero absoluto	No	Sí
Comparaciones de razón o proporcionalidad	No	Sí

Ejemplo 1-3. En la Tabla 3.4 se presentan los cuatro niveles de medida con 4 variables. Elabore otros ejemplos aplicados a su carrera.

Tabla 1.2

Descripción de variables

Variable	Tipo	Nivel de medida
Peso de una persona [kg]	Cuantitativa	Razón
Temperatura corporal [°C]	Cuantitativa	Intervalo
Tipo de institución	Cualitativa	Nominal
Nivel de estudios	Cualitativa	Ordinal

1.2.2 Valores y etiquetas de las variables

Antes de organizar la información es necesario asignar nombres cortos a las variables. Los atributos de las variables cualitativas deben etiquetarse y describirse. Por ejemplo, usando las variables de la Tabla 1.3, se tiene el siguiente detalle:

Tabla 1.3

Codificación de variables

Variable	Código	Valores y Etiquetas
Peso de una persona [kg]	peso	No aplica
Temperatura corporal [°C]	temperatura	No aplica
Tipo de institución	tipoInst	1. Pública 2. Privada
Nivel de estudios	nivEst	1. Grado 2. Posgrado 3. PhD

1.3 Fuentes de información

Las fuentes para recoger se clasifican en primarias y secundarias. Instrumentos como la encuesta, entrevista, observación directa, experimentación, simulación, entre otras fuentes, se consideran fuentes de información primaria. Los instrumentos para obtener esta información secundaria son las fuentes bibliográficas, revistas, diarios y más documentos similares tanto impresos como digitales. A continuación, se describe la encuesta.

1.3.1 La Encuesta

La encuesta es un cuestionario dirigido a individuos, con el fin de conocer su información y opinión con respecto a un tema de interés.

1.3.1.1 TIPOS DE PREGUNTAS

Las preguntas abiertas tienen respuesta libre numérica o de opinión. Las preguntas cerradas de opción múltiple pueden tener selección simple o selección múltiple dentro de un conjunto de atributos o categorías.

1.3.1.2 ELEMENTOS DE LA ENCUESTA

- Encabezado con datos de identificación, código o número de encuesta.
- Objetivo general de la investigación.
- Instrucciones sobre el cuestionario.
- Datos informativos o demográficos del encuestado.
- Cuerpo de la encuesta, organizado en secciones o bloques de preguntas.

1.3.1.3 RECOMENDACIONES PARA ELABORAR LAS PREGUNTAS:

- Lenguaje claro acorde al grupo a quién va dirigida la encuesta.
- Preguntas sin ambigüedad evitando preguntas compuestas.
- Mantener la escala en un solo sentido por ejemplo ascendente.
- Evitar el exceso de preguntas.
- Evitar preguntas que sesguen hacia una respuesta.

Ejemplo 1-4. Encuesta diseñada para adolescentes sobre el uso del tiempo libre con los elementos y recomendaciones dadas.

ENCUESTA SOBRE EL USO DEL TIEMPO LIBRE EN ADOLESCENTES

Objetivo: Determinar las actividades de tiempo libre por preferencia y frecuencia en adolescentes.

Instrucciones: Conteste las preguntas con veracidad, indicando la alternativa que mejor refleje su realidad.

DATOS INFORMATIVOS

- Edad: _____ años ■ Curso: _____
- Género: ☐ M ☐ F ☐ Otro ■ Sector: ☐ Urbano ☐ Rural

SECCIÓN 1: Actividades en el tiempo libre

Marque con una "x" en cada fila el número de días por semana de su actividad

Actividad	0	1-2	3-4	5-7
Ver TV / series	☐	☐	☐	☐
Jugar videojuegos	☐	☐	☐	☐
Practicar deportes	☐	☐	☐	☐
Redes sociales	☐	☐	☐	☐
Lectura	☐	☐	☐	☐
Salidas sociales	☐	☐	☐	☐
Otras actividades	☐	☐	☐	☐

SECCIÓN 2: Percepciones y hábitos

- ¿Cuántas horas al día dedicas a tu tiempo libre?
☐ Menos de 1 hora ☐ 1-2 horas ☐ 3-4 horas ☐ Más de 4 horas
- ¿Sientes que aprovechas bien tu tiempo libre?
☐ Sí ☐ A veces ☐ No
- ¿Te gustaría cambiar la forma en que usas tu tiempo libre?
☐ Sí ☐ No ☐ No estoy seguro/a
- ¿Participas en actividades organizadas fuera del horario escolar?
☐ Sí ☐ No

Gracias por tu participación

1.4 Organización de la información

1.4.1 Matriz de datos

Para recoger la información, se debe diseñar una matriz de datos en la que se organizará la información con el objetivo de hacer operativas las variables. Una vez obtenida la información con el instrumento seleccionado, se registran los datos o se descarga la matriz de datos en un formato adecuado, según corresponda. Observe la Tabla 1.4 que contiene cuatro filas de datos.

Tabla 1.4

Matriz de datos

ID	Peso	Temperatura	Tipo de institución	Nivel de estudios
Id1	80	36.0	Pública	Posgrado
Id2	72	35.5	Privada	Grado
Id3	68	36.1	Privada	Grado
Id4	65	36.3	Pública	Grado

1.4.2 Software estadístico

Es importante elegir el software estadístico apropiado para el análisis, teniendo en cuenta la diversidad de opciones, tanto de software libre como de pago. Para el caso del software libre, se tienen R y R Studio, Python, Jupyter, Kinime, Tableau público y jamovi; mientras que, para el caso de software con licencia, se tienen SPSS, Prism, SAS, Stata y Minitab. Incluso Excel es una aplicación que proporciona importantes funciones de análisis estadístico. Para el desarrollo del texto, se han elegido jamovi y Excel.

1.4.3 Análisis de la información

Para analizar la información, se debe contar previamente con el tipo de investigación que, a su vez, se desprende de la metodología donde se definen las herramientas de análisis estadístico apropiadas. Todas estas herramientas de análisis estadístico se abordarán más adelante, desde su fundamentación teórica hasta la práctica con el apoyo de software estadístico o Excel.

1.5 Laboratorio 1. Introducción a Microsoft Excel

Excel es un programa que forma parte de las aplicaciones de Microsoft. Está formado por hojas de cálculo en las que se puede organizar, analizar y visualizar la información.

1.5.1 Creación de variables con Excel

Para que los datos creados en Excel puedan ser usados apropiadamente y, en muchos casos, exportados a otras aplicaciones de análisis estadístico, se deben crear las variables desde la primera columna y los valores de las variables se ingresan desde la segunda fila. Por el mismo motivo, es recomendable generar los resultados y comentarios en otra hoja. Observe el ejemplo de matriz con datos de la figura 1.2 creada en Excel, a partir de la matriz de datos que se observa en la sección Tabla 1.4

Figura 1.2

Matriz de datos en Excel

	A	B	C	D	E
1	ID	Peso	Temperatura	Tipo de institución	Nivel de estudios
2	Id1	80	36.0	Pública	Posgrado
3	Id2	72	35.5	Privada	Grado
4	Id3	68	36.1	Privada	Grado
5	Id4	65	36.3	Pública	Grado

En Excel se pueden identificar cuatro tipos de datos:

- Numéricos: De utilidad para realizar operaciones matemáticas.
- Fecha y hora: Para datos de fecha con diferentes formatos.
- Texto: Compuesto por una cadena de caracteres de tipo alfabético, numérico y caracteres especiales.
- Lógicos: Se utilizan para identificar los valores de verdadero o falso.

Las variables ID, Tipo de institución y Nivel de estudios, correspondientes a las tres primeras columnas (A,B,C), presentan datos de tipo Texto; mientras que las variables Peso y Temperatura, presentes en las dos últimas columnas (D,E), tienen datos numéricos.

1.6 Laboratorio 2. Introducción a jamovi

Jamovi es un herramienta estadística de código abierto construida sobre el lenguaje de programación R, aunque destaca por ofrecer una interfaz gráfica clara y fácil de usar. A su vez, brinda la posibilidad de integrarse con el entorno de R para usuarios más avanzados, lo que la convierte en una opción versátil. Gracias a sus funcionalidades y a su carácter gratuito, se presenta como una alternativa atractiva frente a otros programas estadísticos de pago.

Figura 1.3

Logo de jamovi



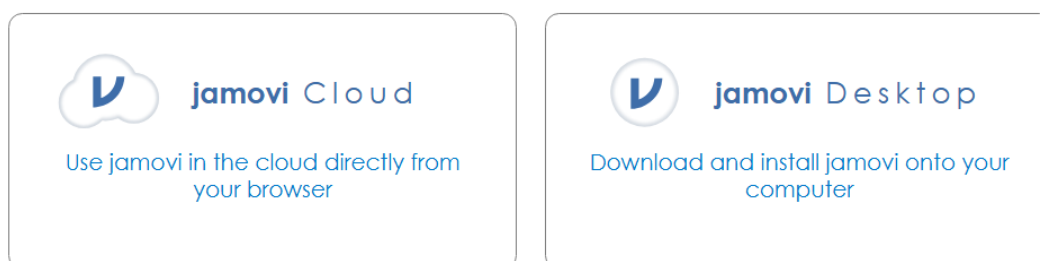
<https://www.jamovi.org/>

1.6.1 Sitio de descarga y registro

En el sitio web oficial de jamovi (<https://www.jamovi.org/>) se observan dos opciones para acceder a jamovi: jamovi cloud o jamovi desktop, como se muestra en la figura 1.4. Luego del registro de una cuenta, se puede usar en línea **jamovi Cloud**, lo que facilita el acceso a los proyectos desde diferentes dispositivos. Por su parte, **jamovi Desktop** permite descargar el software para instalarlo en el computador, con versiones para diferentes sistemas operativos.

Figura 1.4

Portada sitio web de jamovi



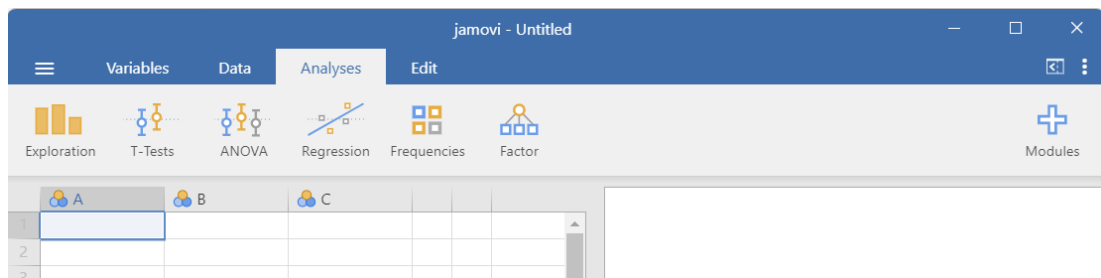
<https://www.jamovi.org/>

En esta edición del libro se utilizará la versión jamovi desktop 2.6.26 SOLID o superior, destinada a la mayoría de los usuarios, sin embargo, se debe considerar los requerimientos de instalación para cada sistema operativo.

Una vez instalado jamovi, al ingresar y se observa la figura 1.5 con dos secciones; la primera sección, ubicada en la parte izquierda, contiene las variables, y en la segunda sección se imprimen los resultados de los análisis. Una vez creada una matriz de datos, esta se guarda en un archivo de extensión ".jmv", el cual contiene los datos y resultados que se hayan generado; pero además, se puede generar un archivo de resultados en formato "pdf" mediante la opción *Exportar*.

Figura 1.5

Pantalla de jamovi



1.6.2 Actualización de la versión de jamovi

La actualización de jamovi consiste en descargar la versión más reciente desde el sitio web oficial e instalarla directamente sobre la versión existente, sin necesidad de desinstalar la versión anterior. Para mantener el software actualizado, se recomienda verificar periódicamente la disponibilidad de nuevas versiones en la página oficial y realizar el proceso de actualización de forma manual.

1.6.3 Menú de jamovi

En la pantalla principal de jamovi, se observa el menú Archivo con el ícono de tres barras horizontales, que permite manipular archivos: crear nuevo, abrir, importar, guardar, guardar como, Exportar. Además, se observa la barra de menús cuyo uso se describe a continuación:

- **Variables:** Crea variables.
- **Data:** Realiza operaciones con variables.
- **Analysis:** Realiza análisis estadísticos.
- **Edit:** Usado para editar el archivo de resultados.

1.6.4 Creación de variables con jamovi

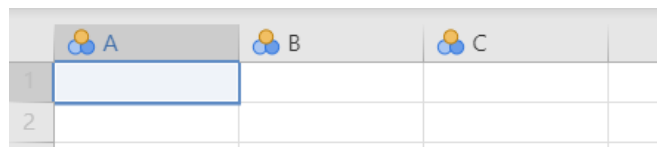
Para crear una nueva variable, se recomienda editar dando un doble clic en la columna correspondiente a las variable de la matriz de datos de la pantalla principal; luego, se agrega la información pertinente:

- **Data variable:** Nombre de la variable sin incluir caracteres especiales.
- **Description:** Descripción de la variable.
- **Measure Type:** Nominal, Ordinal, Continua, ID.
- **Data Type:** Integer, Decimal, Text.
- **Missing values:** Asignación de valores para datos perdidos.

En la pantalla principal de jamovi, se puede dar click en cada columna de la matriz de datos o su vez en el menú **Variables**. Esta matriz tiene tres columnas con los nombres A, B y C asignados por defecto, las cuales pueden ser eliminadas o editadas, tal como se muestra en la figura 1.6

Figura 1.6

Menú variables

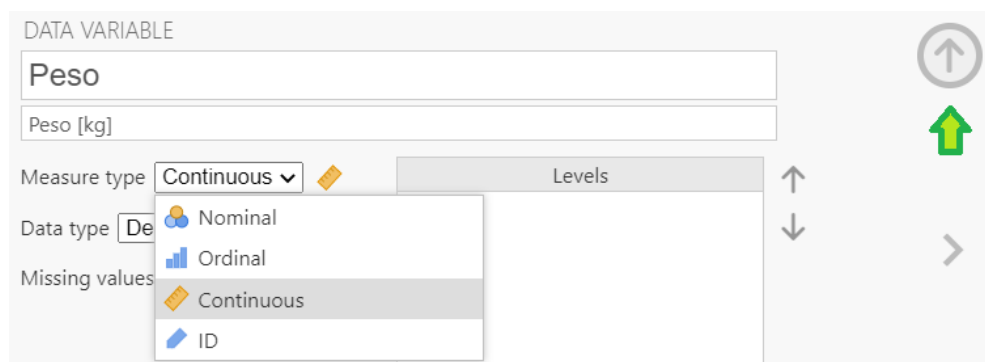


	A	B	C
1			
2			

En la figura 1.7 se ingresa la información de la variable peso, con nombre: Peso, descrita como Peso kg, nivel de medida: continuo, y el tipo de dato: decimal, se completa dando click en la flecha superior derecha de la ventana.

Figura 1.7

Menú Variables



DATA VARIABLE

Peso

Peso [kg]

Measure type: Continuous

Data type: Decimal

Missing values:

Levels:

↑

↓

>

Para el caso de variables cualitativas, además se deben agregar (+) los atributos en el campo **Levels** para agregar los atributos o cualidades de las variables cualitativas correspondientes a preguntas de opción múltiple. Como se puede observar en la definición de las variables en jamovi, no se considera el nivel de medida de intervalo de forma explícita. Por ejemplo, en la figura 1.8 se muestra el detalle de la variable Tipo de institución.

Figura 1.8

Variable cualitativa

DATA VARIABLE

Tipo de institución

Description

Measure type: Nominal

Data type: Text

Missing values:

Levels:

- Pública
- Privada

+

Una vez creadas todas las variables, se procede al ingreso de los datos, obteniéndose una matriz de datos que se visualiza en la figura 1.9. Cada columna tiene un ícono que corresponde al nivel de medida de cada variable.

Figura 1.9

Variable cualitativa

	Variables	Data	Analyses	Edit	
	Paste Clipboard	Edit	Setup	Compute	Transform Variables
	Add	Delete	Filters		
	ID	Peso	Temperat...	Tipo de in...	Nivel de e...
1	Id1	80	36.0	Pública	Posgrado
2	Id2	72	35.5	Privada	Grado
3	Id3	68	36.1	Privada	Grado
4	Id4	65	36.3	Pública	Grado

1.7 Tabla de distribución de frecuencias

Una vez creada la matriz de datos, se puede iniciar el análisis estadístico utilizando tablas de frecuencias.

Una tabla de distribución de frecuencias es la representación de las observaciones individuales o agrupadas en clases con sus respectivas frecuencias. Se usa para variables cualitativas y cuantitativas. Se pueden construir para una y dos variables. Existen algunos tipos de frecuencias que se pueden representar en la tabla.

Una **Frecuencia absoluta** denotada por n_i , representa el número de veces que se repite una observación.

La **frecuencia relativa**, denotada por f_i , es el cociente entre la frecuencia absoluta n_i y el número total de observaciones n , es decir:

$$f_i = \frac{n_i}{n}$$

La acumulación de frecuencias absolutas hasta un valor específico de la variable es la frecuencia absoluta acumulada N_i . La acumulación de frecuencias relativas hasta un valor específico de la variable es la frecuencia relativa acumulada F_i .

Observaciones:

- La suma de frecuencias absolutas es el número total de observaciones n .
- La suma de frecuencias relativas es 1.

1.7.1 Tabla de frecuencias, caso 1

En este caso, se consideran las variables de tipo cualitativo, como por ejemplo, Tipo de institución [Pública, Privada], Nivel educativo [Grado, Posgrado], y las variables numéricas que toman pocos valores, como por ejemplo, Número de hijos por familia [0,1,2,...]. Se representa directamente en la tabla; la primera columna contiene los valores de la variable y las siguientes columnas las respectivas frecuencias.

Ejemplo 1-5. En la Tabla 1.5 se han distribuido las frecuencias, absolutas, relativas y porcentajes para la variable estado civil.

- El 31.82 % son casados y solo el 4.55 % son viudos.
- El 22.73 % son solteros, mientras que el porcentaje restante, es decir el 77.27 %, tienen o han tenido algún tipo de compromiso con otra persona.

Tabla 1.5

Tabla de frecuencias del estado civil

Estado civil	n_i	$f_i = n_i/n$	$f_i \cdot 100 \%$
Soltero	25	0.2273	22.73
Casado	35	0.3182	31.82
Divorciado	15	0.1364	13.64
Viudo	5	0.0455	4.55
Unión de hecho	30	0.2727	27.27
Total: n	110	1.0000	100.00

Ejemplo 1-6. En la Tabla 1.6 se han distribuido las frecuencias absolutas, relativas, relativas acumuladas para una muestra de 36 niños, de los cuales 8 niños son hijos únicos, 5 niños tienen un hermano, 9 niños tienen dos hermanos, 6 niños tienen 3 hermanos y 5 niños tienen 4 o más hermanos.

Tabla 1.6

Frecuencias del número de hermanos

Número de hermanos	n_i	f_i	$F_i \cdot 100 \%$
0	8	0.2222	22.22
1	5	0.1389	36.11
2	12	0.3333	69.44
3	6	0.1667	86.11
4 o más	5	0.1389	100.00
Total	36	1.0000	

Los resultados se pueden interpretar de la siguiente forma:

- Lo común es que las personas tengan 2 hermanos, representados por el 33.33 %.
- Lo menos frecuente es tener un hermano o 4 o más hermanos, representado individualmente por 13.89 % en cada caso.
- La mayoría de las personas tiene un máximo de dos hermanos, representados por el $(22.22+13.89+33.33) \%=66.44 \%$.

1.7.2 Tabla de frecuencias, caso 2

Para las variables cuantitativas que pueden tomar infinitos valores en el conjunto de los números reales, como, por ejemplo, la Edad [años], la Estatura [metros], el Peso [kg], VO_2 máx [ml/kg/min], se recomienda distribuir las observaciones en un número apropiado de intervalos o clases definidos en el conjunto de los números reales y distribuidos en las filas, a partir de los cuales se determinan las frecuencias de cada intervalo. El ejemplo para este caso se realizará con software estadístico.

1.7.3 Tabla de contingencia

La **Tabla de contingencia** es una tabla cruzada entre dos variables que pueden ser cualitativas o cuantitativas de pocos valores (discretas), como resultado se obtienen las frecuencias absolutas también llamados valores observados. Los atributos de la primera variable se representa en las filas y los atributos de la segunda variable en las columnas.

Se pueden representar tablas de contingencia con los valores observados y algunas variantes, tales como la tabla de contingencia con las frecuencias relativas con respecto al total, las frecuencias relativas con respecto al total de las filas y las frecuencias relativas con respecto al total de las columnas; además, puede ser conveniente hallar los respectivos porcentajes.

Cada tipo de tabla cumple una finalidad diferente, por ejemplo, la tabla de contingencia con porcentajes respecto al total de filas sirve para determinar si la variable de fila influye en los atributos de la variable de columna; en otro caso; más adelante se analiza este tipo de tablas en el cálculo de probabilidades.

Ejemplo 1-7. La tabla de contingencia Tabla 1.7 solamente tienen las frecuencias observadas entre Sexo y Deporte de preferencia; los atributos de Sexo se representan en las filas y los atributos de Deporte en las columnas.

- Se aprecia que el grupo de hombres es superior al de las mujeres, así como en la distribución de los resultados en los tres tipos de deportes.
- Tomando como base la Tabla 1.7 se pueden distribuir las frecuencias observadas con respecto al total de personas para determinar de forma conjunta las ocurrencias entre Sexo y Deporte.

Tabla 1.7

Distribución cruzada: Sexo y deporte

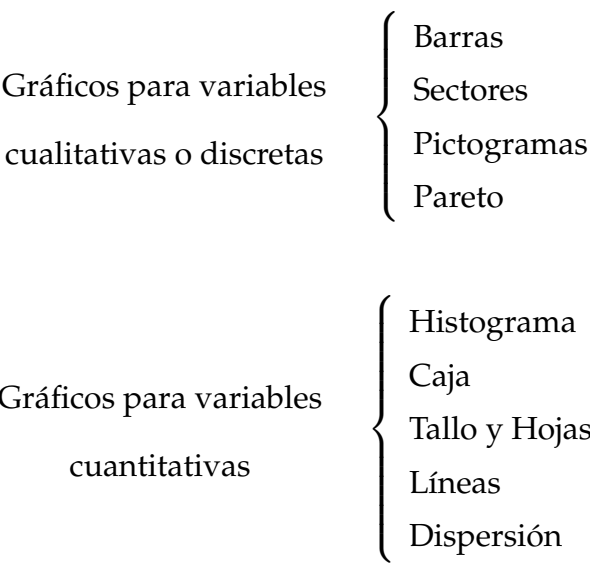
Sexo	Natación	Fútbol	Básquet	Subtotal
Hombre	12	18	10	40
Mujer	10	8	6	24
Total	22	26	16	64

Pero, si el objetivo es determinar la influencia del sexo sobre el deporte de preferencia, se deben hallar los porcentajes con respecto al total de filas, de modo que a cada grupo le corresponda el 100 %, y con ello se pueden realizar comparaciones entre los grupos. Para el caso de los 40 hombres, 12 practican natación, es decir, 0.4 o mas bien 40 %; mientras que en el grupo de 24 mujeres, 10 practican natación, representando el 42 %; es decir, que el sexo no influye en al menos este atributo. De igual forma, se procede con los dos atributos de deporte restantes con respecto a la variable Sexo.

Esta y otras variantes sobre las tablas de contingencia se realizarán con software estadístico en el laboratorio de esta sección.

1.8 Gráficos estadísticos

Los gráficos estadísticos se pueden agrupar por tipo y número de variables.

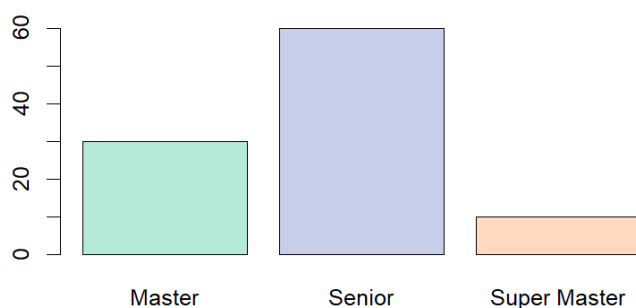


A continuación, se presentan los gráficos para las variables cualitativas o cuantitativas discretas. La figura 1.10 presenta un gráfico de barras, donde las

columnas están separadas y pueden ser de diferentes colores, ya que representan cualidades. Estas cualidades se presentan en el eje horizontal y en el eje vertical se representan las frecuencias. En este ejemplo se visualizan frecuencias absolutas.

Figura 1.10

Gráfico de barras

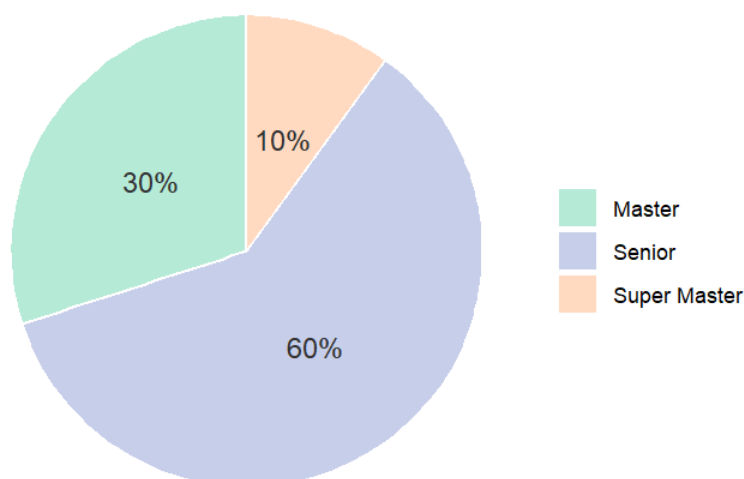


La figura 1.11 presenta un gráfico de sectores, donde los conteos o porcentajes de ocurrencia de las cualidades se distribuyen en los 360 grados de la circunferencia. En este ejemplo se visualizan porcentajes y un cuadro con los atributos.

Figura 1.11

Gráfico de sectores

Distribución por categoría

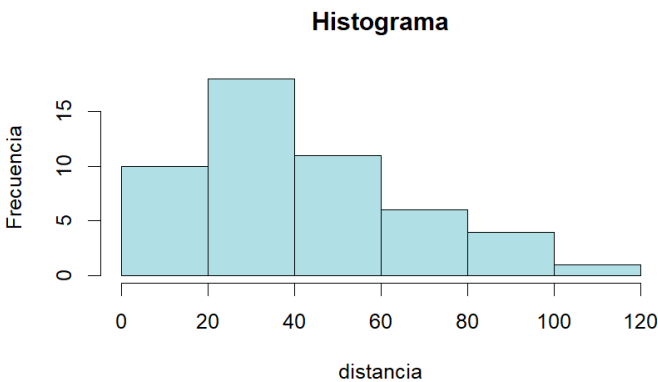


A continuación, se presentan los gráficos para las variables cuantitativas. El histograma de frecuencias que se observa en la figura 1.12 distribuye los intervalos o clases de la variable cuantitativa en el eje horizontal de forma continua

y las frecuencias en el eje vertical.

Figura 1.12

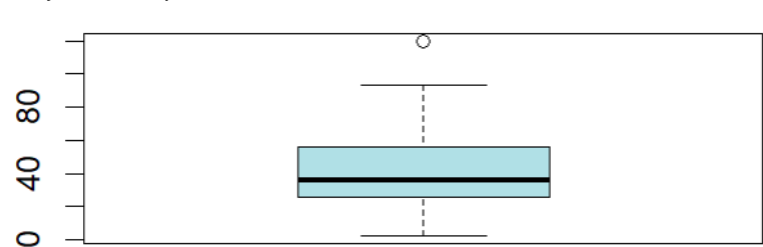
Histograma



El gráfico de cajas es una representación de cinco medidas descriptivas: mínimo, máximo, los tres cuartiles, y se suele agregar los valores atípicos de una distribución de datos. Como se observa en el ejemplo de la figura 1.13, hay un dato atípico.

Figura 1.13

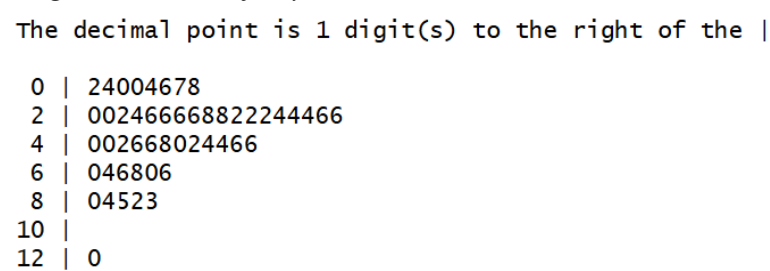
Gráfico de cajas



Un diagrama de tallo y hojas tiene un tallo con los dígitos iniciales (unidades, decenas u otros), y la hoja con los dígitos restantes. En la figura 1.14, hay 7 categorías de 20 unidades cada una.

Figura 1.14

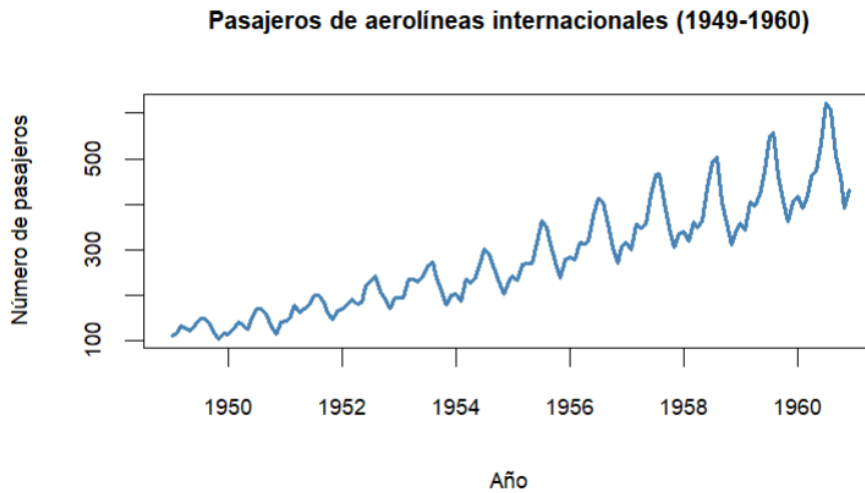
Diagrama de tallo y hojas



El gráfico de líneas puede usarse para representar una variable que depende del tiempo (serie temporal), tal como se muestra en la figura 1.15 .

Figura 1.15

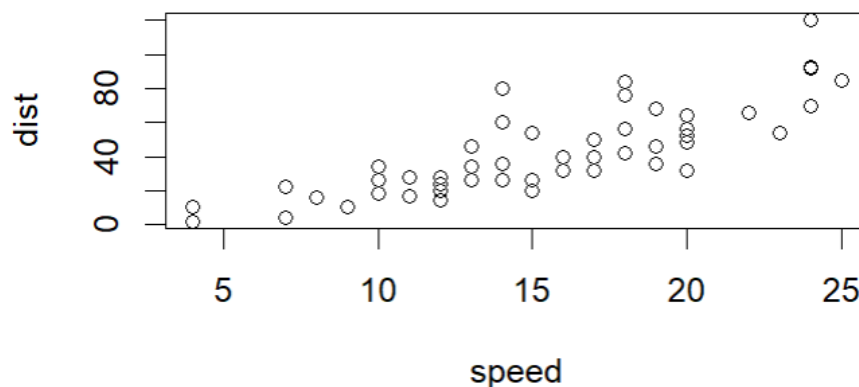
Gráfico de líneas



El gráfico de dispersión en dos dimensiones representa la variable independiente en el eje horizontal y la variable dependiente en el eje vertical, tal como se muestra en la figura 1.16. Se utiliza para estudiar la asociación o grado de relación entre dos variables cuantitativas, pudiendo ser de proporcionalidad directa o inversa según la tendencia creciente o decreciente mostrada en la gráfica.

Figura 1.16

Gráfico de cajas



También se pueden representar gráficos de dispersión tridimensionales, de modo que, en el plano se representan las 2 variables independientes y en el tercer eje, la variable dependiente.

1.9 Laboratorio 3. Tablas y Gráficos

Para los siguientes ejercicios, se usarán los datos del Apéndice A.

1.9.1 Creación de variables en jamovi

Ejemplo 1-8. Crear un archivo con las variables del Anexo A

Opción 1. Resulta conveniente crear el archivo de datos en Excel y luego, en jamovi, elegir la opción Archivo-Importar datos, tras lo cual se personaliza la descripción y etiquetas de las variables según el caso.

Opción 2. Activar el menú Data, para crear las variables: Peso, Sexo, Nivel de actividad Física en jamovi de acuerdo a lo descrito en la sección 1.6.2, obteniendo una matriz de datos como se muestra en la figura 1.17. Estos datos se guardan con un nombre conveniente, con la extensión **omv**; para este ejercicio, el archivo se denominará **PesospruebasFisicas.omv**.

Figura 1.17

Variable cualitativa

	 Peso_lb	 Sexo	 NivelActFi...
1	140	Hombre	Alto
2	153	Hombre	Alto
3	153	Hombre	Bajo
4	154	Hombre	Bajo
5	154	Hombre	Bajo

Observación

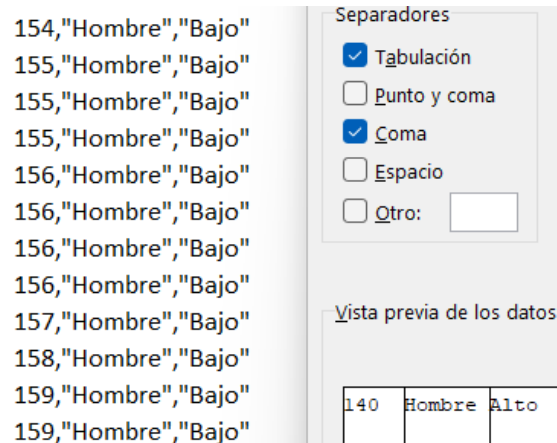
Por el momento, jamovi no tiene la opción de gráfico de sectores. Dicho gráfico se puede realizar con ayuda de Excel, para lo cual se deben copiar los datos de jamovi a Excel, o mas bien, haber tomado la opción 1 de creación del archivo. Una vez que se tiene el archivo en Excel, cuando los datos no están resumidos, resulta conveniente marcar los datos y elegir la opción de tablas y gráficos dinámicos.

Este proceso puede presentar ciertos problemas con los datos que pueden generarse en una sola columna y deben separarse eligiendo Datos/Texto en columnas y seleccionando la opción **delimitados** con el símbolo que corresponda a cada caso. Como se ve en la figura 1.18, el símbolo de separación a elegir es

coma. Luego de esto, los datos se separan en tres columnas y se puede iniciar cualquier análisis.

Figura 1.18

Variable cualitativa



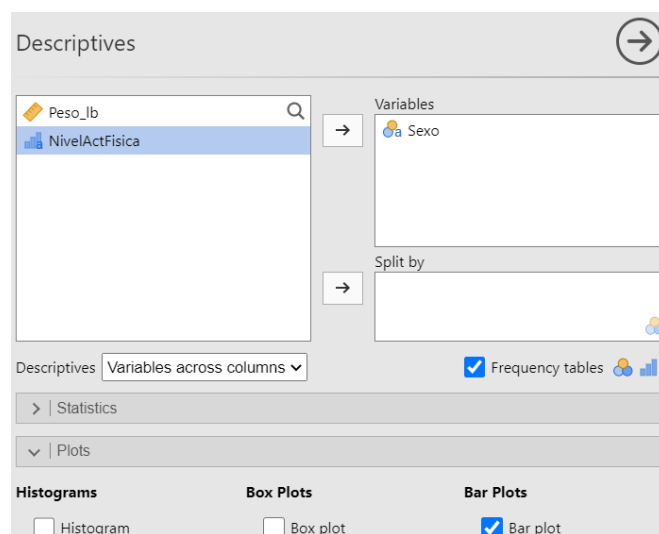
1.9.2 Tablas y gráficos para variables cualitativas o discretas

Ejemplo 1-9. Realizar la tabla de frecuencias y gráfico de barras del sexo con ayuda de jamovi.

Del menú "Explotation", seleccionando **Descriptives** se accede a la ventana de la figura 1.19, elija Statistics y desmarque todas las medidas descriptivas. De las opciones de gráficos (Plot) se elige el gráfico de barras.

Figura 1.19

Menú descriptivos



Los resultados se muestran en la segunda parte de la ventana, la misma que se podría editar antes de exportar en formato pdf y cada vez que se desee, siempre que se hayan guardado los datos en un archivo. Para el caso de la tabla de frecuencias, observe la figura 1.20, que se encuentra en el archivo *Pe-sospruebasFisicas.omv*.

Figura 1.20

Tabla de frecuencias del Sexo

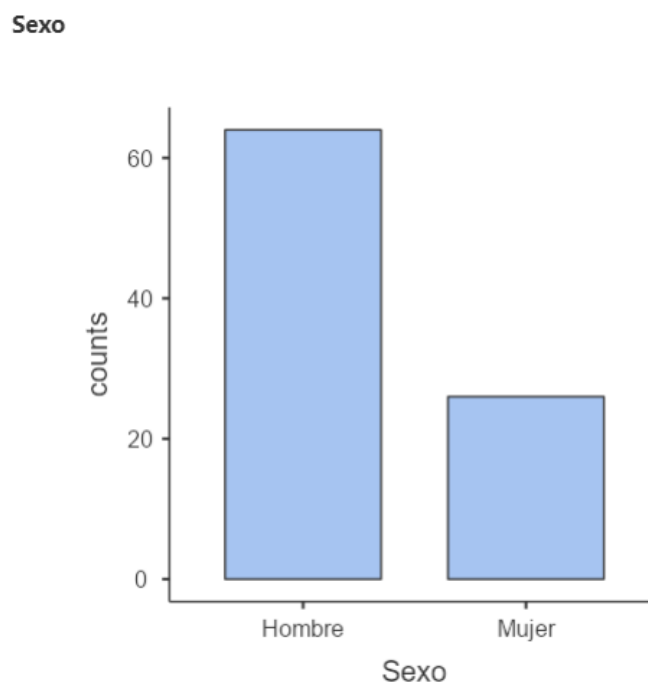
Frequencies

Frequencies of Sexo			
Sexo	Counts	% of Total	Cumulative %
Hombre	64	71.1 %	71.1 %
Mujer	26	28.9 %	100.0 %

En la figura 1.21 se presenta el gráfico de barras para el sexo. En el caso de requerir alguna modificación de colores y etiquetas, entre otros elementos del gráfico, se recomienda integrar el lenguaje de programación R a jamovi.

Figura 1.21

Gráfico de barras de Sexo

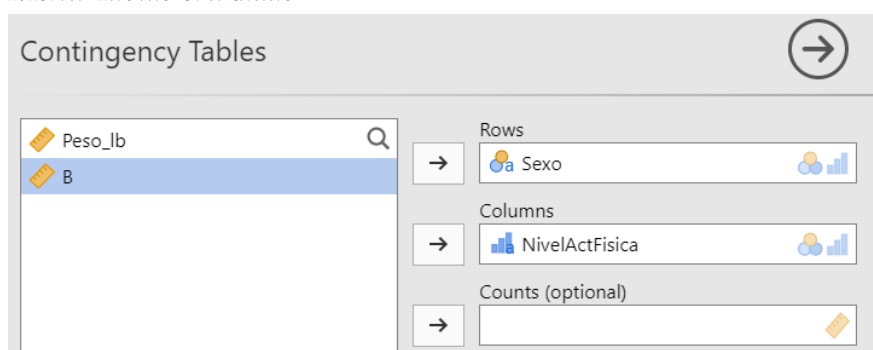


Ejemplo 1-10. Realizar la tabla de contingencia entre Sexo y Nivel de actividad física, usando jamovi

Elegir Analyses, Frequencies, Independent Samples χ^2 test of association y, en la ventana resultante, ingresar la variable Sexo en el campo de las filas y Nivel de actividad física en el campo de las columnas, como se observa en la figura 1.22.

Figura 1.22

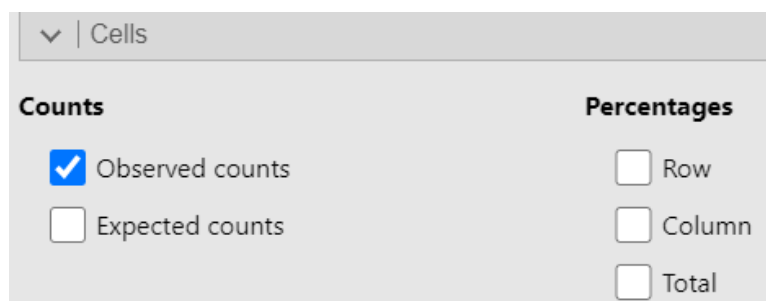
Menu Tablas cruzadas



En la parte inferior del menú se elige el tipo de frecuencias requerido.

Figura 1.23

Sub menu Celdas



a. Presentar las frecuencias observadas (observed counts).

Para este primer caso, en la opción de **Celdas** de la figura 1.23 seleccionar **Total** para obtener la tabla de contingencia representada en la figura 1.24.

Esta tabla permite conocer la distribución de las respuestas en cada cruce de las dos variables. Se deja al lector estudiar el análisis inferencial de estos resultados mediante la prueba Chi cuadrado de independencia.

b. Presentar los porcentajes respecto al total (Total). Nuevamente, desde la opción **Celdas** de la figura 1.23 se desmarcan valores observados en la columna de **porcentajes** y se marca **Total**.

Figura 1.24

Tabla cruzada de valores observados

Contingency Tables

Sexo	NivelActFisica			Total
	Alto	Bajo	Medio	
Hombre	2	38	24	64
Mujer	8	3	15	26
Total	10	41	39	90

Esta tabla permite conocer la distribución de los porcentajes de respuestas en cada cruce de las dos variables con respecto al 100 % total.

Figura 1.25

Tabla cruzada con porcentajes

Contingency Tables

Sexo		NivelActFisica			Total
		Alto	Bajo	Medio	
Hombre	% of total	2.2 %	42.2 %	26.7 %	71.1 %
Mujer	% of total	8.9 %	3.3 %	16.7 %	28.9 %
Total	% of total	11.1 %	45.6 %	43.3 %	100.0 %

En este caso, el 2.2 % de los hombres realiza actividad física a niveles altos, mientras que el 42.2 % la realiza a niveles bajos.

c. Presentar los porcentajes con respecto al total de filas (Rows). Ahora, en la figura 1.23 en la columna de porcentajes, marque **Fila o Row** dependiendo del idioma que use en jamovi, según se observa en la figura 1.26

Figura 1.26

Menu Tablas

Cells

Counts

☐ Observed counts

☐ Expected counts

Percentages

☒ Row

☐ Column

☐ Total

Y el resultado final es la figura 1.27 . En este caso, se observan diferencias porcentuales importantes en el nivel de actividad física entre hombres y mujeres.

Figura 1.27
Tabla cruzada con porcentajes

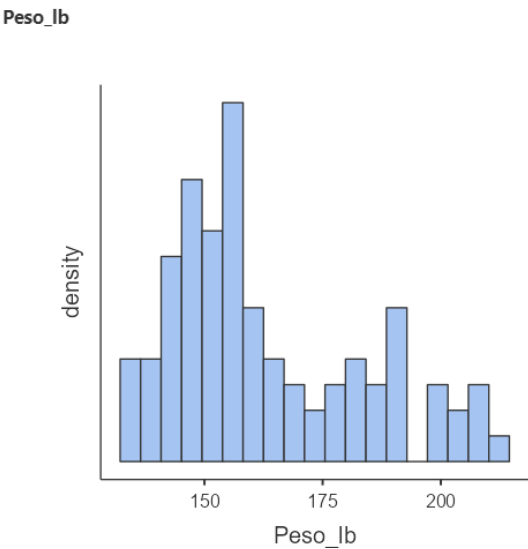
Sexo		NivelActFisica			Total
		Alto	Bajo	Medio	
Hombre	% within row	3.1 %	59.4 %	37.5 %	100.0 %
Mujer	% within row	30.8 %	11.5 %	57.7 %	100.0 %
Total	% within row	11.1 %	45.6 %	43.3 %	100.0 %

Ejemplo 1-11. Realizar el histograma de frecuencias para el peso en jamovi.

El número de clases o intervalos del histograma depende del algoritmo o las fórmulas utilizadas; por lo tanto, los resultados varían entre una aplicación y otra.

La figura 1.28 muestra una distribución del peso en 19 clases o intervalos que van desde aproximadamente 125 a 225 libras. Este gráfico se ha obtenido desde el menú **Exploration** eligiendo **textbfDescriptives** en la sección **Plot** y seleccionando **Histograma** ingrese la variable **Peso** en el campo de variables.

Figura 1.28
Histograma del peso



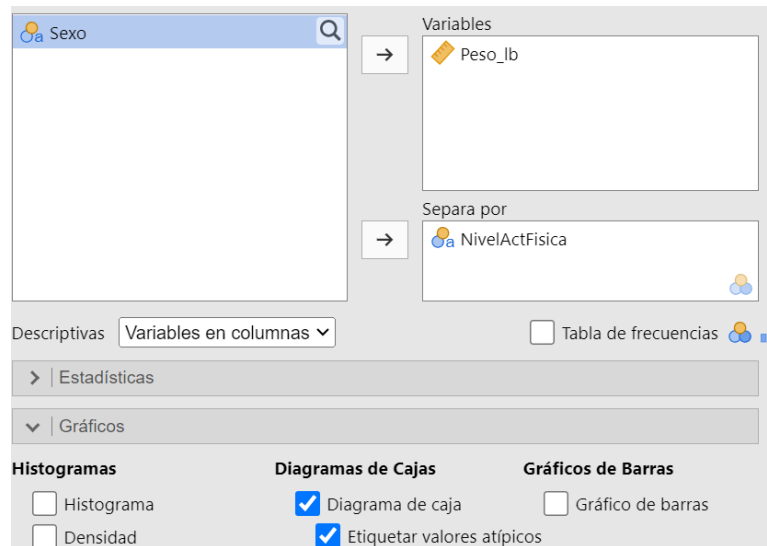
Se observa una mayor acumulación de las observaciones hasta 190 libras, luego, un menor número de personas con pesos superiores a 190 libras.

Ejemplo 1-12. Realizar el gráfico de cajas del peso en función del nivel de actividad física en jamovi.

Desde el menú de exploración y descriptivas, ingrese la variable peso y el nivel de actividad física para agrupar, como se observa en la figura 1.29

Figura 1.29

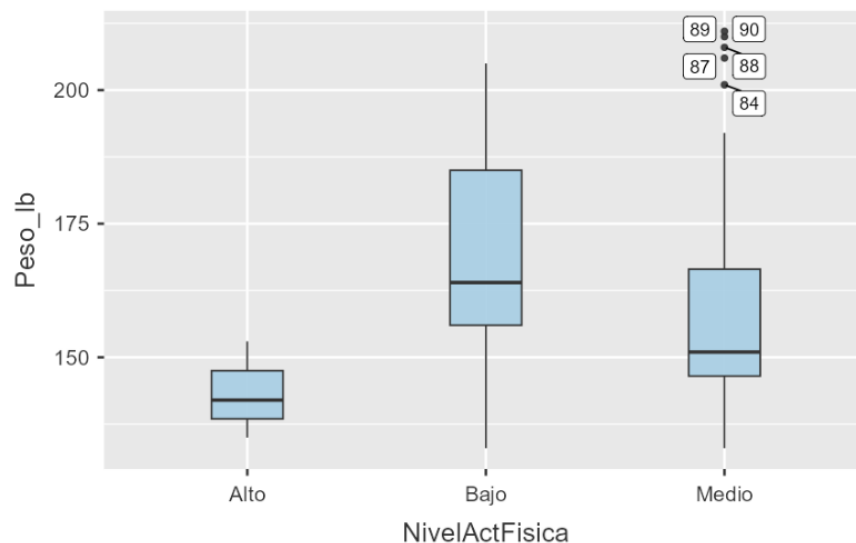
Histograma del peso



La figura 1.30 muestra que, a menor actividad física, mayor es el peso; en el nivel medio de actividad física se presentan pesos extremos entre 190 y 225 libras, identificados por su número de observación: 84, 87, 89 y 90.

Figura 1.30

Histograma del peso

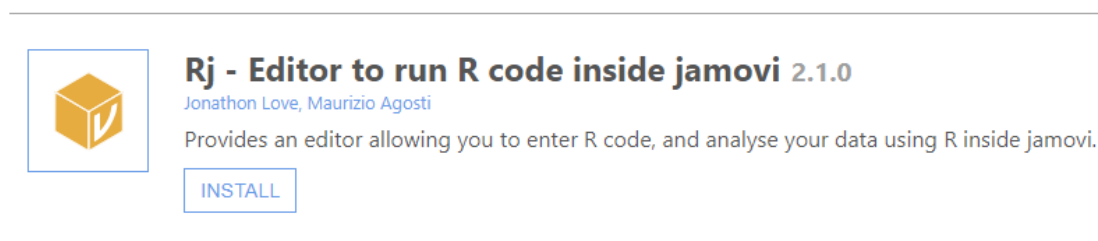


1.9.3 Integración de R en jamovi

En la parte superior de la pantalla principal de jamovi se encuentra un símbolo + con la etiqueta "Módulos". Al dar clic, se accede a las opciones de integración de jamovi a R, que se observan en la figura 1.31 de las cuales se elige la opción

Figura 1.31

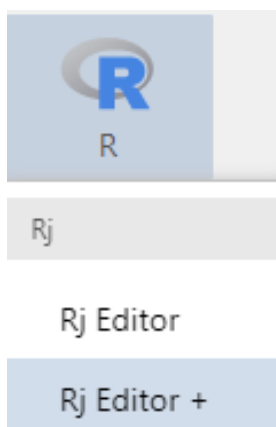
Opciones de instalación de R



Una vez instalado **R**, se elige la opción **Rj** para lograr una integración del código de una forma más sencilla de usar. Pero se debe tener claro que todo lo que se ejecute en dicha sección es mediante códigos de R.

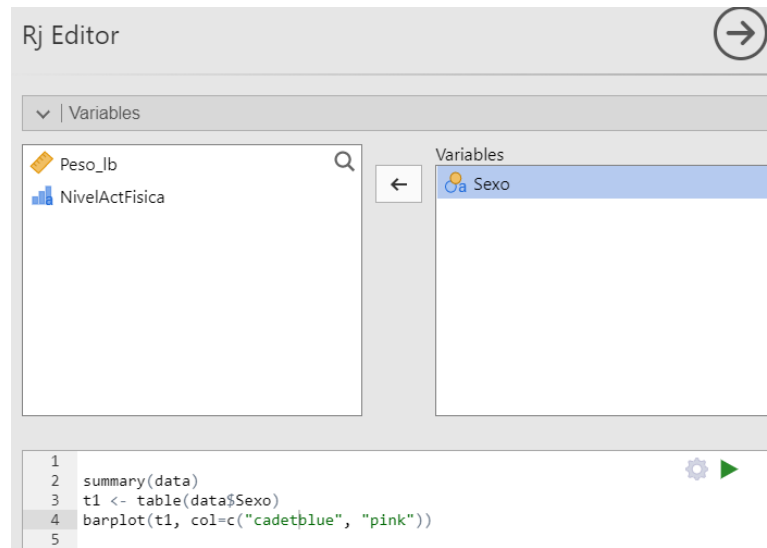
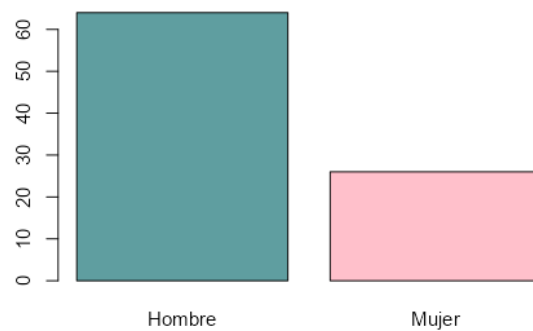
Figura 1.32

Rj Editor



Ejemplo 1-13. Modificar los colores del gráfico de barras.

Como se puede observar en la figura 1.33 en la ventana inferior se escribe el código de R, con la sugerencia del editor al momento de escribirlo. La modificación que se ha realizado se muestra en la figura 1.34, donde se han elegido dos colores diferentes para resaltar visualmente la diferencia entre los resultados.

Figura 1.33*Menú integrado de R y jamovi)***Figura 1.34***Gráfico personalizado***Ejemplo 1-14.** Agregar el gráfico de sectores.

Se agrega el código correspondiente y en los comentarios se antecede el símbolo hashtag. Como se pudo observar en el ejemplo anterior, para utilizar las variables de un conjunto de datos fue necesario introducir este formato: nombre del conjunto de datos, el símbolo de dólares y el nombre de la variable; el código para el gráfico de sectores es `pie`, y los colores se escriben en inglés entre comillas dentro de la función `col=c("color1", "color2")`. Se han agregado porcentajes, realizando las operaciones correspondientes en la tabla.

El código se encuentra en la figura 1.35, y el resultado final está en la figura 1.36. Finalmente, se observa que el 71 % son hombres y solo el 29 % son mujeres.

Figura 1.35

Código final integrado

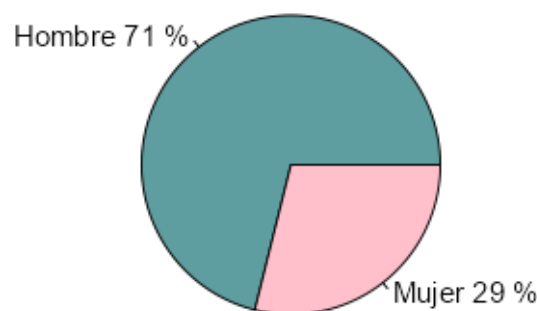
```

1
2 summary(data)
3
4 # Tabla de frecuencias de la variable cualitativa Sexo
5 t1 <- table(data$Sexo)
6 barplot(t1, col=c("cadetblue", "pink"))
7 pie(t1, col=c("cadetblue", "pink"))
8 # Calcular porcentajes
9 porcentajes <- round(t1 / sum(t1) * 100)
10 # Crear etiquetas: nombre + porcentaje
11 etiquetas <- paste(names(t1), porcentajes, "%")
12 # Graficar pastel con colores y etiquetas
13 pie(t1, labels = etiquetas, col = c("cadetblue", "pink"),
14     main = "Distribución por Sexo")
15

```

Figura 1.36

Gráfico de sectores de Sexo

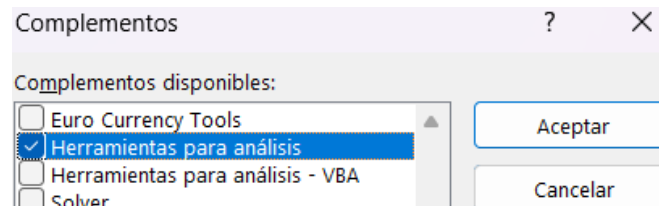


1.9.4 Tablas y gráficos para variables continuas con Excel

Ejemplo 1-15. Instalar el complemento Análisis de datos

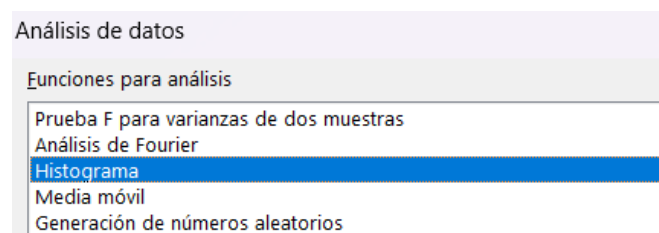
El complemento **Análisis de datos de Excel** proporciona un proceso automatizado y rápido para obtener la tabla de frecuencias y el respectivo histograma de frecuencias con un número conveniente de clases o intervalos, así como otros análisis descriptivos e inferenciales.

Para instalar este complemento, se deben seleccionar los siguientes elementos del menú Archivo: Opciones, Complementos, Complementos de Excel, Ir, Herramientas para el análisis (Marcar). En la figura 1.37 se observa la ventana con el último paso del proceso de instalación. Una vez instalado el complemento, estará disponible en el menú DATOS.

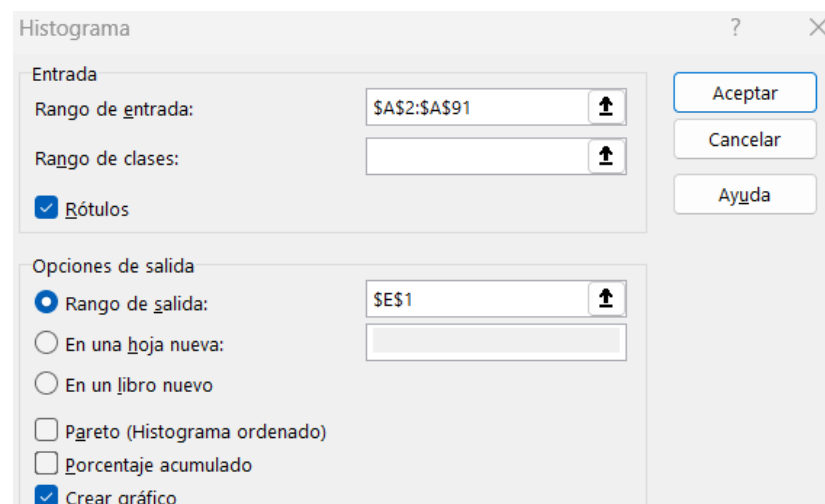
Figura 1.37*Complementos disponibles*

Ejemplo 1-16. Realizar la tabla de frecuencias distribuidas en clase y el histograma de frecuencias para la variable peso

En el menú Datos, selecciona: Análisis de datos, Histograma, Aceptar; tal como se observa en la figura 1.38

Figura 1.38*Opciones Análisis de datos*

Según la figura 1.39, elegir el rango de entrada (pesos), marcar la opción rótulos, el lugar para la salida de los resultados, seleccionar histograma y aceptar.

Figura 1.39*Menú Histograma*

Se obtiene una tabla y un histograma de frecuencias con un número apropiado de clases o intervalos. La tabla de frecuencias que se observa en la figura 1.40 está organizada en 10 clases o intervalos. La primera clase contiene los pesos menores o iguales a 133 libras, la última clase contiene los pesos mayores que 203.33 libras; las demás clases son intervalos semiabiertos, siendo abiertos a la izquierda y cerrados a la derecha; por ejemplo, la segunda clase contiene los pesos mayores que 133 libras pero menores o iguales que 141.67 libras.

Figura 1.40

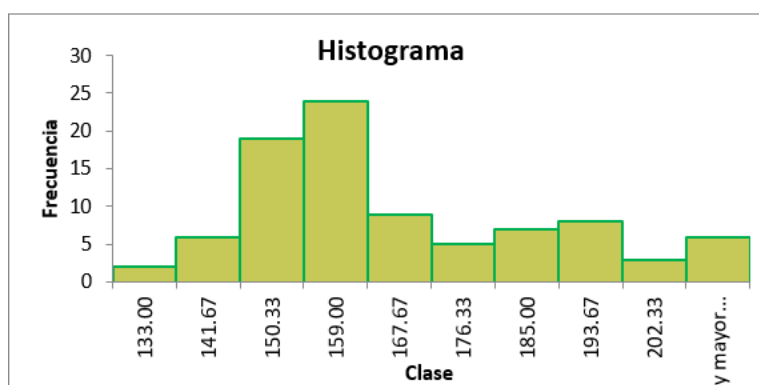
Tabla frecuencias

<i>Clase</i>	<i>Frecuencia</i>
133.00	2
141.67	6
150.33	19
159.00	24
167.67	9
176.33	5
185.00	7
193.67	8
202.33	3
y mayor...	6

Se observa una mayor acumulación de observaciones en la tercera y cuarta clase; después de éstas, las frecuencias disminuyen considerablemente, siendo un indicador de sesgo hacia la derecha; lo que se confirma con el histograma de la figura 1.41, el cual se ha personalizado de modo que el ancho de separación entre columnas sea del 0 % entre otros detalles.

Figura 1.41

Histograma de frecuencias generado por Excel



Es importante aclarar que aplicaciones como Excel, jamovi y otras no utilizan el mismo criterio para determinar el número de clases.

1.10 Medidas descriptivas

Las medidas descriptivas son herramientas estadísticas que permiten sintetizar un gran número de datos en una sola medida para tener una idea cuantitativa aproximada de toda la distribución de las observaciones. A continuación, se desarrollan las medidas descriptivas más representativas.

1.10.1 Medidas de tendencia central

Estas medidas tienden a situarse en el centro del conjunto de datos. Las más importantes son: la media, la mediana y la moda.

Media. Se define como el promedio de las observaciones.

Mediana. Medida que se ubica en el centro de la distribución de los datos ordenados.

Moda. Es la observación de mayor frecuencia.

Las fórmulas correspondientes a las medidas de tendencia central son:

Media poblacional	$\mu = \frac{\sum_{i=1}^m x_i n_i}{N}$
-------------------	--

Media muestral	$\bar{x} = \frac{\sum_{i=1}^m x_i n_i}{n}$
----------------	--

Mediana (n impar)	$\text{Me} = X_{\frac{n+1}{2}}$
-------------------	---------------------------------

Mediana (n par)	$\text{Me} = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$
-----------------	---

Observaciones: n es el tamaño de la muestra, n_i es la frecuencia absoluta y m es el número de observaciones diferentes, observaciones que se aplican en las medidas descriptivas restantes.

1.10.2 Medidas de dispersión

Rango. Determina la amplitud de variación de los datos.

Varianza. Es el promedio de los cuadrados de las diferencias de los datos con respecto a la media.

Desviación estándar. Es la raíz cuadrada de la varianza.

Coficiente de variación. Es el cociente entre la desviación estándar y la media.

Rango	$\text{Rango} = X_{max} - X_{min}$
Varianza poblacional	$\sigma^2 = \frac{\sum_{i=1}^m (x_i - \mu)^2 n_i}{N}$
Varianza muestral	$s^2 = \frac{\sum_{i=1}^m (x_i - \bar{x})^2 n_i}{n - 1}$
Desviación estándar poblacional	$\sigma = \sqrt{\sigma^2}$
Desviación estándar muestral	$s = \sqrt{s^2}$
Coeficiente de variación	$cv = \frac{s}{\bar{x}}$

1.10.3 Medidas de posición

Las medidas de posición delimitan el porcentaje de observaciones que son menores o iguales que esa medida o la vez, delimitan el porcentaje de observaciones que son superiores a esa medida.

Existen medidas específicas muy utilizadas, como son los cuartiles, que dividen el conjunto de datos en cuatro partes iguales; así como también, existen los deciles. En general, se tienen los percentiles, que dividen el conjunto de datos en cien partes iguales. Para hallar los percentiles se deben ordenar los datos y se hallan mediante la siguiente expresión:

$$P_p \% = X_{(n+1)p}$$

En la expresión anterior, p es un número que toma valores entre cero y uno, ya que es una proporción o frecuencia relativa necesaria para hallar el porcentaje. Por ejemplo, para hallar el percentil 5 %, se tiene $p = 0.05$. El término $(n + 1)p$ es la posición del percentil; si resulta en un número entero, esa es la posición del percentil y el percentil será el valor correspondiente; si este número es un decimal terminado en .5, se tienen dos posiciones para el percentil, alrededor de ese número, por lo que el percentil será el promedio de los valores correspondientes a dichas posiciones. Si dicho número toma cualquier otro valor, se redondea y el percentil corresponderá a dicha posición.

1.10.4 Medidas de forma

Coefficiente de asimetría. Determina la forma en cuanto al sesgo tomando valores alrededor del cero. Se presentan tres formas: sesgo a la izquierda ($\alpha_3 < 0$), simétrica ($\alpha_3 = 0$) y sesgo a la derecha ($\alpha_3 > 0$).

El coeficiente de asimetría para la muestra es:

$$\alpha_3 = \frac{\sum_{i=1}^m (x_i - \bar{x})^3 n_i}{ns^3}$$

Coefficiente de curtosis. Determina la forma en cuanto a la puntiagudez y toma valores estandarizados alrededor del cero. Se presentan tres formas: platicúrtica ($\alpha_4 < 0$), mesocúrtica ($\alpha_4 = 0$) y leptocúrtica ($\alpha_4 > 0$).

El coeficiente de curtosis para la muestra es:

$$\alpha_4 = \frac{\sum_{i=1}^m (x_i - \bar{x})^4 n_i}{ns^4}$$

1.11 Ejemplos de reforzamiento

A continuación, se presentan ejemplos de reforzamiento y aplicación de las medidas descriptivas mediante el uso de las fórmulas y calculadora.

Ejemplo 1-17. Hallar todas las medidas descriptivas para el peso [kg], usando los 16 primeros datos de peso del conjunto de datos pesos de apéndice A, y considerarlos como una muestra.

a. Medidas de tendencia central

Media:

$$\begin{aligned}\bar{x} &= \frac{140 + 153(2) + 154(2) + 155(3) + 156(4) + 157 + 158 + 159(2)}{16} = \\ &= 154.75 \text{ libras}\end{aligned}$$

La mediana se halla ordenando los datos y, en este caso, se usa la fórmula para n par:

$$\text{Me} = \frac{X_8 + X_9}{2} = \frac{155 + 156}{2} = 155.5 \text{ libras}$$

La moda es única, con un valor de 156 libras. Las tres medidas de tendencia central son relativamente cercanas, lo que podría indicar baja variabilidad y una posible simetría en la distribución de los datos.

b. Medidas de dispersión

El peso mínimo es 140 libras, el peso máximo es 159 libras, por lo tanto, el rango es: $\text{Rango} = 159 - 140 = 19$ libras

Varianza:

$$s^2 = \frac{(140 - 154.75)^2 + \dots + (158 - 154.75)^2 + (140 - 154.75)^2}{15} = 18.87 \text{ [libras]}^2$$

Desviación estándar:

$$s = \sqrt{18.87} = 4.34 \text{ libras}$$

Coefficiente de asimetría:

$$cv = \frac{4.34}{154.75} = 0.0280$$

Este coeficiente se puede expresar en términos de porcentaje solamente al multiplicar por 100 %.

El rango de variación es de 19 libras, y al hallar el coeficiente de variación, que relaciona la desviación estándar con la media, se confirma una baja variabilidad correspondiente a un 2 %.

c. Medidas de posición

Calcular el percentil 10 %

$$P_{10\%} = X_{(17)0.1} = X_{1.7} = X_2 = 153 \text{ kg}$$

Este resultado, indica que solamente un 10 % de las personas tienen pesos menores o iguales a 153 libras. Es importante resaltar que para una interpretación específica tipo semáforo, se debe conocer el peso recomendable para el grupo de interés.

Calcular el cuartil 2

El cuartil 2 equivale al percentil 50 %

$$P_{50\%} = X_{(17)0.5} = X_{8.5} = \frac{X_8 + X_9}{2} \\ P_{50\%} = 156 \text{ kg}$$

Como se observa, este resultado coincide con la mediana.

El 50 % de las personas tiene pesos menores o iguales a 156 libras, y a la vez tiene pesos superiores a 156 libras.

d. Medidas de forma

En este caso, usaremos las funciones coeficiente de asimetría y kurtosis de Excel para hallar estas medidas, obteniendo los siguientes resultados:

Coeficiente de asimetría

$$\alpha_3 = -2.8017$$

El sesgo es pronunciado hacia el lado izquierdo.

Coeficiente de curtosis

$$\alpha_4 = 9.7694$$

La distribución presenta un pico muy alto.

e. Uso de la calculadora para verificar los resultados.

Actualmente, la mayoría de las calculadoras tienen el menú estadístico, que permite hallar la mayoría de las medidas descriptivas de una forma sencilla y rápida. Consulte el procedimiento que corresponda al modelo de su calculadora y compare los resultados.

Ejemplo 1-18. Hallar las medidas descriptivas del peso de los estudiantes hombres. Sus pesos en kilogramos son: 54, 53, 75, 50, 54, 68, 58

a. Medidas de Tendencia Central

Se halla la media aplicando la fórmula.

$$\bar{x} = \frac{\sum X_i}{n} = \frac{412}{7} = 58.8$$

Para hallar la Mediana, se ordenan los pesos de menor a mayor: 50, 53, 54, 54, 58, 68, 75. Se aplica la fórmula para n impar:

$$X_{\frac{(n+1)}{2}} = X_{\frac{8}{2}}$$

$$X_4 = 54 \text{ kg}$$

b. Medidas de Dispersión

Rango:

$$Rango = X_{max} - X_{min} = 75 - 50 = 25$$

Varianza:

$$s^2 = \frac{\sum_{i=1}^m (x_i - \bar{x})^2 * n_i}{n - 1} = \frac{\sum_{i=1}^m (x_i - 58.8)^2 * n_i}{7 - 1} = 84.14$$

Desviación estándar:

$$s = \sqrt{s^2} = 9.17$$

Coeficiente de variación:

$$cv = \frac{s}{\bar{x}} = \frac{9.17}{58.86} = 0.16$$

Luego= $cv * 100\%$, es decir, $cv = 16\%$, lo que indica una dispersión relativamente baja.

c. Medidas de Posición

Percentil 5 %

$$P_5\% = X_{(7+1)*0.05} = X_{0.4} = X_0 < 50kg$$

El 5 % de los hombres pesa menos de 50 kilogramos.

d. Medidas de Forma

Coeficiente de asimetría

$$\alpha_3 = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{X}}{s} \right)^3 * n_i$$
$$\alpha_{3H} = -0.8958$$

⇒ La distribución es ligeramente sesgada hacia la izquierda

Coeficiente de curtosis:

$$\alpha_4 = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^m \left(\frac{x_i - \bar{X}}{s} \right)^4 * n_i$$
$$\alpha_{4H} = 0.5405 \times 5.018 = 5.5589$$

La distribución presenta un pico muy alto.

1.12 Laboratorio de medidas descriptivas

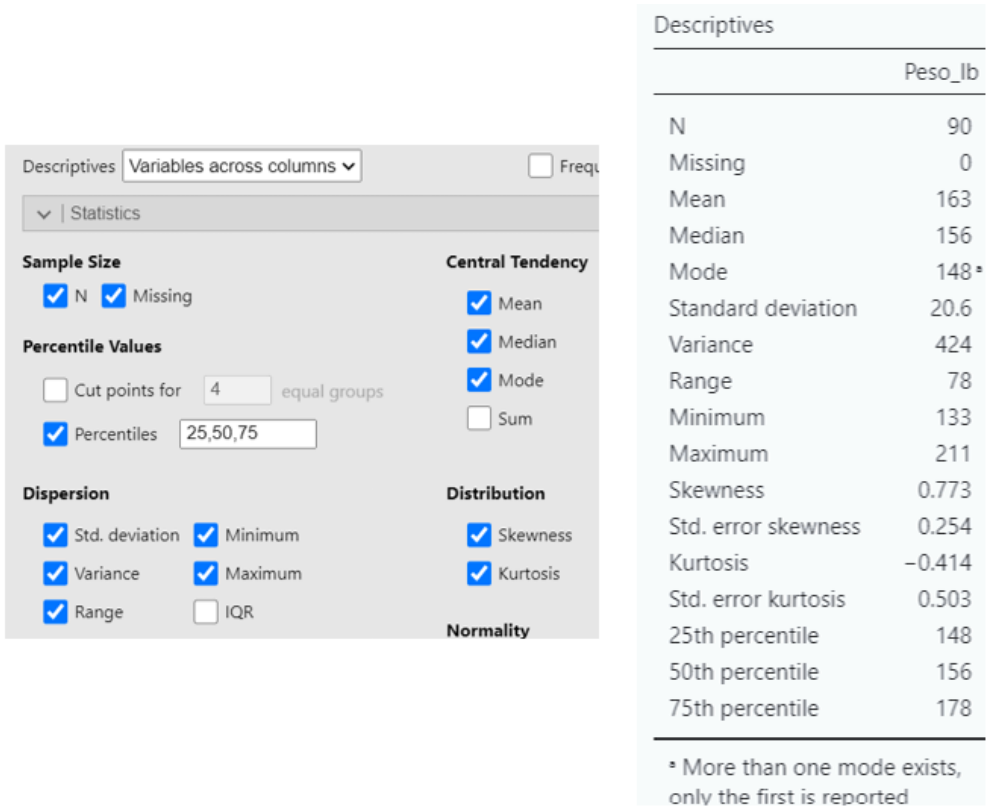
1.12.1 Medidas descriptivas con jamovi

El menú de medidas descriptivas incluye todas las medidas que se definieron anteriormente, excepto el coeficiente de variación.

Ejemplo 1-19. Obtener todas las medidas descriptivas de la variable Peso en libras del conjunto de 90 datos del Apéndice A

Desde el menú **Analyses**, seleccione **Exploration**, elija la opción **Descriptives** y en las opciones ingrese la columna correspondiente al peso, en el campo de las variables; luego, en la opción **Statistics** marque las opciones según la (figura 1.42) que genera el resumen de medidas descriptivas.

Figura 1.42
Menú de opciones y Resultados



La media, mediana y moda son, respectivamente, 163, 156, 148 y a pesar de que el rango de variación es de 78 libras, no es alto debido a la magnitud de los datos. El coeficiente de variación es $cv=20.6/164 = 0.1256$ representa una variabilidad moderada-baja de 12.56 %.

Ejemplo 1-20. Actualizar la versión de jamovi

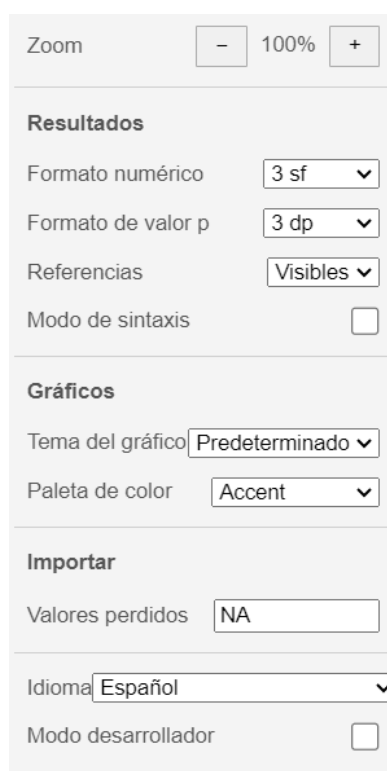
Se recomienda desinstalar la versión anterior antes de instalar la nueva versión de jamovi, así como el módulo Rj, en caso de ser necesario.

Ejemplo 1-21. Personalizar la configuración de jamovi

a. Idioma. En todas las versiones de jamovi se puede cambiar el idioma desde el menú de configuración disponible en la parte superior derecha (tres puntos verticales). Al hacer clic derecho, tenemos las opciones a modificar, tales como el idioma y la sintaxis, entre otras, como se observa en la figura 1.43

Figura 1.43

Configuración



La figura 1.44 visualiza el menú con el cambio a idioma español.

Figura 1.44

Menú en español



b. Sintaxis. Si al realizar un análisis estadístico se requieren los resultados y el código de R, en la configuración de la figura 1.43, se puede marcar el campo **Modo de sintaxis**.

Por ejemplo, para los datos de Temperatura de la Tabla 1.4, se ha agregado el Modo sintaxis, generando el código y los resultados como se observa en la figura 1.45, donde además se observa el cambio en el idioma en los resultados. Como ya se hizo anteriormente, se pueden agregar, de forma opcional, más

Figura 1.45

Sintaxis y resultados

Descriptivas	
<pre>jmv::descriptives(data = data, vars = Temperatura)</pre>	
Descriptivas	
	Temperatura
N	4
Perdidos	0
Media	36.0
Mediana	36.0
Desviación estándar	0.340
Mínimo	35.5
Máximo	36.3

resultados con códigos de R, activando el modulo Rj instalado.

Observaciones

Si un archivo de jamovi fue creado en otra versión, este se puede usar mientras tenga la extensión omv. Si un archivo contiene resultados en cierto idioma, y se requiere un idioma diferente, se debe configurar el nuevo idioma y repetir el análisis correspondiente.

1.12.2 Medidas descriptivas con Excel

Ejemplo 1-22. Con los datos de peso y estatura de la siguiente muestra, obtenga el IMC, y las medidas descriptivas del IMC. Use el menú análisis de datos de Excel.

Primero, se debe hallar el IMC en la columna C; luego, seleccionar: **Datos**, y en la parte superior derecha **Análisis de datos**, ingresar el IMC y demás opciones de acuerdo a la figura 1.46.

Figura 1.46

Datos y opciones

Peso_Kg	Estatura_m	IMC
61.36	1.60	23.97
61.82	1.55	25.73
62.73	1.54	26.45
63.64	1.58	25.49
65.45	1.56	26.89
66.36	1.58	26.58
67.27	1.50	29.90
67.73	1.49	30.51
60.45	1.51	26.51
70.45	1.55	29.32
71.82	1.68	25.45
60.45	1.50	26.87

Estadística descriptiva

Entrada

Rango de entrada: \$C\$1:\$C\$13

Agrupado por: ☒ Columnas ☐ Filas

☒ Rótulos en la primera fila

Opciones de salida

☒ Rango de salida: \$E\$1

☐ En una hoja nueva:

☐ En un libro nuevo

☒ Resumen de estadísticas

En la figura 1.47 se presentan los resultados:

Figura 1.47

Descriptivos IMC

IMC = $\text{Peso} / \text{Estatura}^2$	
Media	26.97
Error típico	0.57
Mediana	26.55
Moda	
Desviación estándar	1.96
Varianza de la muestra	3.85
Curtosis	-0.26
Coefficiente de asimetría	0.67
Rango	6.54
Mínimo	23.97
Máximo	30.51
Suma	323.67
Cuenta	12

1.13 Ejercicios propuestos del capítulo 1

1.1. Identifique el tipo de variable (cuantitativa o cualitativa), el nivel de medida (nominal, ordinal, escala) para las siguientes variables de educación inicial.

Variable	Tipo	Nivel
Sector de residencia del niño (Sur, Centro, Norte)	_____	_____
Tiempo de sueño por noche (horas)	_____	_____
Edad del niño (meses)	_____	_____
Idioma materno del niño (Español, Otros)	_____	_____
Nivel del lenguaje (0. no verbal, ...,)	_____	_____
Puntaje en desarrollo motor (Retraso, Normal)	_____	_____
Género del niño	_____	_____
Nivel socioeconómico del hogar	_____	_____
Tiempo de siesta (min)	_____	_____
Número de hermanos	_____	_____
Color favorito del niño	_____	_____
Tiempo de reacción (milisegundos)	_____	_____

1.2. Identifique el tipo de variable (cuantitativa o cualitativa), el nivel de medida (nominal, ordinal, escala) para las siguientes variables de actividad física.

Variable	Tipo	Nivel
Peso levantado en halterofilia (kg)	_____	_____
Nivel de fatiga (bajo, medio, alto)	_____	_____
Frecuencia cardíaca en reposo (lpm)	_____	_____
Superficie de entrenamiento (césped, arena, tartán)	_____	_____
Porcentaje de grasa corporal	_____	_____
VO ₂ máx (ml/kg/min)	_____	_____
Sexo biológico del deportista	_____	_____
Número de abdominales por minuto	_____	_____
Grupo etario (Sub-14, Sub-16, Sénior)	_____	_____
IMC del deportista (kg/m ²)	_____	_____
Duración del entrenamiento diario (minutos)	_____	_____
Estado de hidratación (deshid, normal, hiperh.)	_____	_____

1.3. Hallar las medidas de tendencia central y dispersión para la variable calificación en la primera unidad de la asignatura de Pedagogía en el periodo académico actual. Las calificaciones de una muestra de estudiantes son: 14, 14, 14, 14, 15, 15, 15, 16, 16, 16, 16, 16, 17, 17, 17, 18, 18, 18

1.4. Hallar las medidas de tendencia central y dispersión para la variable puntuación en la prueba de CI para un grupo de 20 niños de 5 años, con las siguientes observaciones del CI: 89, 98, 96, 102, 94, 106, 121, 85, 115, 114, 97, 84, 90, 85, 93, 109, 96, 100, 113, 100.

1.5. Hallar las medidas de tendencia central y dispersión para la variable evaluación de la motricidad gruesa. Los datos se han obtenido con el Test de Desarrollo Motor Grueso-2 (TGMD-2) estandarizado, aplicado a niños de 3 a 10 años. Las puntuaciones son: 124, 77, 82, 93, 88, 92, 117, 99, 102, 115, 121, 119, 86, 107, 98, 118, 99, 115, 105, 100, 99, 96, 102, 99

1.6. Halle los cuartiles para la variable calificación en la primera unidad de la asignatura de Pedagogía en el periodo académico actual. Las calificaciones de una muestra de 22 estudiantes son: 14, 14, 14, 14, 15, 15, 15, 16, 16, 16, 16, 16, 17, 17, 17, 17, 18, 18, 18

1.7. Las puntuaciones en la prueba de CI para un grupo de 20 niños de 5 años son: ci: 89, 98, 96, 102, 94, 106, 121, 85, 115, 114, 97, 84, 90, 85, 93, 109, 96, 100, 113, 100. a. Hallar los cuartiles de las puntuaciones en la prueba de CI b. Qué porcentaje de niños tiene un Ci superior a 110 puntos?.

1.8. Para evaluar la motricidad gruesa en niños entre 3 y 10 años, se aplica el Test de Desarrollo Motor Grueso-2 (TGMD-2) estandarizado a 20 niños y se obtienen las siguientes puntuaciones: 124, 77, 82, 93, 88, 92, 117, 99, 102, 115, 121, 119, 86, 107, 98, 118, 99, 115, 105, 100, 99, 96, 102, 99. Con esta información determine: a. Los cuartiles, b. el porcentaje de niños con una evaluación de la motricidad superior a 100 puntos, c. el porcentaje de niños con una evaluación de la motricidad mayor o igual a 100 puntos.

1.9. En un estudio con niños de primer grado de educación básica, se ha determinado que 270 niños son diestros, 30 niños son zurdos y no se encontraron niños ambidiestros. Realice la tabla de frecuencias absolutas, frecuencias relativas y porcentajes.

1.10. Con los datos del ejercicio anterior, realice el gráfico de sectores con los porcentajes.

1.11. Los registros de tiempos en minutos de mujeres en una carrera de 5 kilómetros son: 22.5, 21.8, 23.1, 20.9, 24.3, 22.0, 21.5, 23.7, 22.8, 20.5, 21.2, 24.0, 23.3, 22.1, 21.0, 20.7, 23.5, 22.9, 21.7, 24.1. Con estos registros, determine: a. Medidas de tendencia central (Media, Mediana, Moda), b. Medidas de dispersión (Rango, Varianza, Desviación estándar).

1.12. Los registros de anotaciones de baloncesto de 20 jugadores durante un torneo (X: puntos anotados) son: 12, 15, 9, 18, 14, 11, 17, 13, 10, 16, 12, 20, 8, 14, 13, 19, 15, 10, 16, 11. Con esta información, hallar a. Medidas de tendencia central (Media, Mediana, Moda); b. Medidas de dispersión (Rango, Varianza, Desviación estándar); c. medidas de posición (Percentiles 25 y 75); Medidas de forma (Coeficientes de asimetría y curtosis).

1.13. En la ronda de clasificación de los Juegos Olímpicos de París 2024 para el salto alto femenino, se tienen los siguientes registros.

- **Grupo A:** Yaroslava Mahuchikh: 1.95, Eleanor Patterson: 1.95, Christina Honsel: 1.95, Elena Kulichenko: 1.92, Nawal Meniker: 1.92, Valdileia Martins: 1.92, Angelina Topic: 1.92, Rachel Glenn: 1.88, Michaela Hrubá: 1.88, Morgan Lake: 1.88, Mirela Demireva: 1.88, Yelizaveta Matveyeva: 1.88, Daniela Stanciu: 1.88, Rose Amoanimaa Yeboah: 1.88, Maria Zdzik: 1.83, Panagiota Dosi: NM
- **Grupo B:** Nicola Olislagers: 1.95, Iryna Gerashchenko: 1.95, Safina Saddyayeva: 1.95, Tatiana Gusin: 1.92, Buse Savaskan: 1.92, Vashti Cunningham: 1.92, Imke Onnen: 1.92, Airine Palsyte: 1.88, Temitope Simbiat Adeshina: 1.88, Solene Gicquel: 1.88, Lamara Distin: 1.88, Lia Apostolovskii: 1.83, Elisabetta Vandì: 1.83

Fuente: [Resultados oficiales de París 2024 – El País] <https://elpais.com/deportes/resultados/juegos-olimpicos/atletismo/atletismo-salto-de-altura-femenino/>

Con esta información determine a. Las medidas de tendencia central del grupo A, b. Las medidas de tendencia central del grupo B, c. comparar.

1.14. Con los datos del ejercicio 13, hallar: a. la desviación estándar de cada

grupo, b. el coeficiente de variación, c. que grupo presenta mayor variabilidad?.

1.15. Con los datos del ejercicio 13, realiza el histograma de frecuencias de las alturas alcanzadas con todos los datos.

1.16. Realizar un gráfico de barras con los resultados generales de las olimpiadas. Estados Unidos de América: 126, República Popular China: 91, Reino Unido: 65, Francia: 64, Australia: 53, Japón: 45, Italia: 40, Países Bajos: 34, Alemania: 33, República de Corea: 32 Fuente: <https://www.olympics.com/es/olympic-games/paris-2024/medals>

1.17. Realizar una tabla de frecuencias con la lista de los Máximos Goleadores.

- Cristiano Ronaldo: 950 goles en 1,294 partidos oficiales.
- Lionel Messi: 891 goles en 1,131 partidos oficiales.
- Josef Bican: 805 goles en 530 partidos oficiales.
- Pelé: 770 goles en 812 partidos oficiales.
- Romario: 761 goles en 963 partidos oficiales.

1.18. Con los datos del ejercicio 18, agregue una columna para determinar la tasa de efectividad o goles promedio por partido y establezca un nuevo ranking.



Fuente: Freepik

CAPÍTULO 2

Probabilidades

2.1 Introducción

Las probabilidades juegan un papel importante en la inferencia estadística, a pesar de tener un origen en los juegos de azar. Se podría decir que la probabilidad es el grado de certeza con el que un acontecimiento ocurre. A continuación, se presentan algunos conceptos básicos y reglas para entender lo que son las probabilidades.

2.2 Conceptos básicos

- **Experimento aleatorio** es un proceso mediante el cual se obtiene una observación.
- **Espacio muestral** Ω es el conjunto de todos los posibles resultados del experimento aleatorio.
- **Evento** es un subconjunto del espacio muestral. Se lo puede denotar con letras mayúsculas: $E_1, E_2, \dots, E_n$ o $A, B, \dots, Z$.

Definición clásica de probabilidad

Si un experimento que está sujeto al azar resulta de n formas igualmente probables y mutuamente excluyentes, y si n_A de estos resultados tienen un atributo A , la probabilidad de A es:

$$\mathbb{P}(A) = \frac{n_A}{n}, \quad \text{donde } 0 \leq n_A \leq n$$

2.2.1 Axiomas de la probabilidad

Las propiedades fundamentales de la probabilidad son:

- Dado un evento simple E_i , con $i = 1, 2, \dots, n$, se cumple: $0 \leq \mathbb{P}(E_i) \leq 1$
- La suma de las probabilidades es 1: $\sum_{i=1}^n \mathbb{P}(E_i) = 1$
- La probabilidad del espacio muestral es 1: $\mathbb{P}(\Omega) = 1$

A continuación, se presentan ejemplos típicos de probabilidades relacionadas con los juegos de azar que permiten comprender las definiciones básicas y la definición clásica de la probabilidad. Más adelante, se podrá establecer su asociación con ciertos casos prácticos de diferentes disciplinas.

Ejemplo 2-1. Para el experimento aleatorio "Lanzamiento de una moneda", determine el espacio muestral, eventos y las respectivas probabilidades.

Espacio muestral: $\Omega = \{C, S\}$

Eventos C: se observa cara, S: se observa sello

Probabilidades: $P(C) = \frac{1}{2}$, $P(S) = \frac{1}{2}$

Ejemplo 2-2. Para el experimento aleatorio "Lanzamiento de un dado", determine el espacio muestral, eventos y las respectivas probabilidades.

Espacio muestral: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Eventos:

- E_i : se observa el número i , con $i = 1, 2, \dots, 6$
- A : se observa número par
- B : se observa número impar
- C : se observan números menores que 5

Probabilidades:

$$P(E_i) = \frac{1}{6}, P(A) = \frac{3}{6}, P(B) = \frac{3}{6}, P(C) = \frac{4}{6}$$

Ejemplo 2-3. Para el experimento aleatorio "Extracción de una carta", determine el espacio muestral, eventos y las respectivas probabilidades.

Espacio muestral: $\Omega = \{As\heartsuit, \dots, K\heartsuit, As\diamondsuit, \dots, K\diamondsuit, As\clubsuit, \dots, K\clubsuit, As\spadesuit, \dots, K\spadesuit\}$

Eventos: Algunos ejemplos de eventos simples (un solo resultado) y eventos compuestos (al menos un resultado) son:

- $As\heartsuit$: se observa carta As de corazón rojo
- K : se observa una carta K
- $\heartsuit$: se observa una carta de corazón rojo

Probabilidades:

$$P(As\heartsuit) = 1/52, P(K) = 4/52, P(As) = 13/52$$

Ejemplo 2-4. De un estudio realizado a 200 estudiantes, se determinó que 60 de ellos practican cualquier deporte un día por semana. ¿Cuál es la probabilidad de que al seleccionar un estudiante al azar este practique cualquier

deporte un día por semana?

Se define el evento A : estudiante practica cualquier deporte un día por semana, y se tiene $n_A = 60$, con $n=200$, entonces:

$$\mathbb{P}(A) = \frac{60}{200} = 0.30$$

Ejemplo 2-5. A 80 niños de entre 20 y 30 meses se les pide reconocer las partes de su cuerpo (Cabeza, Brazos, Piernas, Tronco, Manos, Pies) mediante un juego y en la primera ronda se determina que 52 niños reconocen 4 o más partes de su cuerpo, 16 niños reconocen entre 2 y 3 partes de su cuerpo y 12 niños solo reconocen una parte de su cuerpo.

a. Identifique los eventos

R_1 : El niño reconoce una parte de su cuerpo

R_2 : El niño reconoce entre 2 y 3 partes de su cuerpo

R_3 : El niño reconoce 4 o más partes de su cuerpo

b. La probabilidad de que el niño reconozca 4 o más partes de su cuerpo.

$$\mathbb{P}(R_1) = \frac{52}{80} = 0.65$$

c. El porcentaje de niños que reconocen 4 o más partes de su cuerpo.

$$0.65 * 100 \% = 65 \%$$

El 65 % de los niños reconocen 4 o más partes de su cuerpo.

d. La probabilidad de que el niño reconozca entre 2 y 3 partes de su cuerpo.

$$\mathbb{P}(R_2) = \frac{16}{80} = 0.20$$

e. La probabilidad de que el niño reconozca solamente una parte de su cuerpo.

$$\mathbb{P}(R_3) = \frac{12}{80} = 0.15$$

f. Verifique la segunda propiedad fundamental de la probabilidad

$$\sum_{i=1}^n \mathbb{P}(R_i) = R_1 + R_2 + R_3 = 0.65 + 0.20 + 0.15 = 1$$

Ejemplo 2-6. Determine el espacio muestral del experimento aleatorio: "Lanzamiento de dos monedas".

Un diagrama de árbol permite representar la ocurrencia de eventos sucesi-

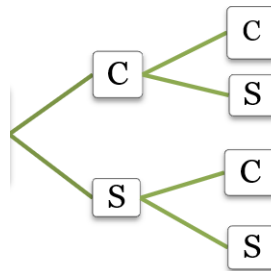
vos. En este caso, el espacio muestral del lanzamiento de dos monedas es:

$$\Omega = \{CC, CS, SC, SS\}$$

Este resultado también puede hallarse con ayuda del diagrama de árbol siguiente figura 2.1.

Figura 2.1

Diagrama de árbol



Elaboración propia

Ejemplo 2-7. Determine el número de elementos del espacio muestral de los siguientes experimentos aleatorios.

a. Experimento aleatorio de n lanzamientos sucesivos de monedas. Se halla el resultado de forma inductiva.

- Al lanzar 2 monedas, se tienen 4 resultados que se pueden expresar: 2^2 .
- Al lanzar 3 monedas, se hallan 8 resultados que se pueden expresar: 2^3 .
- Al lanzar n monedas, se hallan 8 resultados que se pueden expresar: 2^n .

Como se observa, de forma general, la base 2 corresponde al número de resultados del experimento, en este caso: cara y sello, mientras que n corresponde al número de lanzamientos sucesivos de la moneda o repeticiones del experimento aleatorio.

b. Experimento aleatorio de n lanzamientos sucesivos de dados.

Por analogía con el razonamiento anterior, ahora la base de la potencia es 6, que corresponde al número de resultados del dado en cada lanzamiento, por lo tanto, se tienen 6^n posibilidades en n lanzamientos.

2.3 Técnicas de conteo

Cuando el espacio muestral tiene muchos elementos y no es posible enumerar los elementos, resulta conveniente usar las técnicas de conteo, que nos permiten hallar el número de elementos del espacio muestral y sus elementos. A continuación, se exponen tres técnicas de conteo.

Regla $m \times n$

Sea k_i el número de elementos del grupo i , teniendo n grupos con $i=1,2,\dots, n \dots, kn$, entonces el número de arreglos con un elemento de cada grupo es: $k_1 \times k_2 \times \dots \times k_n$.

Permutación El número de formas distintas en las que se pueden ordenar n objetos tomados r a la vez ($r \leq n$), cuando interesa el orden es una “permutación de n objetos tomados r a la vez”.

$${}_nP_r = \frac{n!}{(n-r)!}$$

Combinación El número de formas distintas en que se pueden ordenar n objetos tomados r a la vez ($r \leq n$) cuando no interesa el orden, es una “combinación de n objetos tomados r a la vez”.

$${}_nC_r = \frac{n!}{r!(n-r)!}$$

Ejemplo 2-8. ¿Cuántas parejas de un hombre y una mujer que se pueden formar con 4 hombres y 3 mujeres?

Se aplica la regla $m \times n$, por lo que se obtienen $4 \times 3 = 12$ parejas

Ejemplo 2-9. ¿Cuántas ternas se pueden formar al lanzar 3 dados?

Se aplica la regla $m \times n$, por lo que se obtiene $6 \times 6 \times 6 = 6^3$ ternas.

Ejemplo 2-10. ¿Cuántos arreglos diferentes se pueden formar con las 5 vocales, cuando interesa el orden de las letras?

Se aplica una permutación de $n = 5$ vocales tomadas $r = 5$ a la vez, es decir:

$${}_nP_r = {}_5P_5 = \frac{5!}{(5-5)!} = \frac{5!}{0!} = 5! = 120$$

arreglos de 5 letras. donde $0!=1$, $1!=1$, y $5!=5*4*3*2*1*0!=120$ A continuación se enumeran los primeros resultados:

aeiou
aeiuo
aeoiu
aeoui
aeuoi ...

Ejemplo 2-11. ¿Cuántos arreglos diferentes de dos vocales se pueden formar con las 5 vocales, cuando interesa el orden de las letras?

Se aplica una permutación de $n = 5$ vocales tomadas $r = 2$ a la vez, y se obtienen 20 arreglos de 2 letras.

$${}_nP_r = {}_5P_2 = \frac{5!}{(5-2)!} = \frac{5!}{3!} = \frac{5 * 4 * 3!}{3!} = 20$$

Estos resultados pueden hallarse directamente con la calculadora mediante la función permutación y, para el siguiente caso, con la función combinación. A continuación se enumeran los primeros resultados:

ae ea ia oa ua
ai ei, ...
ao eo,
au eu iu ou

Ejemplo 2-12. Cuantos arreglos diferentes de dos vocales se pueden formar con las 5 vocales, cuando no interesa el orden de las letras?

Se aplica una combinación de $n = 5$ vocales tomadas $r = 2$ a la vez, y se obtienen 10 arreglos de 2 letras.

$${}_nC_r = {}_5C_2 = \frac{5!}{(5-2)!2!} = \frac{5!}{3!2!} = \frac{5 * 4}{2 * 1} = 10$$

A continuación se enumeran los primeros resultados:

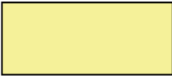
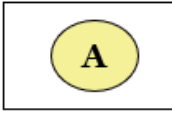
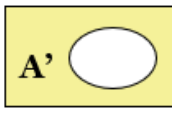
ae
ai ei, ...
ao eo,
au eu io ou

2.4 Reglas de probabilidad

A continuación, se presentan las operaciones entre conjuntos y también son válidas para los eventos.

Figura 2.2

Diagrama de árbol

Evento		Representación
Ω	Evento seguro	 $p(\Omega) = 1$
$\emptyset$	Evento nulo	 $p(\emptyset) = 0$
A	Evento A	 $p(A) = \frac{n_A}{n}$
A^c A'	No sucede A	 $p(A^c) = 1 - p(A)$

2.4.1 Regla de adición de eventos

La regla de probabilidad conjunta requiere la comprensión de los eventos mutuamente excluyentes, y el último diagrama de Venn visualiza conjuntos intersecantes y no intersecantes que ahora se llamaran respectivamente incluyentes o mutuamente excluyentes.

Dos eventos A y B son **mutuamente excluyentes** cuando no tienen resultados comunes, es decir: $A \cap B = \emptyset$, por lo que, $P(\emptyset) = 0$.

Regla de adición. Sean A y B dos eventos cualesquiera, entonces la probabilidad de que sucedan A o B es:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Caso particular. Si los eventos A y B son mutuamente excluyentes, se cumple que:

$$P(A \cup B) = P(A) + P(B)$$

Ejemplo 2-13. Se realiza un estudio a 50 niños de 5 años sobre la Coordinación motora fina y la resistencia cardiovascular y se determina que 20 niños tienen buena coordinación motora fina, 22 niños tienen buena resistencia cardiovascular y 8 niños tienen ambas cualidades

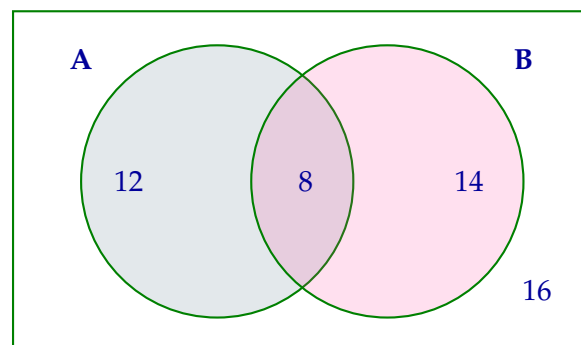
- Represente los datos del problema en un diagrama de Venn.
- Represente las probabilidades del problema en un diagrama de Venn.
- ¿Cuál es la probabilidad de que no tenga coordinación motora fina?
- ¿Cuál es la probabilidad de que un niño seleccionado al azar tenga buena motricidad fina o buena resistencia cardiovascular?
- ¿Cuál es la probabilidad de que no tenga ninguna de las dos?

Solución ítem a)

Puesto que, 8 niños tienen las dos cualidades, del total de niños con cada cualidad se resta 8, para repartir en la cada sección del círculo como se observa en la figura 2.3, es decir $20-8=12$, 8, $22-8=14$, y por fuera de estas secciones se encuentra lo que falta para completar el grupo de 50 niños, es decir $50-12-8-14=16$.

Figura 2.3

Diagrama de Venn: Coordinación motora fina y resistencia cardiovascular en 100 niños



Solución ítem b)

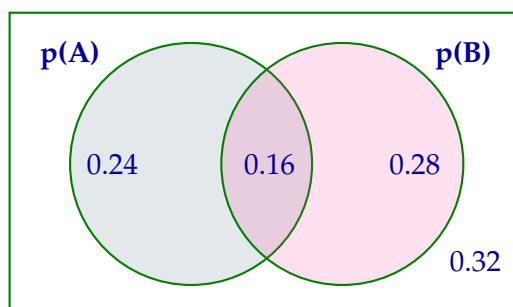
Los resultados anteriores ahora son asociados con sus probabilidades: Se tienen primero las probabilidades del problema:

- A : niño tiene buena coordinación motora fina, $P(A) = \frac{20}{50} = 0.40$
- B : niño tiene buena resistencia cardiovascular, $P(B) = \frac{22}{50} = 0.44$
- $A \cap B$: niños con dos cualidades, por lo que, $P(A \cap B) = \frac{8}{50} = 0.16$

Como $P(A \cap B) = 0.16$, entonces, la probabilidad de A , sin la intersección, es $0.40 - 0.16 = 0.24$, mientras que la probabilidad de B , sin la intersección, es $0.44 - 0.16 = 0.28$; por lo tanto, el diagrama de Venn queda representado en la figura 2.4.

Figura 2.4

Diagrama de Venn: Probabilidades



Solución item c)

$$P(A^c) = 1 - P(A) = 1 - 0.40 = 0.60$$

Solución item d)

Se aplica la regla general de adición para dos eventos cualquiera.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = 0.40 + 0.44 - 0.16 = 0.72$$

Es decir, la probabilidad de que un niño tenga al menos una de las dos cualidades es 0.72.

Solución item e)

El evento solicitado es el evento contrario de $P(A \cup B)$, por lo tanto, la probabilidad de que el niño no tenga ni coordinación ni resistencia es:

$$P((A \cup B)^c) = 1 - P(A \cup B) = 1 - 0.72 = 0.28$$

2.4.2 Regla de probabilidad conjunta

La regla de probabilidad conjunta requiere la comprensión de la independencia de eventos y la probabilidad condicional.

- La **probabilidad condicional** de que suceda el evento B dado que ocurrió el evento A, se denota por $p(B|A)$.
- Dos eventos A y B son **independientes**, cuando la ocurrencia o probabilidad de uno de ellos no afecta en la ocurrencia o probabilidad del otro evento.
- **Observación.** Si A y B son independientes, se cumple una de las siguientes proposiciones:
 - i. $P(A|B) = P(A)$
 - ii. $P(B|A) = P(B)$

Regla de probabilidad conjunta. Sean A y B dos eventos cualesquiera, la probabilidad de que ocurran A y B al mismo tiempo o probabilidad conjunta entre A y B, se define como:

$$P(A \cap B) = P(A)P(B|A)$$

Caso particular: Si A y B son eventos independientes, se cumple lo siguiente:

$$P(A \cap B) = P(A)P(B)$$

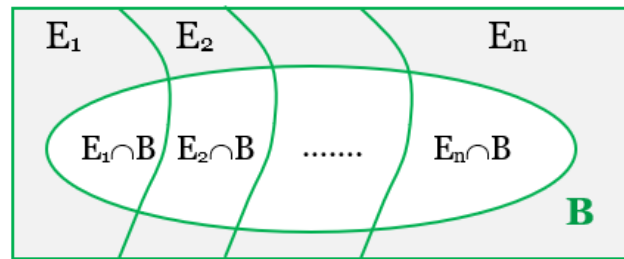
2.4.3 Regla de probabilidad marginal

La probabilidad marginal se entiende como la probabilidad para un evento que es el resultado de la unión de intersecciones con otros conjuntos o eventos. Esto se puede explicar con ayuda de la figura 2.5 y luego con la regla denominada regla de probabilidad marginal.

$$B = (E_1 \cap B) \cup (E_2 \cap B) \cup \dots \cup (E_n \cap B)$$

$$P(B) = P(E_1 \cap B) \cup P(E_2 \cap B) \cup \dots \cup P(E_n \cap B)$$

$$P(B) = P(E_1)P(B|E_1) + P(E_2)P(B|E_2) + \dots + P(E_n)P(B|E_n)$$

Figura 2.5*Probabilidad marginal del evento B***Regla de probabilidad marginal del evento B.**

Suponga que, $E_1, E_2, \dots, E_n$, son eventos mutuamente excluyentes, y B un evento cualquiera, entonces, la probabilidad marginal de que suceda B es:

$$P(B) = \sum_{i=1}^n P(E_i \cap B)$$

Al aplicar la regla de probabilidad conjunta en la regla anterior, se tiene

$$P(B) = \sum_{i=1}^n P(E_i)P(B|E_i)$$

Ejemplo 2-14. Suponga el hecho de que una persona se encuentre satisfecha en la vida está asociada con el número de hijos que tiene. Ante esta duda se ha planteado una encuesta que arroja los siguientes resultados:

- La probabilidad de que una persona no tenga hijos es 0.25
- La probabilidad de que una persona tenga un hijo es 0.35
- La probabilidad de que una persona tenga 2 o mas hijos es 0.40
- La probabilidad de que una persona se sienta satisfecha en la vida cuando no tiene hijos es 0.46
- La probabilidad de que una persona se sienta satisfecha en la vida cuando tiene un hijo es 0.32
- La probabilidad de que una persona se sienta satisfecha en la vida cuando tiene 2 o mas hijos es 0.17.

Con esta información, realice las siguientes actividades:

- a. Transforme las expresiones verbales de probabilidad de los datos en notación estadística adecuada.

$$P(H_1) = 0.25$$

$$P(H_2) = 0.35$$

$$P(H_3) = 0.40$$

$$P(S) = ?$$

$$P(S|H_1) = 0.46$$

$$P(S|H_2) = 0.32$$

$$P(S|H_3) = 0.17$$

- b. Determine la probabilidad de que una persona tenga uno o mas hijos.

Puesto que, H_1 y H_2 son eventos mutuamente excluyentes, se tiene:

$$P(H_2 \cup H_3) = P(H_2) + P(H_3) = 0.32 + 0.17 = 0.49$$

- c. Determine la probabilidad de que una persona tenga un hijo y se encuentre satisfecha con su vida.

$$P(H_1 \cap S) = P(H_1)P(S|H_1) = 0.25 * 0.115$$

$$P(H_1 \cap S) = 0.02875$$

- d. Determine la probabilidad de que una persona este satisfecha con su vida.

$$P(S) = P(H_1 \cap S) + P(H_2 \cap S) + P(H_3 \cap S)$$

$$P(S) = P(H_1)P(S|H_1) + P(H_2)P(S|H_2) + P(H_3)P(S|H_3)$$

$$P(S) = 0.25 * 0.46 + 0.35 * 0.32 + 0.40 * 0.17$$

$$P(S) = 0.295$$

- e. Probabilidad de que una persona no esté satisfecha con su vida.

$$P(S^c) = 1 - P(S)$$

$$P(S^c) = 1 - 0.295$$

$$P(S^c) = 0.705$$

2.4.4 Tablas de contingencia y diagramas de árbol para probabilidades

Las tablas de contingencia, diagramas de árbol y los diagramas de Venn ayudan a representar de forma gráfica un problema de probabilidad.

Ejemplo 2-15. Después de una competencia de 5km, se ha aplicado una encuesta a 120 corredores para establecer la relación entre las variables categoría de los atletas participantes y la hidratación durante esta carrera relativamente corta (semi fondo). Los resultados obtenidos se representan en la Tabla 2.1.

Tabla 2.1

Cruce Hidratación y Categoría

Hidratación	Categoría			Subtotal
	Juvenil	Senior	Máster	
Sí	16	20	16	52
No	20	40	8	68
Subtotal	36	60	24	120

Los eventos para el factor de hidratación se representan:

H1: El corredor se hidrata

H2: El corredor no se hidrata

Los eventos para el factor categoría del corredor se representan:

J: El corredor pertenece a la categoría Juvenil

S: El corredor pertenece a la categoría Senior

M: El corredor pertenece a la categoría Máster

En los literales que se desarrollan a continuación, se supone que se elige un corredor al azar y para este se piden las probabilidades de ocurrencia de cada uno de los siguientes eventos en la competencia 5 km.

a. Probabilidades marginales.

Probabilidad de que se hidrate:

$$P(H_1) = \frac{52}{120}$$

Probabilidad de que no se hidrate:

$$P(H_2) = \frac{68}{120}$$

La suma de las probabilidades marginales de hidratación es 1.

Probabilidad de que pertenezca a la categoría Juvenil: $P(J) = \frac{36}{120}$
Probabilidad de que pertenezca a la categoría Senior: $P(S) = \frac{60}{120}$
Probabilidad de que pertenezca a la categoría Máster: $P(M) = \frac{24}{120}$
La suma de la probabilidades marginales de categoría suman es 1.

b. Probabilidades conjuntas.

Probabilidad de que se hidrate y sea Juvenil: $P(H_1 \cap J) = \frac{16}{120}$
Probabilidad de que se hidrate y sea Senior: $P(H_1 \cap S) = \frac{20}{120}$
Probabilidad de que se hidrate y sea Máster: $P(H_1 \cap M) = \frac{16}{120}$

Probabilidad de que no se hidrate y sea Juvenil: $P(H_2 \cap J) = \frac{20}{120}$
Probabilidad de que no se hidrate y sea Senior: $P(H_2 \cap S) = \frac{40}{120}$
Probabilidad de que no se hidrate y sea Máster: $P(H_2 \cap M) = \frac{8}{120}$
La suma de todas las probabilidades conjuntas es 1.

c. Probabilidades condicionales.

Probabilidad de que se hidrate dado que es Juvenil: $P(H_1 | J) = \frac{16}{36}$
Probabilidad de que no se hidrate dado que es Juvenil: $P(H_2 | J) = \frac{20}{36}$

Probabilidad de que se hidrate dado que es Senior: $P(H_1 | S) = \frac{20}{60}$
Probabilidad de que no se hidrate dado que es Senior: $P(H_2 | S) = \frac{40}{60}$

Probabilidad de que se hidrate dado que es Máster: $P(H_1 | M) = \frac{16}{24}$
Probabilidad de que no se hidrate dado que es Máster: $P(H_2 | M) = \frac{8}{24}$

Probabilidad de que sea Juvenil dado que se hidrata: $P(J | H_1) = \frac{16}{52}$
Probabilidad de que sea Senior dado que se hidrata: $P(S | H_1) = \frac{20}{52}$
Probabilidad de que sea Máster dado que se hidrata: $P(M | H_1) = \frac{16}{52}$

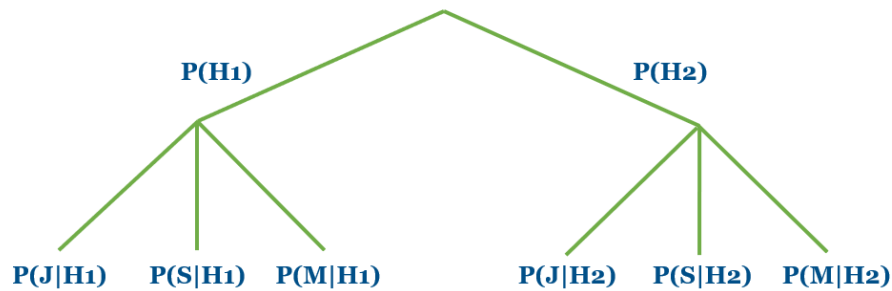
Probabilidad de que sea Juvenil dado que no se hidrata: $P(J | H_2) = \frac{20}{68}$
Probabilidad de que sea Senior dado que no se hidrata: $P(S | H_2) = \frac{40}{68}$
Probabilidad de que sea Máster dado que no se hidrata: $P(M | H_2) = \frac{8}{68}$
La suma de cada grupo de probabilidades condicionales es 1.

d. Realice el diagrama de árbol del problema. Una tabla de contingencia con todos los datos permite representar el diagrama de árbol, empezando con

los eventos de fila o de columna, en este caso se inicia con los eventos relacionados con hidratación que están en las filas de la tabla de contingencia.

Figura 2.6

Diagrama de árbol de problema



Elaboración propia

Ejemplo 2-16. La Tabla 2.2 es uno de los resultados del estudio sobre el desarrollo psicomotor en función de la edad realizado a una muestra de 119 niños. Con esta información, determine las probabilidades marginales, conjuntas y condicionales utilizando la notación estadística apropiada para cada caso.

Tabla 2.2

Déficit del desarrollo psicomotor en una muestra de 119 niños

Escala de Bayley				
Edad (meses)	Riesgo	Retraso	Sin déficit	Subtotal
8 meses	10 (23.2 %)	3 (7 %)		43
18 meses	3 (7.7 %)	0 (0 %)		39
30 meses	1 (2.7 %)	0 (0 %)		37
Subtotal	14 (11.8 %)	3 (2.5 %)		119

Elaboración propia

A partir de la tabla se pueden identificar los siguientes eventos:

Edad

G_1 : Edad 8 meses

G_2 : Edad 18 meses

G_3 : Edad 30 meses

Déficit del desarrollo psicomotor

I : Riesgo

E : Retraso

N : Sin déficit

a. Probabilidades marginales.

$$P(G_1) = \frac{43}{119} \quad P(G_2) = \frac{39}{119} \quad P(G_3) = \frac{37}{119}$$

Con respecto a la edad, se verifica: $P(G_1) + P(G_2) + P(G_3) = 1$

Se deja al lector hallar las probabilidades:

$$P(I) = \frac{1}{119} \quad P(E) = \frac{1}{119} \quad P(N) = \frac{1}{119}$$

b. Probabilidades conjuntas.

$$P(G_1 \cap I) = \frac{10}{119}$$

$$P(G_1 \cap E) = \frac{3}{119}$$

$$P(G_1 \cap N) = \frac{30}{119}$$

$$P(G_2 \cap I) = \frac{3}{119}$$

$$P(G_2 \cap E) = \frac{0}{119}$$

$$P(G_2 \cap N) = \frac{29}{119}$$

$$P(G_3 \cap I) = \frac{1}{119}$$

$$P(G_3 \cap E) = \frac{0}{119}$$

$$P(G_3 \cap N) = \frac{36}{119}$$

La suma de todas las probabilidades conjuntas es 1.

c. Probabilidades condicionales.

$$P(G_1|I) = \frac{10}{14}$$

$$P(G_2|I) = \frac{3}{14}$$

$$P(G_3|I) = \frac{1}{14}$$

$$P(G_1|E) = \frac{3}{3}$$

$$P(G_2|E) = \frac{0}{3}$$

$$P(G_3|E) = \frac{0}{3}$$

$$P(G_1|N) = \frac{30}{102}$$

$$P(G_2|N) = \frac{36}{102}$$

$$P(G_3|N) = \frac{36}{102}$$

La suma de las probabilidades condicionales en cada grupo es 1.

Hallar las probabilidades condicionales cuando se conoce la edad:

$$P(I|G_1)$$

$$P(E|G_1)$$

$$P(N|G_1)$$

$$P(I|G_2)$$

$$P(E|G_2)$$

$$P(N|G_2)$$

$$P(I|G_3)$$

$$P(E|G_3)$$

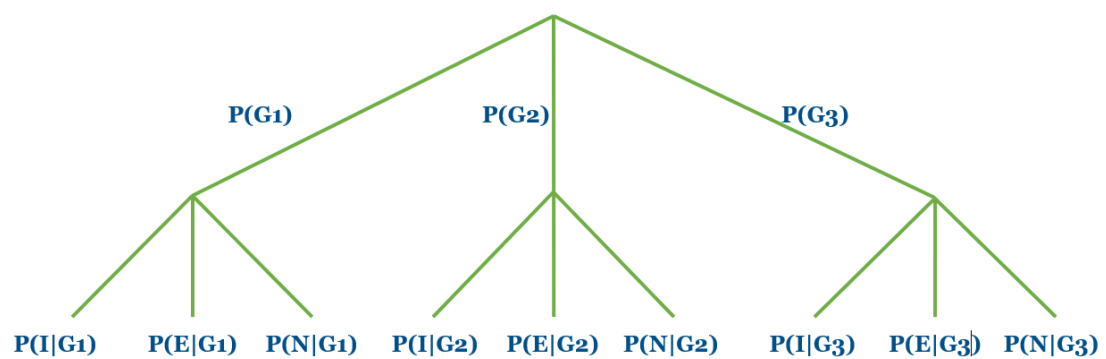
$$P(N|G_3))$$

d. Diagrama de árbol empezando con la edad (meses) del niño.

Al igual que en el ejemplo anterior, se podría iniciar el diagrama con las variables de fila o columna y sus eventos, en este caso se ha pedido iniciar con la edad del niño, como se observa en la figura 2.7.

Figura 2.7

Diagrama de árbol



Fuente: Elaboración propia

2.4.5 Probabilidad condicional *a posteriori*

La probabilidad condicional se puede hallar mediante el siguiente teorema.

Teorema de Bayes

Suponga que los eventos $E_1, E_2, \dots, E_n$ son mutuamente excluyentes, B es un evento cualquiera y dadas las probabilidades condicionales *a priori* $P(B|E_i)$, entonces, la probabilidad condicionales *a posteriori* de que ocurra el evento B dado que ocurrió el evento E_i es:

$$\frac{p(E_i)p(B|E_i)}{p(B)}$$

Al desarrollar tanto el numerador como el denominador, se tiene:

$$p(E_i|B) = \frac{p(E_i)p(B|E_i)}{p(E_1)p(B|E_1) + p(E_2)p(B|E_2) + \dots + p(E_n)p(B|E_n)}$$

Ejemplo 2-17. Recuerde el ejemplo 2.14 que relaciona la satisfacción en la vida con el número de hijos de una persona. Como continuación del ejemplo, determine las siguientes probabilidades condicionales, use el teorema de Bayes.

- a. Probabilidad de que una persona no tenga hijos, dado que esta satisfecha con su vida

$$p(H_1|S) = \frac{P(H_1 \cap S)}{P(S)} = \frac{P(H_1)P(S|H_1)}{P(S)}$$

$$p(H_1|S) = \frac{0.25 * 0.46}{0.295} = 0.3898$$

- b. Probabilidad de que una persona tenga un hijo, dado que esta satisfecha con su vida.

$$p(H_2|S) = \frac{P(H_2 \cap S)}{P(S)} = \frac{P(H_2)P(S|H_1)}{P(S)}$$

$$p(H_2|S) = \frac{0.35 * 0.32}{0.295} = 0.3796$$

- c. Probabilidad de que una persona tenga dos o mas hijos, dado que esta satisfecha con su vida. Puesto que la suma de probabilidades dentro de una condicional o nodo es 1, entonces:

$$p(H_3|S) = 1 - (p(H_1|S) + p(H_2|S)) =$$

$$1 - (0.3898 + 0.3796) = 0.2306$$

Ejemplo 2-18. Aplicación de las probabilidades en una prueba de natación de 50 metros estilo libre, donde los eventos son la técnica de la brazada y el cumplimiento de una marca del tiempo referencial.

Figura 2.8

Prueba de natación



<https://pixabay.com/>

Durante el entrenamiento en las pruebas cronometradas de 50 metros, se han identificado tres técnicas:

- T_1 : Alta rotación de brazos – Baja longitud de brazada (Menos eficiente).
- T_2 : Rotación media – Alta longitud de brazada – Tracción con codo alto (Mas eficiente).
- T_3 : Rotación irregular – Codo bajo – Tracción deficiente (Deficiente).

El 40 % de los nadadores practican la técnica 1, el 35 % la técnica 2 y el 25 % la tecnica 3, por lo tanto:

- $P(T_1) = 0.40$
- $P(T_2) = 0.35$
- $P(T_3) = 0.25$

Al incorporar el cumplimiento de la meta (tiempo menor que 28 segundos), se determina que al aplicar la técnica, la probabilidad de que el tiempo sea inferior a 28 segundos es 0.60; si aplica la técnica 2, la probabilidad de que el tiempo sea inferior a 28 segundos es 0.90; si aplica la técnica 3, la probabilidad de que el tiempo sea inferior a 28 segundos es 0.20. Lo que se traduce como:

$$P(R|T_1) = 0.60$$

$$P(R|T_2) = 0.90$$

$$P(R|T_3) = 0.20$$

- a. ¿Cuál es la probabilidad de que un nadador seleccionado al azar establezca una marca inferior a 28 segundos?

Para el evento I: Tiempo o marca inferior a 28 segundos, se aplica la regla de probabilidad marginal:

$$P(I) = P(T_1)P(I|T_1) + P(T_2)P(I|T_2) + P(T_3)P(I|T_3)$$

$$P(I) = (0.40)(0.60) + (0.35)(0.90) + (0.25)(0.20)$$

$$P(I) = 0.24 + 0.315 + 0.05 = 0.605$$

- b. Si un nadador termina su carrera de 50 metros en menos de 28 segundos, ¿cuál es la probabilidad de que lo haya logrado con la técnica 1?

Se aplica el teorema de Bayes:

$$P(T_1|I) = \frac{P(T_1)P(I|T_1)}{P(I)}$$

$$P(T_1|I) = \frac{(0.40)(0.60)}{0.605}$$

$$P(T_1|I) = 0.3967$$

La probabilidad de usar la técnica 1 cuando hizo una marca inferior a 28 segundos es 0.3967.

- c. Si un nadador termina su carrera de 50 metros en menos de 28 segundos, ¿cuál es la probabilidad de que lo haya logrado con la técnica 2?

Se aplica el teorema de Bayes:

$$P(T_2|I) = \frac{P(R|T_2)P(T_2)}{P(I)}$$

$$P(T_2|I) = \frac{(0.35)(0.90)}{0.605} =$$

$$P(T_2|I) = 0.5207$$

- d. Si un nadador termina su carrera de 50 metros en menos de 28 segundos, ¿cuál es la probabilidad de que lo haya logrado con la técnica 3?

Considerando que, $P(T_1|I) + P(T_2|I) + P(T_3|I) = 1$, entonces,

$$P(T_3|I) = 1 - P(T_1|I) - P(T_2|I)$$

$$P(T_3|I) = 1 - 0.3967 - 0.5207 = 0.0826$$

- e. Se deja al lector el cálculo de las probabilidades relacionadas con la aplicación de la técnica bajo la condición M : tiempo o marco mayor o igual que 28 segundos, que es el evento contrario de I , es decir:

$$P(T_1|M)$$

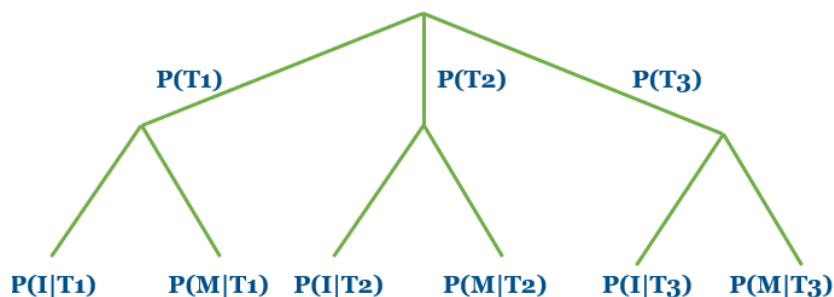
$$P(T_2|M)$$

$$P(T_3|M)$$

- f. Realizar el diagrama de árbol comenzando con los eventos T_1, T_2, T_3

Figura 2.9

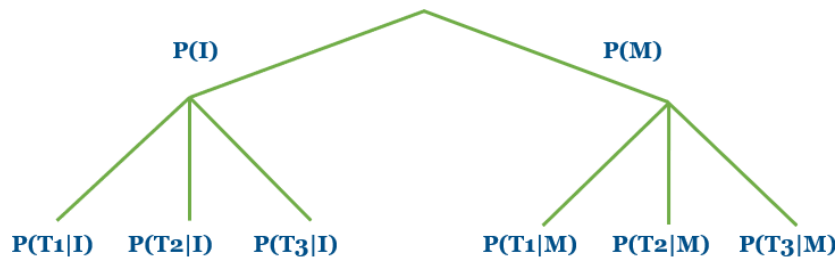
Diagrama de árbol 1



- g. Realizar el diagrama de árbol comenzando con los eventos I y M

Figura 2.10

Diagrama de árbol 2



2.5 Laboratorio de probabilidades en Jamovi

La comprensión de las definiciones, axiomas, reglas y teoremas es fundamental para el cálculo de probabilidades y su aplicación con software estadístico. En el capítulo 1, laboratorio 3 de jamovi, se presentó un ejemplo de una tabla de contingencia con porcentajes con respecto al total, por filas y con respecto a las columnas, las cuales son de utilidad para verificar las reglas de probabilidad.

Ejemplo 2-19. Con los datos del ejemplo 2.15, hallar tablas de contingencia para determinar probabilidades..

- Crear el archivo de datos en Jamovi. Crear la variable hidratación con los atributos: Si, No; y la variable categoría del atleta con atributos: Juvenil, Senior, Máster, de acuerdo a las frecuencias observadas de la Tabla 2.1.

Figura 2.11

Datos en Jamovi

Hidrata	Categ
Si	Juvenil
Si	Juvenil
Si	Juvenil

- Realizar la tabla de contingencia con las frecuencias observadas.

La figura 2.12 presenta grupos heterogéneos en cantidad ya sea por eventos de hidratación o categoría del atleta.

Figura 2.12

Hidratación vs Categoría

Tablas de Contingencia

Categ	Hidrata		Total
	No	Si	
Juvenil	20	16	36
Máster	8	16	24
Senior	40	20	60
Total	68	52	120

- c. Realizar la tabla de contingencia respecto al total general.

La figura 2.12 es una captura de la tabla obtenida con jamovi que permite distribuir de forma general los atletas en los diferentes eventos y sus cruces. Al dividir por 100 % los subtotales de fila y columna, se obtienen las probabilidades marginales de la categoría (fila): $P(Juvenil) = 0.30$, $P(Máster) = 0.20$, $P(Senior) = 0.50$ y las probabilidades marginales de la hidratación (columna): $P(Si) = 0.57$, $P(No) = 0.43$.

Figura 2.13

Hidratación vs Categoría (%) respecto al total

Tablas de Contingencia

Categ		Hidrata		Total
		No	Si	
Juvenil	% del total	16.67 %	13.33 %	30.00 %
Máster	% del total	6.67 %	13.33 %	20.00 %
Senior	% del total	33.33 %	16.67 %	50.00 %
Total	% del total	56.67 %	43.33 %	100.00 %

Las probabilidades conjuntas se hallan con esta tabla, al dividir cada celda por 100 %, de modo que: $P(Juvenil \cap No) = 0.1667$, $P(Juvenil \cap Si) = 0.1333$, $P(Máster \cap No) = 0.0667$, $P(Máster \cap Si) = 0.1333$, $P(Senior \cap No) = 0.3333$, $P(Senior \cap Si) = 0.1667$.

- d. Realizar la tabla de contingencia con respecto al total de filas.

En este caso, debe modificar la opción de las celdas al elegir "Filas", como se observa en la figura 2.14, dando lugar a la tabla de la figura 2.15.

Figura 2.14

Menú de Frecuencias en una tabla de contingencia

Frecuencias	Porcentajes
☐ Frecuencias observadas	☒ Fila
☐ Frecuencias esperadas	☐ Columna
	☐ Total

Como se puede observar cada fila representa el 100 % de observaciones. Al dividir por 100 % se obtienen las probabilidades condicionales de hidratación con respecto a cada categoría del atleta, las mismas que son:

Figura 2.15

Porcentaje de hidratación en cada categoría

Tablas de Contingencia				
Categ		Hidrata		Total
		No	Si	
Juvenil	% de fila	55.56 %	44.44 %	100.00 %
Máster	% de fila	33.33 %	66.67 %	100.00 %
Senior	% de fila	66.67 %	33.33 %	100.00 %
Total	% de fila	56.67 %	43.33 %	100.00 %

$P(\text{No}|\text{Juvenil}) = 0.5557$, $P(\text{Si}|\text{Juvenil}) = 0.4444$, $P(\text{No}|\text{Máster}) = 0.3333$,
 $P(\text{Si}|\text{Máster}) = 0.6667$, $P(\text{No}|\text{Senior}) = 0.6667$, $P(\text{Si}|\text{Senior}) = 0.3333$.

Como se puede observar en cada columna las diferencias de los resultados de cada fila en cuanto al evento hidratarse tiene diferencias superiores al 10 %; situación que es evidente tanto en la columna de "No se hidrata" "Si se hidrata", donde es más marcada la diferencia, esto puede ser un indicador de que la Categoría del atleta influye sobre la Hidratación.

Ejemplo 2-20. Realice el gráfico correspondiente a los resultados anteriores. Personalice el estilo y colores

Resulta conveniente analizar la relación entre las variables del ejemplo anterior de forma visual, dentro de una misma escala que no este influenciada por el tamaño de la muestra de cada grupo.

El proceso consiste en marcar las opciones del menú Gráfico según la figura 2.16. Donde se debe marcar la opción del tipo de gráfico "Barras", el eje y "Porcentaje"; mientras que, para el eje X se marca "Filas", esto con la intención indicada en el párrafo anterior.

Figura 2.16

Menú gráfico Tabla Contingencia

Gráficos

☐ Gráfico de Barras

Tipo de Barras

☒ Al lado
 ☐ Alineados

Eje Y

☐ Frecuencias
 ☒ Porcentajes

de filas

Eje X

☒ Filas
 ☐ Columnas

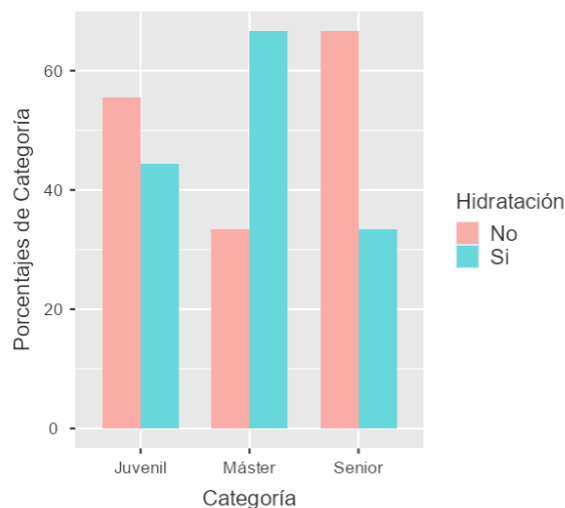
Para personalizar los estilos de gráfico y color, usar el menú de opciones de jamovi ubicado en la parte superior derecha, el mismo que dispone de las opciones que se presentan en la figura 2.17 se ha elegido el estilo **Hadley** para generar el gráfico correspondiente a la figura 2.18.

Figura 2.17

Menú Estilos

Figura 2.18

Gráfico cruzado: Hidratación vs Categoría



Ejemplo 2-21. Caso: Aprendizaje del idioma en niños extranjeros que llegan a Canadá

El corpus Paradis es un estudio realizado en el año 2002 en Edmonton, Canadá, con una muestra de lenguaje naturalista de 25 niños que apren-

den inglés como segunda lengua. Como parte del estudio, se grabaron videos con los niños conversando con un asistente de investigación estudiantil, con quien, durante 45 minutos, se plantearon preguntas planificadas, además de temas impuestos por los niños mientras la conversación avanzaba sin plantear preguntas sobre ellos. La información completa del corpus longitudinal se puede encontrar en el sitio web de CHILDES, aquí: <http://childes.talkbank.org/access/Biling/Paradis.html>. Las variables de este ejercicio se encuentran en el anexo B, de las cuales solo tomaremos las variables X_2 : género y A: aprendizaje alrededor de 10 meses. Con esta información, determine:

- a. La tabla de contingencia con los porcentajes respecto al total de filas, colocando el género en las filas. figura 2.19

Figura 2.19

Porcentaje del Periodo de aprendizaje vs Género

Tablas de Contingencia

gender		Periodo_Alrededor 10 meses		Total
		Máximo 10 m	Más de 10 m	
F	% de fila	75.00 %	25.00 %	100.00 %
M	% de fila	62.50 %	37.50 %	100.00 %
Total	% de fila	66.67 %	33.33 %	100.00 %

- b. Identifique los eventos marginales, sin calcular.
 - A1: Aprender inglés en un máximo de 10 meses
 - A2: Aprender inglés en mas de 10 meses
 - F: femenino
 - M: masculino
- c. Extraiga las probabilidades condicionales de que el niño aprenda inglés en un máximo de 10 meses dado que es de género femenino y la probabilidad de que el niño aprenda inglés en más de 10 meses dado que es de género femenino.

$$P(A_1|F) = 0.75$$

$$P(A_2|F) = 0.25$$

2.6 Ejercicios propuestos del capítulo

2.1. Dado el experimento aleatorio de extracción de una carta (juego de naipes de 52 cartas; 4 tipos de cartas: $\clubsuit$, $\diamond$, $\heartsuit$, $\spadesuit$; 13 en cada grupo), determine las siguientes probabilidades cuando se selecciona una carta al azar:

a. $P(\clubsuit)$ b. $P(\diamond)$

2.2. Utilizando los datos del ejercicio anterior, determine:

a. $P(\heartsuit)$ b. $P(\spadesuit)$

2.3. Si en el juego de cartas mencionado en el primer ejercicio, ahora se realiza la extracción de una segunda carta sin reposición de la primera carta, determine:

a. $P(\clubsuit|5\clubsuit)$ b. $P(\clubsuit|5\diamond)$

2.4. Utilizando los datos y condiciones del ejercicio anterior, determine:

a. $P(As|5\heartsuit)$ b. $P(As\spadesuit|As\heartsuit)$

2.5. Si $P(X) = 0.38$, $P(Y) = 0.21$, $P(X \cap Y) = 0.06$ determine:

a. $P(X^c)$ b. $P(X \cup Y)$

2.6. Si $P(X) = 0.38$, $P(Y) = 0.21$, $P(X \cap Y) = 0.06$ determine:

a. $P(Y^c)$ b. $P(X \cap Y^c)$

2.7. Se ha realizado un estudio exploratorio con 64 niños de 5 años sobre el estilo de aprendizaje del lenguaje. La información se toma mediante la observación directa. Sobre el caso particular, se ha determinado que 26 niños disfrutan de canciones, rimas y cuentos hablados, 16 niños siguen instrucciones verbales sin dificultad, 10 niños recuerdan palabras o frases luego de haberlas escuchado, 12 niños responden mejor a conversaciones y juegos de palabras. Si se selecciona un niño al azar,

- a. ¿Cuál es la probabilidad de que el niño disfrute de canciones, rimas y cuentos hablados?

- b. ¿Cuál es la probabilidad de que el niño no disfrute de canciones, rimas y cuentos hablados?

2.8. Utilizando los datos del ejercicio anterior, halle:

- a. ¿Cuál es la probabilidad de que el niño siga instrucciones verbales sin dificultad?
- b. ¿Cuál es la probabilidad de que el niño disfrute de canciones, rimas y cuentos hablados o responda mejor a conversaciones y juegos de palabras?

2.9. De un estudio realizado a nadadores novatos, se determina que, aparte del entrenamiento de natación, el 30 % realiza entrenamiento en seco (fuerza y core), el 15 % realiza trabajo cardiovascular (correr y bicicleta), y el 4 % realiza ambos tipos de entrenamiento. Determinar la probabilidad de que:

- a. No realice entrenamiento seco
- b. Realice entrenamiento seco o trabajo cardiovascular

2.10. Utilizando los datos del ejercicio anterior, halle:

- a. Solo realice entrenamiento seco.
- b. Solo realice trabajo cardiovascular.

2.11. Con los datos sobre el aprendizaje del idioma en niños extranjeros que llegan a Canadá, disponibles en el anexo B, determine las siguientes probabilidades al seleccionar un niño al azar:

- a. La probabilidad de que su primer idioma sea Mandarin
- b. La probabilidad de que su primer idioma no sea Mandarin
- c. La probabilidad de que el primer idioma sea Mandarin o coreano.

2.12. Utilizando los datos de la figura 2.19, determine la probabilidad condicional de que el niño aprenda inglés en máximo 10 meses dado que es de género masculino.

2.13. La probabilidad de que un jugador de fútbol esté fatigado es 0.25, esta condición puede influir sobre la incidencia de lesiones deportivas. Por los

datos históricos de lesiones, se conoce que si el futbolista esta fatigado, la probabilidad de que se lesione es 0.65, y si el futbolista no esta fatigado, la probabilidad de que se lesione es 0.20.

- a. ¿Cuál es la probabilidad de que el futbolista se lesione?
- b. ¿Cuál es la probabilidad de que el futbolista se fatigue dado que está lesionado?

2.14. Utilizando los datos del ejercicio anterior, determine:

- a. La probabilidad de que el futbolista no se lesione
- b. La probabilidad de que el futbolista se fatigue dado que no está lesionado

2.15. Una maestra realiza actividades lúdicas para determinar el disfrute (D) de niños de 5 años con tres tipos de actividades: Físicas (juegos al aire libre) con una probabilidad de 0.40; cognitivas (acertijos), con una probabilidad de 0.35; y actividades sociales y emocionales (juegos de interacción y empatía), con una probabilidad de 0.25. La probabilidad de que el niño disfrute, sabiendo que realiza actividad física es 0.84; la probabilidad de que disfrute cuando realiza actividad cognitiva es 0.45, y la probabilidad de que disfrute dado que realiza actividad de tipo emocional es 0.72,

- a. ¿Cuál es la probabilidad de que el niño realice actividad física y disfrute?
- b. ¿Cuál es la probabilidad de que el niño realice actividad cognitiva y disfrute?

2.16. Con la información del ejercicio anterior, determine:

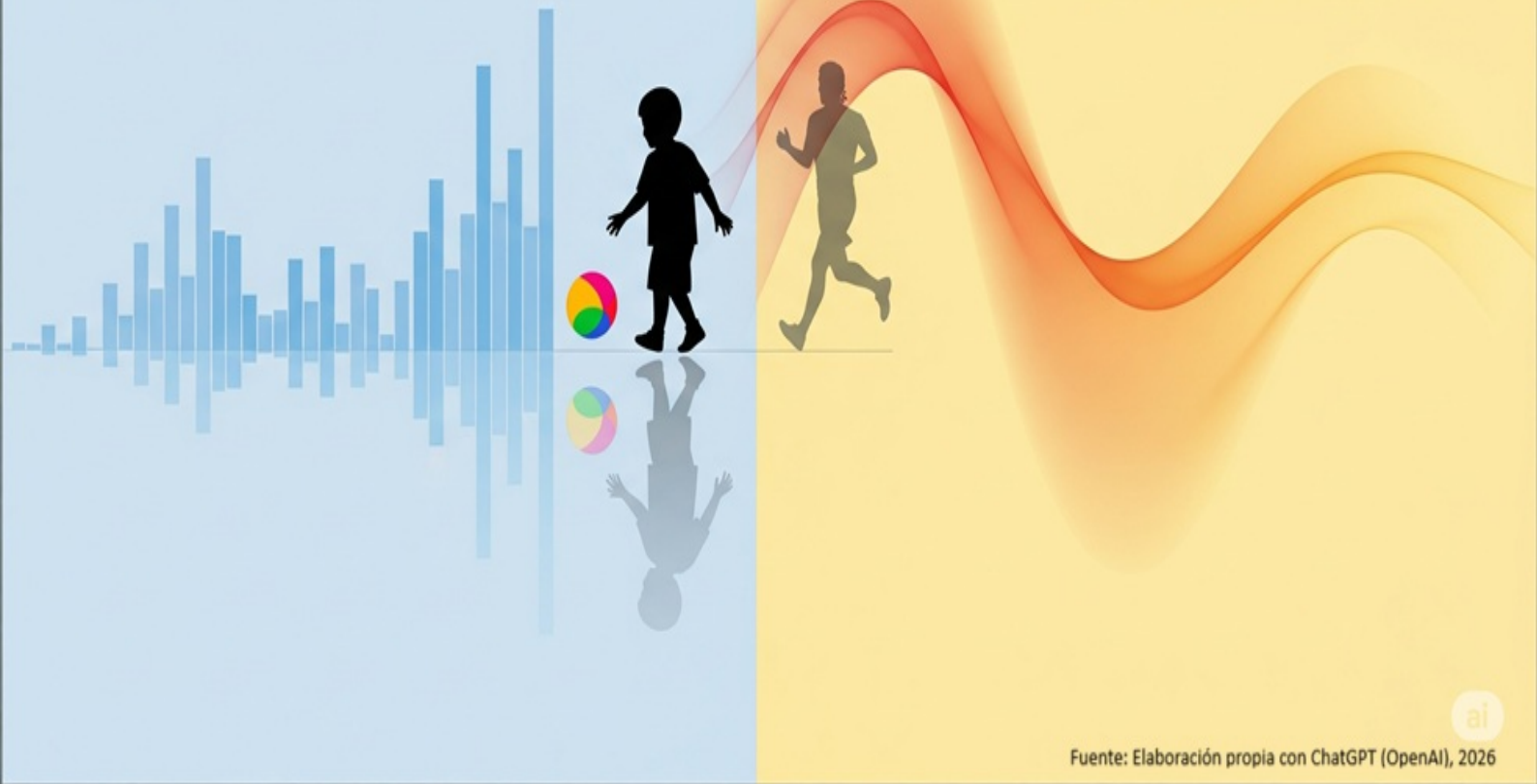
- a. ¿Cuál es la probabilidad de que el niño disfrute?
- b. ¿Cuál es la probabilidad de que el niño haya realizado una actividad cognitiva, dado que disfruta?

2.17. Utilizando los datos del ejercicio anterior, determine:

- a. La probabilidad de que el niño no disfrute
- b. La probabilidad de que el niño haya realizado una actividad cognitiva, dado que no disfruta.

Discrete Distributions

Continuous Distributions



Fuente: Elaboración propia con ChatGPT (OpenAI), 2026

CAPÍTULO 3

Variables aleatorias

3.1 Introducción

Las variables aleatorias son útiles en el campo de las probabilidades para modelar problemas mediante funciones que permiten el cálculo de probabilidades, sujetas al cumplimiento de las características que requiere cada modelo de probabilidad.

Una **variable aleatoria** (v.a) es una función que asigna un número a los resultados del espacio muestral.

$$X : \Omega \rightarrow \mathbb{R}$$

3.1.1 Tipos de variables aleatorias

Las variables aleatorias son de tipos: discretas y continuas.

- Una **Variable aleatoria discreta** toma un número finito de valores en el conjunto de los números reales; por lo general, se usan números enteros.
- La **variable aleatoria continua** toma infinitos valores en el conjunto de los números reales.

Ejemplo 3-1. A continuación se presentan variables aleatorias discretas y su rango de valores.

- Lanzamiento de dos monedas X: "número de caras en el lanzamiento de dos monedas"

Ω	X
SS	0
SC, CS	1
CC	2

- X: Número de canastas después de 5 lanzamientos (intentos) en juego de tiros al aro de basket; $X=0,1,2,3,4,5$
- X: Número de chips defectuosos en una muestra de 30; $X=0,1,2,\dots, 30$
- X: Número de futbolistas con doping positivo en una muestra de 10 futbolistas. ; $X=0,1,2,\dots, 10$
- X: número de botellas de bebida hidratante con fecha de caducidad expirada, entonces, la variable aleatoria toma valores $X= 0,1, \dots ,20$

Ejemplo 3-2. A continuación se presentan variables aleatorias continuas y su rango de valores.

- X: Tiempo de sueño diario de un niño de 6 meses (horas); $0 \leq X \leq 12$
- X: Temperatura corporal normal de un niño ($^{\circ}\text{C}$); $35.5 \leq X \leq 37.5$
- X: Velocidad máxima que puede alcanzar un jugador delantero de fútbol profesional durante un sprint (distancia corta, 20-30 metros); $32 \leq X \leq 34$
- Tiempo registrado en una carrera de 50 metros para novatos, estilo mariposa (segundos); al no especificar el rango de edad, este puede tomar valores $X \geq 35s$.
- X: pH de una bebida hidratante; puede variar entre 0 y 14.

3.1.2 Modelos de distribución de probabilidad

Una de las aplicaciones de las variables aleatorias es un modelo de probabilidad, que permiten calcular probabilidades para los valores de la variable aleatoria o un rango de valores en el dominio de la variable.

Algunos modelos de probabilidad para variables aleatorias discretas son:

- Binomial
- Hipergeométrica
- Poisson

Algunos modelos de probabilidad para variables aleatorias continuas son:

- Normal
- t-Student
- Exponencial

3.2 Variables aleatorias discretas

Para todas las variables aleatorias, ya sean discretas o continuas, se definen conceptos importantes tales como parámetros, función de probabilidad, función de distribución de probabilidad, así como se establecen las características del experimento aleatorio asociado y sus parámetros.

3.2.1 Función de probabilidad para una variable aleatoria discreta

Sea X una variable aleatoria discreta, la **función de probabilidad** se define como la probabilidad de que X tome un valor específico x dentro del dominio de la función, es decir:

$$f(x) = P(X = x)$$

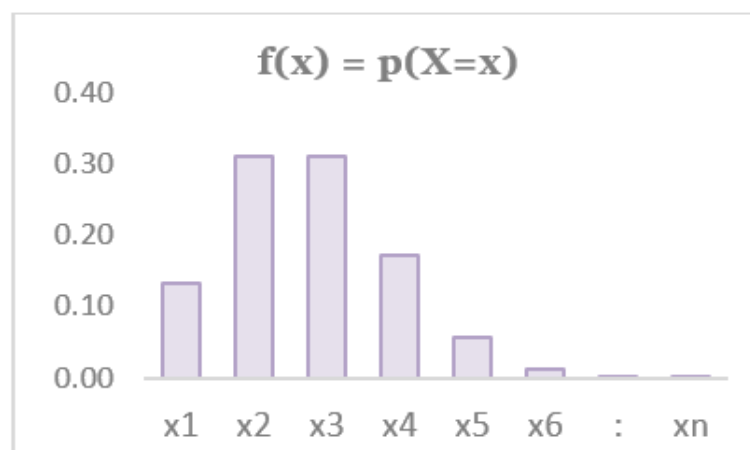
El dominio es el conjunto de valores de X y recorrido es el intervalo $[0, 1]$.

Esta función satisface dos propiedades:

$$\begin{aligned} (1) \quad & 0 \leq f(x) \leq 1 \quad \forall x \in X \\ (2) \quad & \sum_X f(x) = 1 \quad \text{si } X \text{ es v.a. discreta} \end{aligned} \quad (3.1)$$

Figura 3.1

Grafica de una v.a. discreta



3.2.2 Función de distribución para una variable aleatoria discreta

Sea X una variable aleatoria discreta, la **función de distribución** se define como la probabilidad de que X tome valores menores o iguales que un valor específico x dentro del dominio de la función, es decir:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

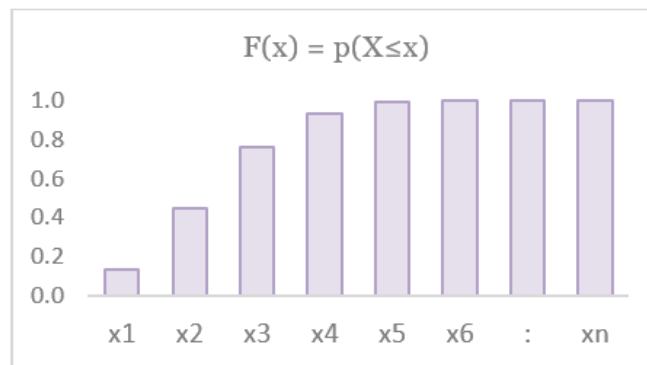
El dominio es el conjunto de valores de X y recorrido es el intervalo $[0, 1]$.

Esta función satisface cuatro propiedades:

$$\begin{aligned}
 (1) \quad & 0 \leq F(x) \leq 1 && \forall x \in X \\
 (2) \quad & F(x_i) \leq F(x_j) && \forall x_i < x_j \\
 (3) \quad & P(X > x) = 1 - P(X \leq x) && = 1 - F(x) \\
 (4) \quad & P(x_i \leq X \leq x_j) && = F(x_j) - F(x_{i-1})
 \end{aligned} \tag{3.2}$$

Figura 3.2

Función de distribución discreta



3.2.3 Media y varianza de una variable aleatoria discreta

Como parte del estudio de las variables aleatorias discretas y sus funciones de probabilidad, se puede definir la media como una función denominada valor esperado o esperanza matemática.

El valor esperado de una variable aleatoria discreta X es el promedio de X denotado por $E(x)$ o μ_X .

$$\mu_X = E(x) = \sum_X x \cdot f(x) \tag{3.3}$$

Propiedades de la Esperanza Matemática

Sea X es una variable aleatoria discreta y a, b constantes, se cumple:

- $E(a) = a$
- $E(aX) = aE(X)$
- $E(aX + bY) = aE(X) + bE(Y)$

Las propiedades de la esperanza matemática y de la varianza son de utilidad para hallar las medidas descriptivas con datos ponderados.

La varianza de una variable aleatoria discreta X se define como un valor esperado del cuadrado de las diferencias entre cada valor de la variable y su media.

$$V(X) = \sigma_X^2 = E(X - \mu_x)^2 = \sum_X (x - \mu_X)^2 \cdot f(x) \quad (3.4)$$

Propiedades de la Varianza

Sea X es una variable aleatoria discreta y a, b constantes, se cumple:

- $Var(a) = 0$
- $Var(aX) = a^2 Var(X)$
- $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$

Ejemplo 3-3. Se ha realizado un estudio sobre el número de comidas que toma un niño por día, y se determina que respecto a la variable aleatoria discreta X : número de comidas por día, con valores 3,4,5,6 le corresponden los siguiente números de niños en cada valor: 5,15,20. Con esta información determine:

- a. Hallar las funciones de probabilidad y distribución.

Tabla 3.1

Funciones de probabilidad y distribución

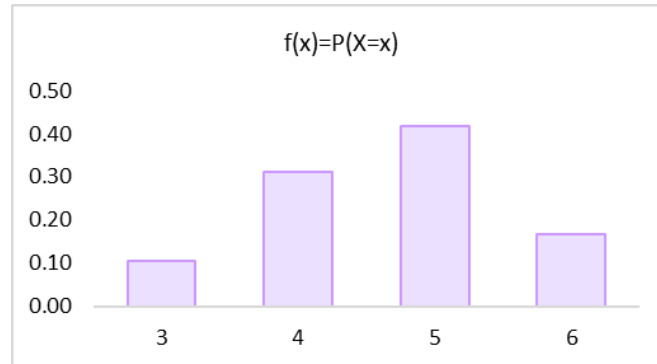
X	n_i	$f(x) = P(X = x)$	$F(x) = P(X \leq x)$
3	5	$\frac{5}{48} = 0.1042$	0.1042
4	15	$\frac{15}{48} = 0.3750$	0.4792
5	20	$\frac{20}{48} = 0.4167$	0.8959
6	8	$\frac{8}{48} = 0.1667$	1.0626

- b. Realizar la gráfica de la función de probabilidad utilizando Excel.

Lo común es observar a niños tomando entre 4 y 5 comidas (incluidas) con una probabilidad de 0.7917

Figura 3.3

Función de probabilidad



c. Hallar la esperanza matemática y la varianza

$$\mu = E(X) = \sum_x x \cdot f(x) = 3 \frac{5}{48} + 4 \frac{15}{48} + 5 \frac{20}{48} + 6 \frac{8}{48} = \frac{223}{48}$$

$$\mu = E(X) = 4.65$$

$$\sigma^2 = E(X - \mu)^2 = (3 - 4.65)^2 \frac{5}{48} + \dots + (6 - 4.65)^2 \frac{8}{48}$$

$$\sigma^2 = 0.7704$$

Ejemplo 3-4. Para realizar un estudio sobre percepción sobre la actividad física, se seleccionarán de forma aleatoria 2 adolescentes de grupo de 6 y 1 un niño de un grupo de 4. Con esta información determine las funciones de probabilidad y medidas descriptivas para la variable aleatoria X: número de adolescentes en la terna.

a. Hallar las funciones de probabilidad y distribución.

- El número total de ternas que se pueden formar con 10 personas es una combinación, ya que no interesa el orden de cada miembro en el grupo, que resulta ser: $10C3 = 120$.
- Los valores de la variable aleatoria número de adolescentes en la terna son: 0,1,2,3
- Las ternas se forman con adolescentes y niños, y para hallar la cantidad combinada de elementos se usará la regla *mxn*, donde cada termino es una combinación con la cantidad requerida para cada grupo.
- La probabilidad asociada a cada valor de la variable aleatoria se halla con la definición clásica de probabilidad: $P(A) = \frac{n_A}{n}$, siendo el denominador 120. Finalmente, se construye la función de probabilidad.

Tabla 3.2

Funciones de probabilidad y distribución

X	ni	$f(x) = P(X = x)$	$F(x) = P(X \leq x)$
0	$6C0 * 4C3 = 4$	$P(X = 0) = \frac{4}{120}$	$P(X \leq 0) = \frac{4}{120}$
1	$6C1 * 4C2 = 36$	$P(X = 1) = \frac{36}{120}$	$P(X \leq 1) = \frac{40}{120}$
2	$6C2 * 4C1 = 60$	$P(X = 2) = \frac{60}{120}$	$P(X \leq 2) = \frac{100}{120}$
3	$6C3 * 4C0 = 20$	$P(X = 3) = \frac{20}{120}$	$P(X \leq 3) = \frac{120}{120}$

b. Hallar la esperanza matemática y la varianza

$$\mu = \sum_x x \cdot f(x) = 0 \cdot \frac{4}{120} + 1 \cdot \frac{36}{120} + 2 \cdot \frac{60}{120} + 3 \cdot \frac{20}{120} = \frac{216}{120}$$

$$\mu = 1.8$$

$$\sigma^2 = E(X - \mu)^2 = (0 - 1.8)^2 \frac{4}{120} + (1 - 1.8)^2 \frac{36}{120} + (2 - 1.8)^2 \frac{60}{120} + (3 - 1.8)^2 \frac{20}{120}$$

$$\sigma^2 = \frac{14}{25} = 0.56$$

$$\sigma = \sqrt{0.56} = 0.75$$

Resumen de resultados:

- En promedio, hay 1.8 adolescentes en la terna
- La varianza es 0.56
- La desviación estándar de la terna es 0.75
- La probabilidad de que no haya adolescentes en la terna es 4/120
- La probabilidad de que 1 adolescente conforme la terna es 36/120
- La probabilidad de que 2 adolescentes conforme la terna es 60/120
- La probabilidad de que 3 adolescentes conforme la terna es 20/120

3.3 Variables aleatorias continuas

En este libro no se pretende desarrollar las variables aleatorias continuas desde el aspecto estadístico-matemático, ya que no todos los lectores tienen el requisito de los cursos de cálculo para abordar el tema desde esa perspectiva.

Ejemplo 3-5. Las siguientes variables aleatorias se consideran continuas

- Velocidad [km/h]
- Diámetro abdominal [cm]
- Tiempo de reacción ante un estímulo [segundos]
- peso de un niño de 12 meses [kg]

3.3.1 Resumen de funciones para una variable aleatoria continua X

Si X es una variable aleatoria continua, con dominio el conjunto de los números reales y recorrido el intervalo $[0, 1]$, se definen:

Función de distribución

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(x) dx$$

Esperanza matemática

$$E(x) = \mu = \int_{-\infty}^x f(x) dx$$

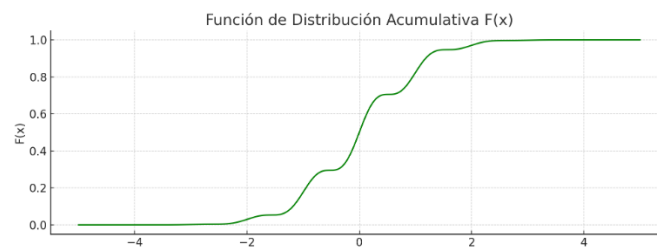
Varianza

$$\text{Var}(x) = \sigma^2 = \int_{-\infty}^x f(x) dx$$

Si X es una variable aleatoria continua, cumple las tres primeras propiedades de la función $F(x)$ del caso discreto, utilizando ahora integrales. Pero en la propiedad para el caso continuo se expresa de la siguiente manera:

$$P(x_i \leq X \leq x_j) = F(x_j) - F(x_i) \quad (3.5)$$

A continuación se presentan las gráficas de la función de probabilidad continua que se suele llamar función de densidad $f(x)$ y que se utilizará solamente para conocer su comportamiento, y la función de distribución $F(x)$ con la que se hallarán las probabilidades acumulativas.

Figura 3.4*Función de densidad***Figura 3.5***Función de distribución*

3.3.2 Resumen elementos clave de las variables aleatorias continuas

Tabla 3.3*Diferencias con lo estudiando en las variables discretas*

Concepto	Variables continuas
Rango de Valores	Infinitos valores en el conjunto de números reales
Gráfica de $f(x)$	Curva continua
Interpretación de $F(x)$	Área bajo la curva
$f(x) = P(X = x)$	Es cero
$F(x) = P(X \leq x)$	Se halla con integrales
$E(X)$, $V(X)$	Se hallan con integrales
Tablas de probabilidad	En algunos casos se usan para evitar la integración

Observación: Más adelante, en la sección de modelos de distribución para variables aleatorias continuas, se realizarán los ejercicios de aplicación.

3.4 Distribución binomial

3.4.1 Características del experimento aleatorio binomial

Un experimento aleatorio binomial consiste de n ensayos repetidos bajo las mismas condiciones, de modo que:

1. Los ensayos son independientes.
2. Cada ensayo produce dos posibles resultados mutuamente excluyentes, denominados éxito (E) y fracaso (F).
3. La probabilidad de obtener un éxito se denota por p y permanece constante. Por lo tanto, la probabilidad de fracaso, que es $1 - p = q$.

Bajo estas condiciones, se define la variable aleatoria binomial

La **Variable aleatoria binomial** se define como X : número de éxitos en n ensayos, o X : “número de éxitos en una muestra de tamaño n (con reposición)”. Los parámetros de la variable aleatoria binomial son:

- n : número de ensayos
- p es la probabilidad de éxito

La función de probabilidad es:

$$f(x) = P(X = x) = nCx p^x (1 - p)^{n-x}, \quad X = 0, 1, 2, \dots, n$$

La función de distribución de probabilidad es

$$F(x) = P(X \leq x) = \sum_X f(x), \quad x = 0, 1, 2, \dots, n$$

3.4.2 Medidas descriptivas de la variable aleatoria binomial

Si X es variable aleatoria discreta, se definen respectivamente la media y la varianza de la siguiente forma:

$$\mu = E(x) = np$$

$$\sigma^2 = Var(x) = npq$$

Ejemplo 3-6. Variables aleatorias binomiales

- X: Número de respuestas correctas en un examen de 10 preguntas independientes.
- X: Número de palabras nuevas escritas correctamente por un niño de un conjunto de 8 palabras.

Ejemplo 3-7. Un niño está aprendiendo la escritura de 5 palabras nuevas independientes. La probabilidad de escribir correctamente una palabra nueva es 0.75. Si se selecciona un niño al azar, determinar:

- a. La probabilidad de que escriba correctamente una palabra nueva.

$$P(X = 1) = 5C1 * 0.75^1 * 0.25^4$$

- b. La probabilidad de que escriba correctamente máximo una palabra nueva.

$$P(X \leq 1) = P(0) + P(1) =$$

$$5C0 * 0.75^0 * 0.25^5 + 5C1 * 0.75^1 * 0.25^4 =$$

- c. El número promedio de palabras nuevas escritas correctamente

$$\mu = np = 5 * 0.75 =$$

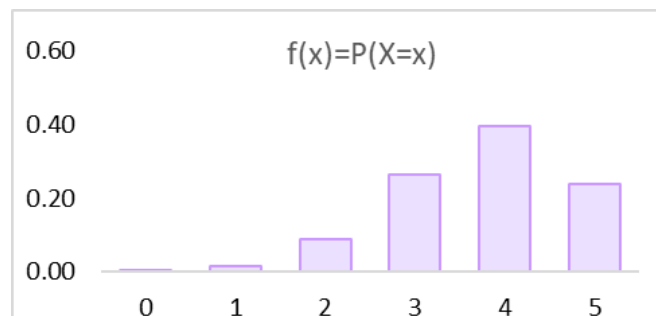
- d. La desviación estándar del número de palabras nuevas escritas correctamente.

$$\sigma^2 = npq = 5 * 0.75 * 0.25 =$$

- e. Realizar la gráfica de la función de probabilidad

Figura 3.6

Función de probabilidad



En la gráfica de $f(x)$ realizada con Excel, se observa sesgo a la izquierda, puesto que, la probabilidad de acertar 0,1,2 palabras es baja.

3.5 Distribución Hipergeométrica

Para comprender el significado de la variable aleatoria hipergeométrica, recuerde el ejemplo 3.4, donde se tomó la variable aleatoria X : número de adolescentes en una terna, teniendo dos grupos de personas para seleccionar la terna. Sin embargo, en la construcción de la función de probabilidad, se tomaron en cuenta los elementos de los dos grupos, los mismos que en cantidad se complementaron para conservar siempre el tamaño de la muestra $n = 3$. La variable aleatoria del ejemplo 3.4 es hipergeométrica.

3.5.1 Características del experimento aleatorio hipergeométrico

Un experimento aleatorio hipergeométrico consiste en n ensayos bajo las siguientes condiciones:

1. Los ensayos son dependientes (debido a la no reposición de los elementos extraídos).
2. Cada ensayo produce dos posibles resultados mutuamente excluyentes, denominados éxito (E) y fracaso (F).

La **Variable aleatoria hipergeométrica** se define como X : número de éxitos en n ensayos, o también, X : “número de éxitos en una muestra de tamaño n (sin reposición)”.

Los parámetros de la variable aleatoria son

- N : Tamaño de la población
- k : Número de éxitos en la población
- n : Tamaño de la muestra

La función de probabilidad es:

$$f(x) = P(X = x) = \frac{kCx(N-k)C(n-x)}{NCn}$$

$$X = 0, 1, 2, \dots, \min(k, n)$$

La función de distribución es:

$$F(x) = P(X \leq x) = \sum_x \frac{kCx(N-k)C(n-x)}{NCn}$$

$$X = 0, 1, 2, \dots, mn(k, n)$$

3.5.2 Medidas descriptivas para la variable aleatoria hipergeométrica

La media es:

$$\mu = E(x) = np$$

donde $p = \frac{k}{N}$ es la proporción de éxitos.

La varianza es:

$$\sigma^2 = Var(x) = npq * \frac{N-n}{N-1}$$

donde $q = 1 - p$ es la proporción de fracasos.

Ejemplo 3-8. Un Gimnasio, debe contratar a 6 entrenadores, para lo cual dispone de 20 candidatos finalistas, todos con el mismo nivel. Se conoce que 12 de candidatos son hombres y 8 son mujeres. Interesa determinar las probabilidades sobre el número de mujeres en el grupo seleccionado de forma aleatoria.

- a. ¿Qué tipo de experimento aleatorio corresponde a la información proporcionada?

Es un experimento aleatorio hipergeométrico, ya que no existe reposición de los candidatos una vez seleccionados, por lo tanto cada nueva selección afecta en probabilidad a la siguiente selección al disminuir en uno cada vez en el espacio muestral de candidatos. Es decir, se cumple la dependencia, pero en cada selección se mantienen dos posibles resultados: Hombre o Mujer.

- b. ¿Cuál es la variable aleatoria y sus parámetros? Variable aleatoria:

X : número de mujeres seleccionadas, con $X=0,1,\dots$, $\text{mín } n = 6, k = 8=6$

Parámetros

- $N = 20$ candidatos
- $k = 8$ candidatas mujeres
- $n = 6$ número de candidatos a ser seleccionados

- c. ¿Cuál es la probabilidad de que la selección no este formada por mujeres?

$$P(X = 0) = \frac{8C0 * 12C6}{20C8} = 0.0238$$

- d. ¿Cuál es la probabilidad de que la selección (muestra) esté formada solo por mujeres?

$$P(X = 6) = \frac{8C6 * 12C0}{20C8} = 0.0007$$

- e. ¿Cuál es la probabilidad de que la selección tenga una mujer?

$$P(X = 1) = \frac{8C1 * 12C5}{20C8} = 0.1635$$

- f. ¿Cuál es la probabilidad de que la selección contenga al menos una mujer?

Sin aplicar propiedades, se debe hallar:

$$P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6)$$

El número de operaciones a realizar resulta inconveniente, por lo que se recomienda aplicar las propiedades de la función de distribución $F(x)$, que en este caso corresponde al evento contrario.

$$P(X \geq 1) = 1 - P(X < 1) = P(X = 0)$$

$$1 - 0.0238 = 0.9762$$

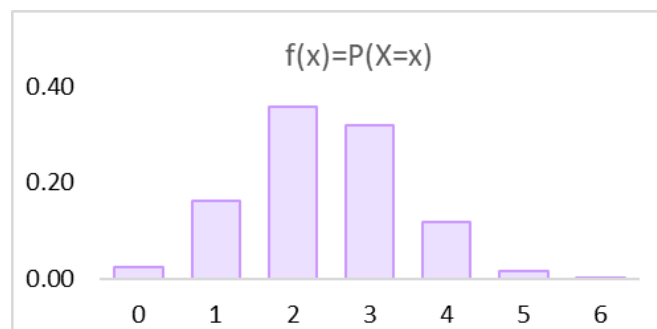
- g. ¿Cuál es el número promedio de mujeres en la selección?

$$\mu = np = n * \frac{k}{N} = 6 * \frac{8}{20} = 2.4$$

- h. Graficar $f(x)$

Figura 3.7

Función de probabilidad



3.6 Distribución Poisson

En la educación de niños y educación en general puede ser de interés determinar el número de ocurrencias de un evento en un determinado intervalo de tiempo, u otro tipo de intervalo. A este tipo de variable se le llama variable aleatoria Poisson.

Ejemplo 3-9. Variables definidas en un intervalo

- Número de niños que llegan a la formación cada minuto.
- Número de corredores que se hidratan cada kilómetro.
- Número de ocasiones en las que el niño realiza interrupciones verbales en la clase.
- Número de goles por partido de futbol
- Número de abdominales por minuto

3.6.1 Características del experimento aleatorio Poisson

1. El número de sucesos hallados en los intervalos son independientes entre sí.
2. El número promedio de sucesos por intervalo es proporcional al tamaño del intervalo.
3. La probabilidad de observar más de un suceso en un subintervalo tiende a cero.

Ejemplo 3-10. Se ilustra la característica 2

Si el número promedio de sucesos por intervalo es 2.5 unidades cuando el tamaño del intervalo es uno, ¿cuál sería el número promedio de sucesos por intervalo cuando el tamaño del intervalo es la mitad del intervalo? Se observa que hay una constante de proporcionalidad, que en este caso es la mitad; por lo tanto, el número promedio de sucesos por intervalo también será la mitad, es decir $\frac{2.5}{2} = 1.25$

La **Variable aleatoria Poisson** se define como X : “Número de sucesos por intervalo”, $X=0,1,2,\dots$. Donde $f(x)$ tiende a cero para algún valor de X en el dominio de la función.

Esta variable aleatoria tiene un solo parámetro

λ : Número promedio de suceso por intervalo.

La función de probabilidad es:

$$f(x) = P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

$$X = 0, 1, 2, \dots,$$

La función de distribución es:

$$F(x) = P(X \leq x) = \sum_x \frac{e^{-\lambda} \lambda^x}{x!}$$

$$x = 0, 1, 2, \dots$$

La variable aleatoria Poisson no depende del tamaño de la muestra, por tal motivo, no se indica cuál es el valor límite de la variable aleatoria X . Sin embargo, la gráfica es asintótica respecto al eje horizontal; eso significa que, a medida que la variable X crece, la probabilidad se acerca a cero. En la figura 3.8 se aprecia lo dicho.

Ejemplo 3-11. Gráficas de $f(x)$ y $F(x)$, con $\lambda = 1.25$

Como se puede observar en la gráfica de $f(x)$, cuando X crece, la función tiende a cero (la función es asintótica respecto al eje horizontal).

Figura 3.8

Función de probabilidad

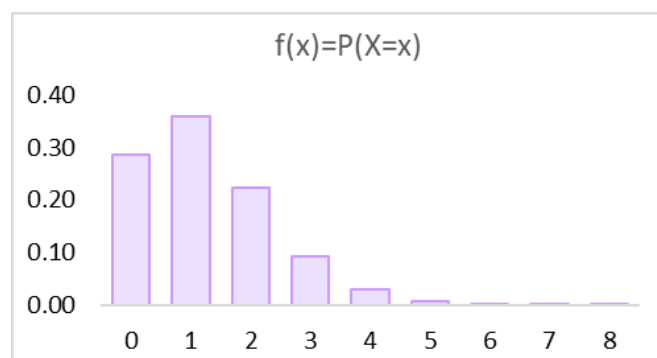
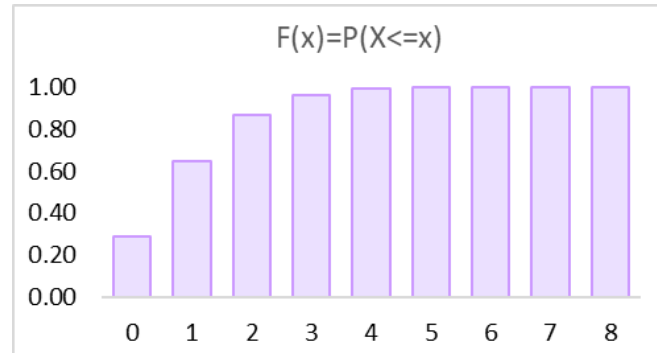


Figura 3.9*Función de probabilidad*

3.6.2 Medidas descriptivas para la variable aleatoria Poisson

La media y la varianza respectivamente:

$$\mu = E(x) = \lambda$$

$$\sigma^2 = Var(x) = \lambda$$

Ejemplo 3-12. Una maestra de primer grado, ha registrado que, su alumno interrumpe la clase en promedio 2 veces por hora. Estas interrupciones suelen ser: hablar o levantarse sin permiso. La maestra quiere conocer

- a. ¿Cuál es la probabilidad de que un niño interrumpa la clase 4 veces durante una hora?

$$P(X = 4) = \frac{e^{-2}2^4}{4!} = 0.0920$$

- b. ¿Cuál es la probabilidad de que un niño interrumpa la clase 4 veces durante media hora?

Puesto que, el intervalo de tiempo se ha reducido a la mitad, entonces el valor del parámetro cambia con la misma proporción, por lo tanto, el nuevo valor del parámetro es:

$$\lambda = \frac{2}{2} = 1$$

Y con este nuevo valor de λ se halla la probabilidad.

$$P(X = 4) = \frac{e^{-1}1^4}{4!} = 0.0153$$

- c. ¿Cuál es la probabilidad de que un niño interrumpa la clase 2 máximo dos veces durante 15 minutos?

Se debe hallar el parámetro para el nuevo intervalo. Ya que 15 minutos es

la cuarta parte de una hora, entonces:

$$\lambda = \frac{2}{4} = 0.5$$

Además, la probabilidad solicitada es acumulativa, por lo tanto, para este caso se aplica dicha definición:

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$P(X \leq 2) = \frac{e^{-0.5}0.5^0}{0!} + \frac{e^{-0.5}0.5^1}{1!} + \frac{e^{-0.5}0.5^2}{2!}$$

$$P(X \leq 2) = e^{-0.5} * (1 + 0.5 + 0.125)$$

$$P(X \leq 2) = 0.98561$$

Ejemplo 3-13. Un militar hombre debe cumplir con la prueba de abdominales cada año y se conoce que para cierto grupo etario debe realizar un promedio de 50 abdominales por cada minutos y medio. Al oficial a cargo de las pruebas le interesa conocer:

- a. ¿Cuál es la probabilidad de que un militar realice 56 abdominales por cada minuto y medio?

En este caso, el parámetro es $\lambda = 50$ cada 1.5 minutos.

$$P(X = 56) = \frac{e^{-50}50^{56}}{56!} = 0.0376$$

- b. ¿Cuál es la probabilidad de que un militar realice 30 abdominales en un minuto?

Se debe hallar un nuevo parámetro usando la regla de tres, de modo que si en 1.5 minutos, se realizan 50 abdominales, ¿Cuántas abdominales realiza en un minuto?

Aplicando una regla de tres se tiene: $\lambda = \frac{50}{1.5}$,

luego: $\lambda = \frac{50}{1.5} = 33.33$

$$P(X = 30) = \frac{e^{-33.33}33.33^{30}}{30!} = 0.0611$$

- c. ¿Si el intervalo es de 1 minuto, cuál es la varianza?

$$\sigma^2 = Var(x) = \lambda = 33.33$$

3.7 Distribución normal

3.7.1 Introducción

La distribución normal es un modelo matemático que permite estudiar el comportamiento de una variable aleatoria X , cuya función de densidad tiene la forma de una campana, conocida como la Campana de Gauss, en honor al matemático Carl Friedrich Gauss (1777-1855), quien desarrolló esta función. Una variable normal tiene como parámetros la media poblacional μ y la varianza poblacional σ^2 , y se denota por: $X \sim N(\mu, \sigma^2)$.

Observación. Su uso depende del conocimiento de sus parámetros. Más adelante, se contará con la distribución t-Student como alternativa a la distribución normal cuando no se conoce la desviación estándar σ .

3.7.2 Funciones de densidad y distribución

Si X es una variable aleatoria normal, su función de densidad es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

con $-\infty < X < \infty$, $e=2.71828\dots$, $\pi = 3.14159$

La función de densidad $f(x)$ con gráfica representada en la figura 3.10 tiene tres características:

- La gráfica de la función de densidad $f(x)$ es simétrica.
- La gráfica de la función de densidad $f(x)$ es mesocúrtica.
- $f(x)$ tiende a cero cuando X tiende a infinito.

Figura 3.10

Función de densidad



Si X es una variable aleatoria normal, su función de distribución es:

$$F(x) = P(X \leq x)$$

Para calcular esta probabilidad acumulativa, se debe usar la integral, que no es el propósito de este curso por lo que se utiliza un cambio de variables para determinar las probabilidades mediante una tabla estandarizada para cualquier variable.

3.7.3 Variable aleatoria normal estándar

Si X es una variable aleatoria normal con media μ_x y desviación estándar σ_x , se define la variable aleatoria

$$Z = \frac{X - \mu_X}{\sigma_X}$$

denominada variable aleatoria normal estándar con media $\mu_z = 0$ y desviación estándar $\sigma_z = 1$.

Esta variable resulta de una traslación de la media hacia el origen de un nuevo sistema de referencia, donde la función de densidad es $f(z)$ y la función de distribución $F(z) = P(Z \leq z)$, para la cual se define una tabla de probabilidades acumulativas que se encuentra en el Anexo B.

3.7.4 Distribución normal estándar

Si Z es una variable aleatoria normal estándar con media 0 y varianza 1, se denota por $Z \sim N(0, 1)$ y su función de distribución es:

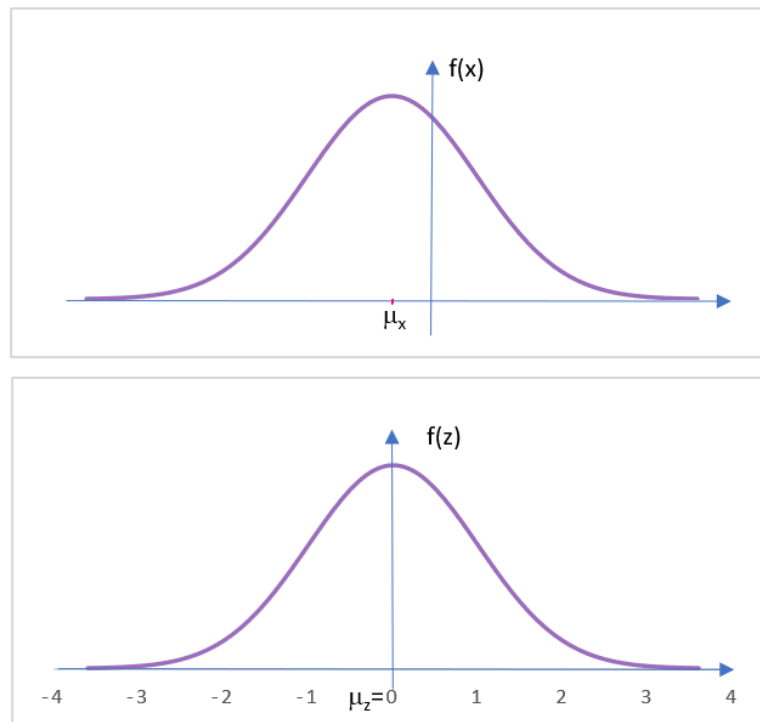
$$F(z) = P(Z \leq z)$$

La utilidad de esta nueva función radica en la facilidad de determinar la probabilidad para la variable aleatoria X con ayuda de la variable aleatoria normal estándar Z , ya que las probabilidades son iguales, pero se hallan en un nuevo sistema de referencia. De modo que:

$$F(x) = P(X \leq x) = P(Z \leq z) = F(z)$$

$F(x)$ tiene dos propiedades basadas en la simetría de la distribución.

- i. $P(Z < z) = 1 - P(Z \leq z) = P(Z \geq -z)$
- ii. $P(-z \leq Z \leq z) = 1 - 2P(Z \leq -z) = 2P(Z \leq z) - 1$

Figura 3.11*Funciones de densidad $f(x)$ y $f(z)$* 

3.7.5 Tabla de distribución acumulativa $F(z)$

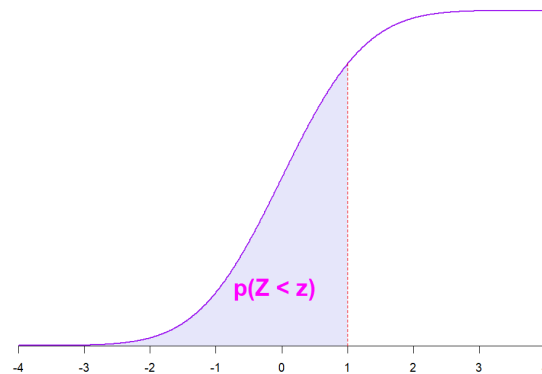
La tabla de distribución $F(z)$ se ha construido justamente para probabilidades en cola izquierda, es decir, $P(Z < z)$. Luego, las probabilidades en cola izquierda se hallan directamente en la tabla; mientras que para las probabilidades en cola derecha debe aplicarse la propiedad de evento contrario o la propiedad de simetría. La tabla $F(z)$ del Anexo B se ha construido con valores de Z , desde -3.69 hasta 3.69. Para valores fuera de este rango, las probabilidades pueden ser cero o uno, dependiendo de si se pide a la derecha o a la izquierda de estos valores extremos.

Observación. En los ejemplos que se presentan a continuación, se utiliza la representación gráfica del problema, pero en lugar de usar la función de distribución $F(x)$ o $F(z)$, se usa simplemente la función de densidad para comprender mejor la aplicación de las propiedades.

Ejemplo 3-14. Si Z es una variable aleatoria normal estándar con media 0 y varianza 1, hallar las siguientes probabilidades:

Figura 3.12

Función de distribución $F(z)$

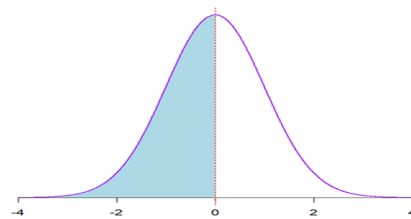


a. $P(Z < 0)$

La media de Z es cero, ubicada en el medio del eje horizontal; por lo tanto, la probabilidad a la izquierda de cero es 0.5

Figura 3.13

$P(Z < 0)$



b. $P(Z < -4)$

En la figura 3.13, se ha representado el valor mínimo de Z usado en las tablas, luego, $p(Z < -4)$ es prácticamente cero.

c. $P(Z < 1.96)$

En la tabla de la distribución $F(z)$ y en las filas de la primera columna de Z positivos se busca el valor $Z = 1.9$ y desde la segunda columna, se busca la centésima, es decir, 0.06, como se observa en la figura 3.14 .

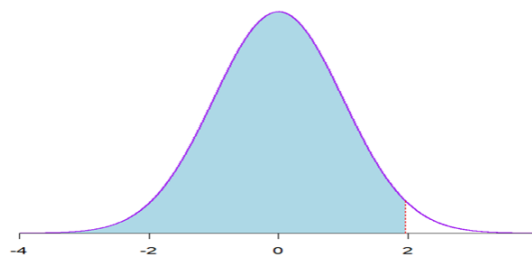
Luego, la probabilidad buscada corresponde al cruce entra la fila y columna indicadas:

$$P(Z < 1.96) = 0.975$$

La representación gráfica se observa en la figura 3.15.

Figura 3.14*Parte de la Tabla $F(z)$*

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558

Figura 3.15 $P(Z < 1.96)$ d. $P(Z > 1.2)$

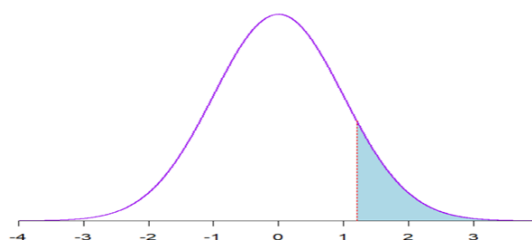
Para hallar esta probabilidad, es necesario aplicar propiedades para expresar la probabilidad en términos de cola izquierda, ya que la tabla $F(z)$ no proporciona probabilidades en cola derecha; por lo tanto,

Si se aplica la propiedad de evento contrario, se tiene:

$$P(Z > 1.2) = 1 - P(Z < 1.2) = 1 - 0.88493 = 0.11507$$

Si se aplica la propiedad de simetría se tiene el mismo resultado:

$$P(Z > 1.2) = P(Z < -1.2) = 0.11507$$

Figura 3.16 $P(Z > 1.2)$ 

e. $P(-1.2 < Z < 1.96)$

Al aplicar propiedades, se tiene:

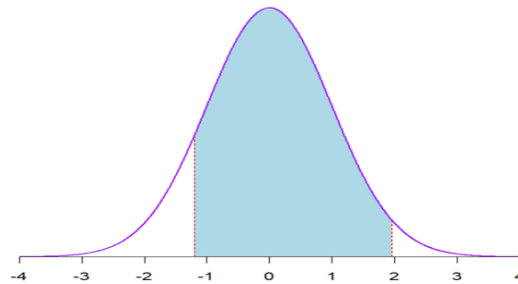
$$P(-1.2 < Z < 1.96) = P(Z < 1.96) - P(Z < -1.2)$$

$$P(-1.2 < Z < 1.96) = 0.97500 - 0.11507 = 0.85493$$

Este resultado se representa en la campana de Gauss como sigue:

Figura 3.17

$$P(-1.2 < Z < 1.96)$$



Ejemplo 3-15. Si Z es una variable aleatoria normal estándar con media 0 y varianza 1, hallar los puntos críticos correspondientes a la probabilidad en la distribución normal estándar. Es decir Z tal que cumpla con la probabilidad dada.

En este caso, se busca la probabilidad en la tabla, y luego se determina el valor de Z correspondiente en la fila y columna de la probabilidad. No siempre se tiene el valor exacto de la probabilidad dada, por ello, se realiza el promedio de los valores de Z correspondientes a las probabilidades que se acerquen tanto por la izquierda como por la derecha. Recuerde que, antes de hallar el valor de Z , la probabilidad debe estar expresada en términos de $F(z) = P(Z < z)$.

a. $P(Z > z) = 0.025$

Se puede aplicar la propiedad de simetría, pero debe considerar el signo de Z . Pero en este caso resulta conveniente aplicar la propiedad de evento contrario, para expresar en términos de $<$

$$P(Z > z) = 1 - P(Z < z)$$

$$0.025 = 1 - P(Z < z)$$

$$P(Z < z) = 1 - 0.025$$

$$P(Z < z) = 0.975$$

El valor correspondiente de Z es único: $Z = 1.96$, su interpretación gráfica corresponde a la figura 3.15, ya que se ha realizado el proceso inverso del ejemplo 3.14, literal c.

b. $P(Z < z) = 0.95$

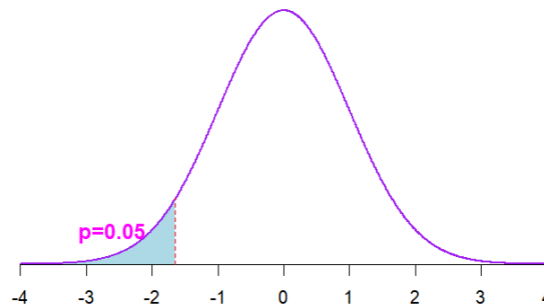
Puesto que la probabilidad dada es de cola izquierda ($<$), se procede a buscar directamente el valor de Z correspondiente en la tabla. Sin embargo, no existe un valor único de probabilidad, como se observa en la figura 3.17, teniendo dos valores de probabilidad que se acercan tanto por la izquierda 0.94950 como por la derecha 0.95053, por lo tanto, se tienen dos valores de Z correspondientes 1.64 y 1.65 con los que se halla el promedio para obtener el valor de Z crítico buscado.

$$Z = \frac{1.64 + 1.65}{2} = 1.645$$

La representación gráfica del problema muestra la probabilidad en la cola izquierda correspondiente al valor de Z o punto crítico hallado.

Figura 3.18

$$P(Z < z) = 0.05$$



Ejemplo 3-16. Educación inicial. Las funciones cognitivas son de vital importancia en el desarrollo de un niño. Para niños entre 0 y 42 meses existe la prueba desarrollada por quien mediante un test ha establecido que el valor promedio de la prueba es 100 puntos con una desviación estándar de 15 puntos.

Los estadísticos proporcionados se asumen como los parámetros teóricos (población) con media poblacional $\mu = 100$ y desviación estándar $\sigma = 15$. Si se aplica la prueba o test a un niño al azar, determine lo siguiente:

- a. La probabilidad de que el niño tenga una puntuación inferior a la media. No es necesario hallar el valor estandarizado z correspondiente a la media, debido a que la media se encuentra en el centro de la distribución, su probabilidad es 0.5.
- b. La probabilidad de que el niño tenga una puntuación inferior a 90 puntos. El valor de $x = 90$, y los parámetros media poblacional $\mu = 100$ y desviación estándar $\sigma = 15$ se reemplazan en la fórmula de la variable aleatoria normal estándar Z , de modo que:

$$Z = \frac{x - \mu}{\sigma}$$
$$Z = \frac{90 - 100}{15} = -0.67$$

Luego se calcula la probabilidad correspondiente al valor de z , es decir:

$$P(X \leq x) = P(Z \leq z)$$
$$P(X \leq 90) = P(Z \leq -0.67)$$
$$P(Z \leq -0.67) = 0.25143$$

El resultado obtenido se puede leer como sigue: la probabilidad de que un niño obtenga una puntuación inferior a 90 puntos es 0.25143, o también se puede escribir esta interpretación en términos de porcentaje, luego, el 25.14 % de los niños obtiene puntuaciones inferiores a 90 puntos.

- c. La probabilidad de que el niño tenga puntuaciones entre 90 y 110. Se debe hallar

$$P(90 \leq X \leq 110)$$

Previamente determine los valores estandarizados z_1 y z_2 correspondientes a 90 y 110.

$$Z_1 = \frac{90 - 100}{15} = -0.67$$
$$Z_2 = \frac{110 - 100}{15} = 0.67$$

Estos valores se reemplazan en la probabilidad, de modo que se tiene:

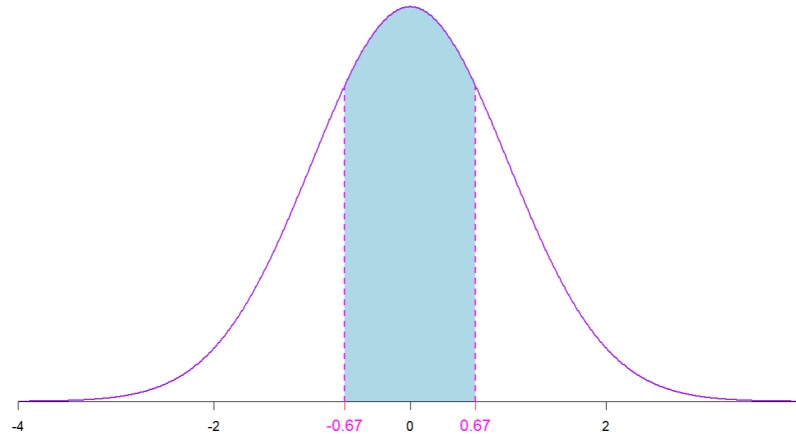
$$P(90 \leq X \leq 110) = P(-0.67 \leq Z \leq 0.67) =$$

Luego se aplica la regla de probabilidad 4 para cualquier variable continua,

a pesar de que también se puede aplicar la propiedad de simetría 2 para la variable Z , como se observa en la figura 3.19, ya que los valores de z hallados son simétricos.

Figura 3.19

Probabilidad entre dos valores



$$\begin{aligned} P(-0.67 \leq Z \leq 0.67) &= P(Z \leq 0.67) - P(Z \leq -0.67) \\ &= 0.74857 - 0.25143 = 0.49714 \end{aligned}$$

La probabilidad de que un niño obtenga una puntuación entre 90 y 110 es 0.49714. En términos de porcentaje, el 49.71 % de los niños tienen puntuaciones entre 90 y 110.

- d. ¿Cuál es la puntuación a la que máximo llega el 95 % de los niños?

Ahora se debe hallar un valor de x que corresponde a la probabilidad, que también corresponde a un valor de z en la distribución normal estándar, de modo que: La probabilidad solicitada es una probabilidad a la izquierda de x

$$P(X < x) = 0.95$$

$$P(Z < z) = 0.95$$

El valor de z correspondiente a esa probabilidad en cola izquierda es 1.645 que fue hallado en el ejemplo anterior. Este valor de z se reemplaza en la fórmula de la variable aleatoria normal estándar Z para despejar el valor de x .

$$Z = \frac{x - \mu}{\sigma}$$

$$1.645 = \frac{x - 100}{15}$$

$$1.645 * 15 = x - 100$$

$$x = 100 + 1.645 * 15 = 124.675$$

La puntuación máxima a la que llega el 95 % de los niños es 124.675 puntos.

Ejemplo 3-17. Actividad Física. La fuerza pico de reacción del suelo (GRF) o recepción con salto con unidad de medida múltiplos del peso corporal (BW). La variable que se estudia en deportistas mujeres sub 18 en las mismas condiciones de superficie e y ejecución sigue una distribución normal con parámetros media $\mu = 3.2$ y varianza $\sigma^2 = 0.5^2$.

Figura 3.20

La fuerza pico de reacción



Imagen generada con IA

Se han determinado las categorías de la fuerza de reacción del suelo en la siguiente tabla.

Con $X \sim \mathcal{N}(\mu = 3.2, \sigma^2 = 0.5^2)$, determinar las probabilidades y el valor de la fuerza que se indica a continuación.

- ¿Cuál es la probabilidad de que la fuerza pico de reacción del suelo en una recepción sea baja? De acuerdo a las categorías definidas, esto se traduce como:

$$P(X < 2.5)$$

$$Z = \frac{x - \mu}{\sigma}$$

Tabla 3.4*Fuerza pico de reacción del suelo [BW]*

Fuerza Pico	Categorías [BW]
Baja carga en recepción	< 2.8
Zona Aceptable	$2.8 < X < 3.6$
Carga superior	> 3.6

$$Z = \frac{2.8 - 3.2}{0.5} = -0.80$$

Luego se calcula la probabilidad correspondiente al valor de z , es decir:

$$P(X \leq x) = P(Z \leq z)$$

$$P(X \leq 2.8) = P(Z \leq -0.80)$$

$$P(Z \leq -0.80) = 0.21185$$

- b. Determine la probabilidad de que la carga de recepción sea aceptable

$$P(2.8 \leq X \leq 3.6)$$

se hallan los valores estandarizados z_1 y z_2 correspondientes a 90 y 110.

$$Z_1 = \frac{2.8 - 3.2}{0.5} = -0.80$$

$$Z_2 = \frac{3.6 - 3.2}{0.5} = 0.80$$

Estos valores se reemplazan en la probabilidad, de modo que se tiene:

$$P(2.8 \leq X \leq 3.2) = P(-0.80 \leq Z \leq 0.80) =$$

$$= 0.788145 - 0.21185 = 0.56960$$

La probabilidad de obtener una carga de recepción aceptable es 0.56960.

- c. Determine el porcentaje de que la carga de recepción sea superior a 3.6. Para hallar $P(X > 3.6)$, usar Z , donde a $x = 3.6$ le corresponde $z = 0.8$ que se halló en el literal anterior; por lo tanto

$$P(Z > 0.8) = P(Z < -0.8) = 0.21185$$

El 21.185 % de las deportistas supera una carga de recepción de 3.6 (BW)

- d. Encuentre el valor de la carga de recepción x que es superado en el 10 % de las ocasiones (**Umbral de alto impacto 10 %**).

Se debe hallar x tal que $P(X > x) = 0.10$, para lo cual se usa la distribución normal estándar, con la cual a la misma probabilidad le corresponde un valor de z ; de modo que $P(Z > Z) = 0.10$. Luego, aplicando propiedades, se tiene:

$$1 - P(Z < Z) = 0.10$$

$$P(Z < Z) = 0.90$$

Esta probabilidad del 0.90 se busca en la tabla de distribución $F(z)$ del Anexo B, pero en este caso no existe el valor exacto; por lo que, para hallar z , se realiza un promedio de los valores z correspondientes a las probabilidades que se aproximan a 0.90, por la izquierda y por la derecha:

$$Z = \frac{1.28 + 1.29}{2} = 1.285$$

Dicho valor se reemplaza en la formula de la variable aleatoria normal estándar Z , para luego despejar x

$$Z = \frac{x - \mu}{\sigma}$$

$$1.285 = \frac{x - 3.2}{0.5}$$

$$1.285 * 0.5 = x - 3.2$$

$$x = 3.8425$$

El umbral de alto impacto del 10 % para deportistas de voley sub 18 es 3.8425 BW. O dicho de otro modo, el 10 % de los deportistas superan una carga de recepción de 3.8425 BW.

Figura 3.21

Z crítico para $p=0.1$



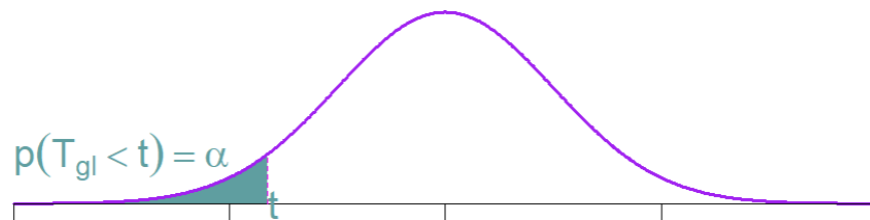
3.8 Distribución t-Student

3.8.1 Introducción

La **distribución t de Student**, desarrollada por William Sealy Gosset bajo el seudónimo "Student" se utiliza cuando se desconoce la desviación estándar poblacional σ .

Figura 3.22

Distribución t-Student



3.8.1.1 CARACTERÍSTICAS DE LA GRÁFICA DE LA DISTRIBUCIÓN $F(t)$

Su gráfica tiene características similares a la distribución normal:

- Simétrica.
- Tiene mayor dispersión o colas más anchas para $n < 30$.
- Asintótica.

A medida que se incrementa el tamaño de la muestra, la distribución t-Student se aproxima a la distribución normal.

3.8.1.2 PARÁMETROS DE LA DISTRIBUCIÓN $F(t)$

- Grados de libertad: $g.l.$, que dependen del tamaño de la muestra y las restricciones.
- Significancia o probabilidad α

3.8.2 Tabla de distribución t-Student

El diseño de la tabla de distribución $F(t)$, siendo t una variable estandarizada, tiene una cantidad limitada de valores de probabilidad (significancia), pero sí una cantidad considerable de puntos críticos. La tabla impresa se encuentra en el Anexo C y cuenta con probabilidades tanto en la cola derecha como en la cola izquierda. Una parte de esta tabla se observa en la figura 3.23

Figura 3.23*Parte de la Tabla distribución t-Student*

	α						
g.l	0.0005	0.001	0.0025	0.005	0.01	0.015	0.02
1	-636.619	-318.309	-127.321	-63.657	-31.821	-21.205	-15.895
2	-31.599	-22.327	-14.089	-9.925	-6.965	-5.643	-4.849
3	-12.924	-10.215	-7.453	-5.841	-4.541	-3.896	-3.482
4	-8.610	-7.173	-5.598	-4.604	-3.747	-3.298	-2.999
5	-6.869	-5.893	-4.773	-4.032	-3.365	-3.003	-2.757

Ejemplo 3-18. Para la variable t-Student determine:

- a. Probabilidad de que t sea menor que -5.841 con 3 grados de libertad: g.l

En la figura 3.23, ubique la fila g.l=3 y $t=-5.841$; luego, en la columna correspondiente, se halla la probabilidad, de modo que:

$$P(T_3 < -5.841) = 0.005$$

- b. Probabilidad de que t sea menor que -3 con 5 grados de libertad. En la g.l=5 de la figura 3.23 no existe $t = -3$, pero si se encuentran dos valores que se acercan por la izquierda y por la derecha, con probabilidades correspondientes: 0.015 y 0.020, por lo tanto:

$$-3.003 < T_5 < -2.757$$

$$0.015 < P(T_5 < t) < 0.020$$

Este procedimiento es de utilidad cuando solamente se dispone de la tabla de distribución. Sin embargo, en la siguiente sección esta probabilidad se halla fácilmente con las funciones de Excel.

- c. Hallar el valor de t , de modo que la probabilidad a la izquierda de t con 4 grados de libertad es 0.015

$P(T_4 < t) = 0.015$, corresponde a un valor de t o punto crítico $t=-3.298$, como se observa en la figura 3.23 dado por el cruce de la fila de 4 grados y la columna de probabilidad = 0.015. Es decir, $t=-3.298$ satisface a:

$$P(T_4 < t) = 0.015$$

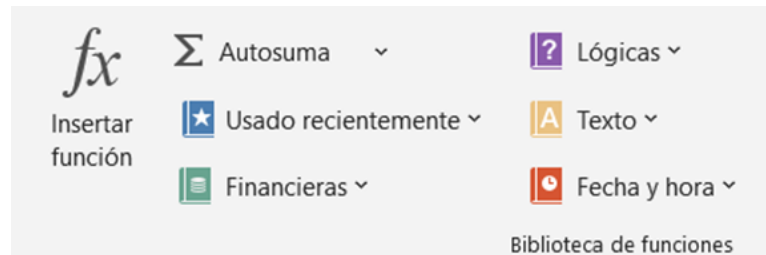
3.9 Laboratorio de variables aleatorias

3.9.1 Distribuciones de probabilidad discretas y continuas con Excel

Mediante el ícono de funciones de Excel figura 3.24, se puede acceder al Menú, en el cual se selecciona la categoría Estadísticas .

Figura 3.24

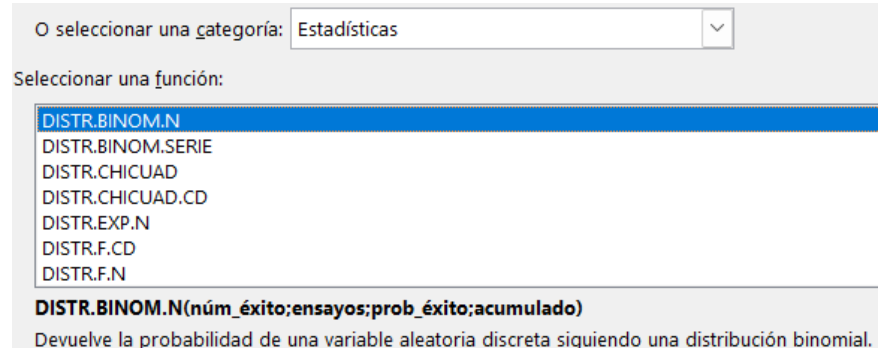
Icono de Funciones de Excel



Luego, se seleccionan las funciones de probabilidad y sus funciones inversas del menú de la figura 3.25 según se requiera

Figura 3.25

Menú Estadísticas Distribuciones



Ejemplo 3-19. Si X es una variable aleatoria binomial con parámetros $n = 8$ ensayos, y probabilidad de éxito $p = 0.15$, determine:

- La probabilidad de encontrar 4 éxitos, es decir $P(X = 4)$

Con las opciones mostradas en la figura 3.25 se elige la distribución binomial para ingresar los datos de los parámetros, según se observa en la figura 3.26. En este caso, es importante indicar que el valor lógico es cero, usado para valores no acumulativos de la probabilidad, es decir, cuando se desea determinar $P(X = x)$, donde $x = 4$. En la misma imagen se observa el resultado de la probabilidad: $P(X = 4) = 0.01850$

Figura 3.26

Menú Estadísticas Distribuciones

Argumentos de función

DISTR.BINOM.N

Núm_éxito	4	=	4
Ensayos	8	=	8
Prob_éxito	0.15	=	0.15
Acumulado	0	=	FALSO

= 0.018498596

Devuelve la probabilidad de una variable aleatoria discreta siguiendo una distribución binomial.

- b. La probabilidad de encontrar 4 éxitos o menos, es decir $P(X \leq 4)$

Esta probabilidad acumulativa, se puede realizar de dos formás: Hallar la suma de probabilidades puntuales en Excel, tomando como valor acumulado 0

$$P(X \leq 4) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$$

O más bien, usando la función de distribución, pero en el campo de acumulado reemplazar 1 como indicador de verdadero de la figura 3.26, como se observa en la figura 3.27

Figura 3.27

Menú Estadísticas: Distribuciones

DISTR.BINOM.N

Núm_éxito	D2	=	4
Ensayos	8	=	8
Prob_éxito	0.15	=	0.15
Acumulado	1	=	VERDADERO

= 0.997146126

$$P(X \leq 4) = 0.99715$$

- c. La probabilidad de encontrar más de 4 éxitos, es decir $P(X > 4)$

Esta probabilidad acumulativa de cola derecha, se puede hallar con: $P(X > 4) = P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8)$

O simplemente aplicando la propiedad de evento contrario aprovechado el resultado anterior, es decir: $1 - \text{DISTR.BINOM.N}(D2;8;0.15;1)$, luego

$$P(X > 4) = 1 - P(X \leq 4) = 0.00285$$

Ejemplo 3-20. Si X es una variable aleatoria Hipergeométrica con parámetros: número de elementos en la población $N = 25$, número de éxitos en la población $k = 5$ y número de elementos en la muestra $n = 4$, determine

a. $P(X = 3)$

Desde el mismo menú de funciones, categoría Estadísticas y al seleccionar la función correspondiente a la distribución hipergeométrica se genera el menú de la figura 3.27 donde se ingresan el valor de x , los parámetros y el criterio de acumulación 1 y no acumulación 0 que corresponde a este ítem. Y el resultado es $P(X = 3) = 0.01581$

Figura 3.28

Menú distribución hipergeométrica

DISTR.HIPERGEOM.N			
Muestra_éxito	3		= 3
Núm_de_muestra	4		= 4
Población_éxito	5		= 5
Núm_de_población	25		= 25
Acumulado	0		= FALSO
			= 0.015810277

b. $P(X \leq 3)$

Se eligen el procedimiento del ítem anterior y solamente se ingresa en acumulado el valor 1, por lo tanto:

$$P(X \leq 3) = 0.99960$$

c. $P(X \geq 3)$

Solo se aplicará la definición de esta probabilidad para hallarla en Excel, es decir, $P(X \geq 3) = P(X = 3) + P(X = 4)$ En Excel solo se halla la probabilidad que falta $P(X = 4)$, luego,

$$P(X \geq 3) = P(X = 3) + P(X = 4) = 0.01581 + 0.00040 = 0.01621$$

- d. Realizar la gráfica de la función de probabilidad para los parámetros indicados y en el rango de X .

Se crea la tabla de datos representando X en la primera columna con valores desde 0 hasta 3, y en la segunda columna la distribución hipergeométrica ingresando los parámetros y ligando los valores de x de cada celda, dando lugar a la tabla de Excel figura 3.29.

Figura 3.29

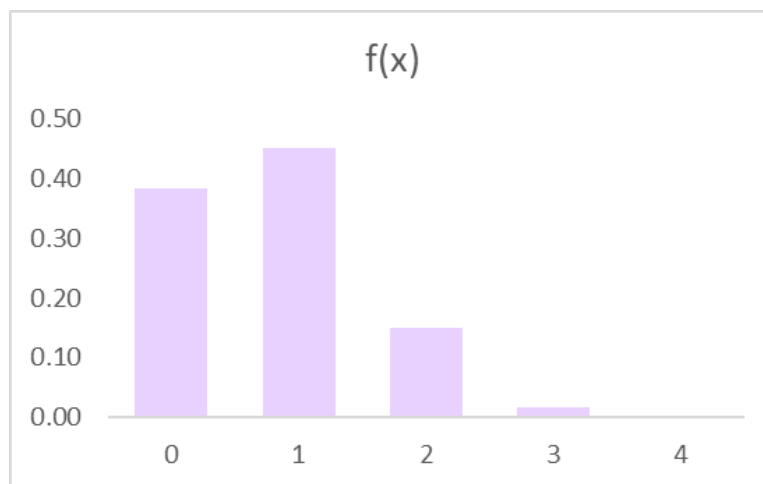
Valores x y $f(x)$

x	$f(x)$
0	0.3830
1	0.4506
2	0.1502
3	0.0158
4	0.0004

Marcar la columna de $f(x)$, dar click derecho, e insertar el gráfico de columnas, antes de finalizar, dar click en el eje horizontal y modificar las etiquetas del eje horizontal, eligiendo la columna X con los valores desde 0 hasta 4; obteniendo la figura 3.30

Figura 3.30

Función de probabilidad $f(x)$



Ejemplo 3-21. X es variable aleatoria Poisson con parámetro λ

Para hallar probabilidades puntuales o acumuladas, se procede de forma similar a lo realizado con las otras distribuciones. Solo recuerde que ésta función es asintótica; es decir, no depende del tamaño de la muestra, de modo que el máximo valor de x se presenta cuando $f(x)$ se hace cero o tiende a cero para una cantidad de decimales elegidos.

Ejemplo 3-22. X es variable normal con parámetros la media poblacional $\mu = 14.5$ y desviación estándar $\sigma = 1.5$, determine

a. $P(X \leq 18)$

Al usar Excel, no es necesario pasar por el cálculo de la variable aleatoria normal estándar; por lo tanto, se obtiene el valor de la probabilidad directamente. Luego, $P(X \leq 18) = 0.99902$

Figura 3.31

Menú de la distribución normal $F(x)$

DISTR.NORM.N				
X	18	↑	=	18
Media	14.5	↑	=	14.5
Desv_estándar	1.5	↑	=	1.5
Acumulado	1	↑	=	VERDADERO
				= 0.990184671

- b. Hallar x tal que $P(X \leq x) = 0.1$

Ahora, se debe elegir la función inversa normal en Excel, cuyo menú es: Como se observa en la captura del Menú de Excel, el resultado es 12.58.

Figura 3.32

Menú de Inversa normal $F(x)$

INV.NORM				
Probabilidad	0.1	↑	=	0.1
Media	14.5	↑	=	14.5
Desv_estándar	1.5	↑	=	1.5
				= 12.57767265

Ejemplo 3-23. Para ciertos procesos, de la Estadística Inferencial, solo se requiere el uso de la variable normal estándar Z , la misma que tienen parámetros media 0 y desviación estándar 1.5. Para esta variable determine:

- a. $P(Z \leq 1.96)$

Una vez más, mediante el ícono de funciones, categoría Estadística, seleccionar distribución normal estándar e ingresar los datos de acuerdo a la figura 3.33. En este caso, también se ingresar el valor acumulado 1.

Luego, $P(X \leq 1.96) = 0.97500$

Figura 3.33Menú Distribución normal estándar $F(z)$

DISTR.NORM.ESTAND.N			
Z	1.96	↑	= 1.96
Acumulado	1	↑	= VERDADERO
			= 0.975002105

- b. Hallar z tal que $P(X \leq x) = 0.1$

Ahora, se debe elegir la función inversa normal estándar en Excel, cuyo menú se observa en la figura 3.34, donde a pesar de que todas las distribuciones continuas son acumulativas, en este caso no se pide el valor de acumulados:

Figura 3.34

Menú de Inversa norma estándar

INV.NORM.ESTAND			
Probabilidad	0.1	↑	= 0.1
			= -1.281551566

Ejemplo 3-24. Probabilidades y puntos críticos en la distribución t-Student

Excel dispone de las funciones de distribución para hallar la probabilidad en cola izquierda, derecha y dos colas, algo que se entenderá mejor más adelante con el estudio de la estadística inferencial. Además, dispone de la función inversa T que permite hallar los puntos críticos correspondientes a una probabilidad dada, brindando opciones para cola izquierda y dos colas; por lo que, para la cola derecha se aplicarán las propiedades de simetría usadas para la distribución normal.

A continuación, se presenta la solución usando Excel de los ejemplos realizados en la sección 3.8.1, para lo cual solo se muestra una captura de pantalla con los parámetros grados de libertad y, según corresponda, el valor del punto crítico T o la significancia α

- a. $P(T_3 < -5.841)$

Observe el resultado en la figura 3.35

- b. $P(T_5 < -3)$

Observe el resultado en la figura 3.36

Figura 3.35*Distribución T en cola izquierda*

DISTR.T.N			
X	-5.841		= -5.841
Grados_de_libertad	3		= 3
Acumulado	1		= VERDADERO
			= 0.004999782

Devuelve la distribución t de Student de cola izquierda.

Figura 3.36*Distribución T en cola izquierda*

DISTR.T.N			
X	-3		= -3
Grados_de_libertad	5		= 5
Acumulado	1		= VERDADERO
			= 0.015049624

Devuelve la distribución t de Student de cola izquierda.

c. $P(T_4 < t) = 0.015$

Observe el resultado en la figura 3.37

Figura 3.37*Punto crítico T en cola izquierda*

INV.T			
Probabilidad	0.015		= 0.015
Grados_de_libertad	4		= 4
			= -3.297629728

Devuelve el inverso de cola izquierda de la distribución t de Student.

- d. Hallar el número de grados de libertad en la distribución t-Student de modo que T y Z sean aproximadamente iguales y que corresponden al área en cola derecha de 0.025 Se puede verificar que, $P(Z < z) = 0.025$ corresponde a un valor $Z = -1.96$

Pero en la tabla de la distribución t-Student, para esa misma probabilidad le corresponden tantos valores de t como grados de libertad tiene la tabla del Anexo C.

En este sentido, se muestra la tabla de la distribución t-Student con una significancia $\alpha = 0.025$, con una primera parte con las cuatro primeras filas de grados de libertad, como se observa en la figura 3.38. los valores de T están bastante alejados del valor de Z .

Figura 3.38

Valores críticos T hasta $gl=4$

g.l	α						
	0.01	0.015	0.02	0.025	0.04	0.05	0.1
1	-31.821	-21.205	-15.895	-12.706	-7.916	-6.314	-3.078
2	-6.965	-5.643	-4.849	-4.303	-3.320	-2.920	-1.886
3	-4.541	-3.896	-3.482	-3.182	-2.605	-2.353	-1.638
4	-3.747	-3.298	-2.999	-2.776	-2.333	-2.132	-1.533

En la figura 3.39 se observa cuanto más se incrementan los grados de libertad, el valor crítico T tiende a acercarse (converge) hacia el punto crítico Z .

Figura 3.39

Valores críticos T para gl grandes

g.l	α						
	0.01	0.015	0.02	0.025	0.04	0.05	0.1
70	-2.381	-2.215	-2.093	-1.994	-1.776	-1.667	-1.294
80	-2.374	-2.209	-2.088	-1.990	-1.773	-1.664	-1.292
90	-2.368	-2.205	-2.084	-1.987	-1.771	-1.662	-1.291
100	-2.364	-2.201	-2.081	-1.984	-1.769	-1.660	-1.290
300	-2.339	-2.180	-2.063	-1.968	-1.757	-1.650	-1.284
500	-2.334	-2.176	-2.059	-1.965	-1.754	-1.648	-1.283
1000	-2.330	-2.173	-2.056	-1.962	-1.752	-1.646	-1.282

Finalmente, en este caso se puede concluir que Z y T son aproximadamente iguales para una probabilidad en cola izquierda de 0.025 para 1000 grados de libertad.

$$Z = -1.96 \approx T = -1.962$$

$$P(Z < z) = 0.025 = P(T_{gl} < t)$$

3.10 Ejercicios del capítulo

3.1. Educación inicial. Recuerde el Test desarrollado por Bayley-III sobre el desarrollo de las funciones cognitivas, en el que se determina una puntuación media de 100 y una desviación estándar de 15 para niños entre 1 y 42 meses. Este estudio establece una categorización del desarrollo de acuerdo a las puntuaciones de la Tabla 3.5

Tabla 3.5

Puntuaciones en el Test Bayley III

Categoría	Rango de puntuaciones
Extremadamente bajo	≤ 69
Límite (Borderline)	70-79
Promedio bajo	80-89
Promedio	90-109
Promedio alto	110-119
Superior	120-129
Muy superior	≥ 130

Con la ayuda de esta tabla, determine el porcentaje de niños con:

- Puntuaciones Extremadamente bajas
- Puntuaciones en el Límite
- Puntuaciones Promedio bajo
- Puntuaciones Promedio
- Puntuaciones Promedio alto
- Puntuaciones Superior
- Puntuaciones Muy superior

3.2. Continuación del ejercicio anterior.

- Determine el valor de la puntuación de modo que solo el 1 % de los niños alcancen esa puntuación.

- b. Determine el valor de la puntuación de modo que solo el 1 % de los niños superen esa puntuación.

3.3. Educación Básica. En una clase de $n = 24$ niños, la meta de lectura diaria es de 20 minutos, lo que se cumple con una probabilidad de ocurrencia de $p = 0.70$. El cumplimiento de la meta de lectura diaria es independiente de cada niño. Si la variable aleatoria es **número de niños que cumplen con la meta de lectura diaria**, determine

- a. El tipo de variable aleatoria y los parámetros de este problema.
- b. ¿Cuál es la probabilidad de que 12 niños cumplan la meta?
- c. ¿Cuál es la probabilidad de que máximo, 12 niños cumplan la meta? =
- d. ¿Cuál es la probabilidad de que entre 10 y 16 niños cumplan la meta?

3.4. En el aula existen 40 libros, de los cuales 18 son libros de cuentos y el resto son de asignaturas. La docente selecciona al azar y sin reemplazo 10 libros para una actividad. Si la variable aleatoria es **el número de libros de cuentos seleccionados en la muestra**, determine:

- a. El tipo de variable aleatoria y los parámetros de este problema.
- b. ¿Cuál es la probabilidad de que 3 libros sean de cuentos?
- c. ¿Cuál es la probabilidad de que 3 o más libros sean de cuentos? =
- d. ¿Cuál es la probabilidad de que más de 2 libros sean de cuentos?

3.5. El **número de faltas ortográficas por página** que comete un niño al tomar un dictado es una variable aleatoria de Poisson, puesto que las faltas ortográficas se presentan de manera independiente. El docente ha registrado que históricamente el número promedio de faltas ortográficas por página es 2.4. Si el docente realiza un dictado, determine:

- a. La probabilidad de que el niño no cometa faltas ortográficas en una página, es decir, $\Pr(X = 0)$.

- b. La probabilidad de que, a lo sumo, el niño cometa 2 faltas ortográficas en una página, es decir, $\Pr(X \leq 2)$ (a lo sumo dos faltas).
- c. Si el docente realiza un dictado de dos páginas, ¿cual seria el número promedio de faltas ortográficas en dos páginas?
- d. La probabilidad de que el niño cometa 2 faltas ortográficas en dos páginas.

3.6. Con los datos del ejercicio 3.4, determine: a. el número promedio de cuentos seleccionados, b. la varianza y c. la desviación estándar.

3.7. Con los datos del ejercicio 3.3, determine: a. el número promedio de niños que cumplen con la meta de lectura, b. la varianza y c. la desviación estándar.

3.8. Presente ejemplos de posibles variables aleatorias relacionadas con la educación de niños, justificando su aplicación en los siguientes casos:

- a. Binomial
- b. Hipergeométrica
- c. Poisson
- d. Normal

3.9. Presente ejemplos de posibles variables aleatorias relacionadas con la actividad física o deportes, justificando su aplicación en los siguientes casos:

- a. Binomial
- b. Hipergeométrica
- c. Poisson
- d. Normal

3.10. Con los datos del ejercicio 3.5, determine el número promedio de faltas ortográficas. Además, halle la varianza y la desviación estándar.

3.11. El tiempo de atención a clase (en minutos) en niños de 4 años sigue una distribución normal con una media $\mu = 12$ minutos y una desviación estándar $\sigma = 3$ minutos. Al seleccionar un niño al azar,

- ¿Cuál es la probabilidad de que un niño tenga un tiempo de atención menor a su promedio?
- ¿Cuál es la probabilidad de que un niño tenga una atención menor a 10 minutos?
- ¿Cuál es la probabilidad de que un niño tenga una atención entre 10 y 14 minutos?
- ¿Qué porcentaje de niños supera los 15 minutos?

3.12. La estatura de los niños de 4 años se distribuye normalmente con una media de 102 cm y una desviación estándar de 5 cm. Al seleccionar un niño al azar,

- ¿Qué proporción de niños mide más de su media?
- ¿Qué proporción de niños mide más de 108 cm?
- ¿Cuál es la estatura mínima del 25 % más bajo?
- Identifica el rango (mínimo y máximo) donde está el 95 % de los niños

3.13. La probabilidad de que atleta élite complete exitosamente un intento técnico es 0.8. Si un atleta realiza 6 intentos, determine:

- Probabilidad de que exactamente 5 sean exitosos
- Probabilidad de que al menos 4 sean exitosos

3.14. Con los datos del ejercicio anterior, determine;

- Probabilidad de que más de un intento sea exitoso
- Número promedio de intentos exitosos

3.15. Un equipo de 18 atletas tiene 10 en un nivel óptimo. Si se toma una muestra de 5 sin reemplazo, determine:

- a. Probabilidad de que exactamente 3 estén en nivel óptimo
- b. Probabilidad de que al menos 4 estén en nivel óptimo

3.16. Un grupo de entrenamiento de 20 deportistas tiene 7 de resistencia. Se seleccionan 6 sin reemplazo para un beneficio económico. Determine:

- a. Probabilidad de que exactamente 2 sean de resistencia
- b. Probabilidad de que ninguno sea de resistencia

3.17. Un boxeador profesional conecta en promedio 3 golpes significativos por minuto en combate.

- a. ¿Cuál es la probabilidad de que conecte exactamente 5 golpes significativos en un minuto?
- b. ¿Cuál es la probabilidad de que conecte menos de 2 golpes significativos en un minuto?

3.18. Continuación: Un boxeador profesional conecta en promedio 3 golpes significativos por minuto en combate.

- a. ¿Cuál es la probabilidad de que conecte más de 2 golpes significativos en 30 segundos?
- b. ¿Cuál es la probabilidad de que conecte entre 2 y 4 golpes significativos en 2 minutos

3.19. El tiempo (en segundos) impuesto por un atleta élite en 100 m sigue una distribución normal con una media de 10.2 segundos y una desviación estándar de 0.15 segundos. Al seleccionar un atleta al azar, determine:

- a. Probabilidad de que registre menos de 10 segundos
- b. Probabilidad de que su registro oscile entre 10 y 10.4 segundos.

3.20. El VO_2 máximo (ml/kg/min) de atletas elite sigue una distribución normal con media 70 y desviación estándar 5. Al seleccionar al azar un atleta elite determine:

- a. Probabilidad de que un atleta tenga más de 75
- b. Probabilidad de que esté entre 65 y 80



CAPÍTULO 4

Estimación de parámetros

4.1 Introducción

Los temas a estudiar en este capítulo son parte de la estadística inferencial aplicada a una sola población y muestra. Los temas seleccionados son las distribuciones muestrales, los intervalos de confianza, el cálculo del tamaño de la muestra y las pruebas de hipótesis. Estos temas son fundamentales para generar investigación en el contexto educativo y de la actividad física. El propósito de este capítulo es presentar estos temas de manera práctica y aplicada, sin la rigurosidad matemática propia de carreras de ingeniería, con ejemplos que resulten cercanos al lector.

Algunas aplicaciones de la estadística inferencial para el ámbito educativo y de actividad física están relacionadas con el estudio de comportamientos, desempeño y rendimiento (académico, físico, entre otros) entre grupos de elementos de cualquier población (estudiantes, deportistas, entre otros), llegando a determinar si los resultados son o no significativos de acuerdo a niveles de significación (probabilidad) establecidos para un estudio. Luego de lo cual es posible tomar decisiones con mayor fundamento que un análisis descriptivo por sí solo.

4.2 Conceptos

La **Estadística inferencial** estudia los métodos para obtener conclusiones sobre una población en base a la información que proporciona la muestra.

Parámetro es una medida descriptiva hallada en la población y se denota por θ . **Estadístico o Estimador** es una medida descriptiva hallada en la muestra y se denota por $\hat{\theta}$.

Algunos parámetros son la media μ , la proporción p , la varianza σ^2 , y los respectivos estadísticos son la media muestral $\bar{x}$, la proporción muestral $\hat{p}$, varianza muestral S^2 .

Existen dos tipos de estimadores: el estimador puntual y el estimador por intervalo.

El **estimador puntual** toma un único valor para estimar el parámetro. Ejemplos de estimadores puntuales son: la media $\bar{x}$, la proporción $\hat{p}$ y la S^2 .

El **estimador por intervalo** es un rango de valores dentro del cual oscila el parámetro.

La diferencia entre parámetro y el estadístico se denomina **Error de estimación**

y se expresa:

$$e = \hat{\theta} - \theta$$

Al tomar el valor absoluto del error de estimación, se obtiene el error absoluto

$$E = |\hat{\theta} - \theta|$$

Y cuando el interés es hallar el error en términos de proporción con respecto al valor del parámetro, se realiza el siguiente cociente, denominado error relativo.

$$E_r = \frac{|\hat{\theta} - \theta|}{\theta}$$

Una **muestra aleatoria** es un conjunto de n variable aleatorias $X_1, X_2, \dots, X_n$, independientes y con la misma distribución de probabilidad.

Estas variables aleatorias se pueden organizar en la siguiente matriz:

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{k1} & x_{k2} & x_{k3} & \cdots & x_{kn} \end{bmatrix}$$

En cada columna se disponen los valores de cada variable aleatoria. Donde n representa el número de columnas o variables y k representa el número de filas. Por ejemplo, la primera columna tiene las observaciones de la variable X_1 , y de esta se puede hallar un representante dependiente del parámetro a estudiar. Por ejemplo, si el parámetro a estudiar es la media, entonces su estimación puntual es:

$$\bar{x}_1 = \frac{\sum_{i=1}^k x_{i1} + x_{21} + x_{31} + \dots + x_{k1}}{k}$$

Esto se realiza con cada columna de la matriz de variables aleatorias y será de utilidad en la distribución muestral de la media.

4.3 Distribución es muestrales

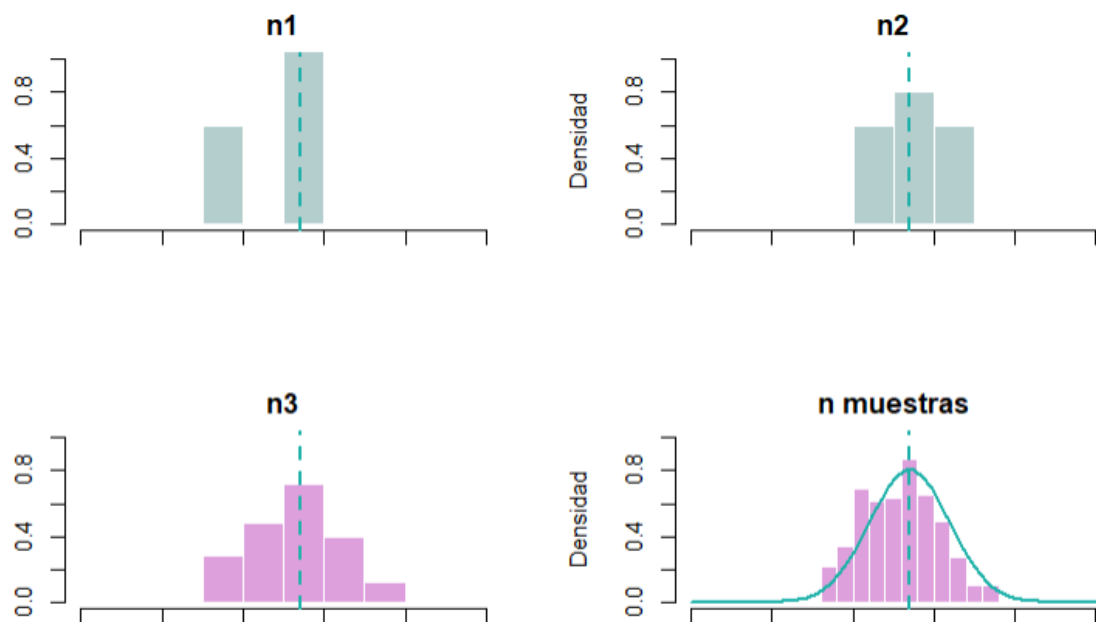
La **distribución muestral de un estimador** es una función de probabilidad construida con los estimadores de cada variable aleatoria de la muestra aleatoria.

Dada una variable aleatoria X en la población de tamaño N , se realiza la distribución muestral de un parámetro, tomando una muestra aleatoria de n variables aleatorias de k elementos. Cada variable aleatoria está representada por un estimador. En teoría, se tienen NCk posibles muestras y estimadores. Mientras más grande sea la muestra aleatoria, el estimador converge a la media poblacional.

En la figura 4.1 se representan 4 distribuciones muestrales, con tamaños que van incrementándose hasta llegar al máximo tamaño de muestra, con el cual se logra la normalidad de la distribución y la estimación tiende a estabilizarse su estimación.

Figura 4.1

Distribución muestral de un estadístico



Este proceso garantiza dos resultados importantes:

- La esperanza matemática del estimador es igual al valor del parámetro.

$$E(\hat{\theta}) = \theta$$

- La varianza de estimación del parámetro es:

$$Var(\hat{\theta}) = \sigma_{\hat{\theta}}^2 = \frac{Var(\theta)}{n} = \frac{\sigma_{\theta}^2}{n}$$

La raíz cuadrada de la varianza se denomina **error estándar de estimación del estimador** y se expresa:

$$S_e(\hat{\theta}) = \sqrt{\frac{\sigma_{\theta}^2}{n}}$$

4.3.1 Distribución muestral de las medias

La distribución de probabilidad de $\bar{X}$ se halla a partir de la variable continua X , para lo cual se toma una muestra aleatoria de tamaño n y se halla la media de cada variable aleatoria con su respectiva función de probabilidad. Donde la media de todas las medias es la media poblacional y la varianza de la distribución muestral de medias se reparte en las n muestras, como se observa a continuación.

$$E(\bar{x}) = \mu$$

$$Var(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n}$$

El error estándar de estimación de la media es:

$$S_e(\bar{x}) = \frac{\sigma_x}{\sqrt{n}}$$

Donde la varianza de la distribución muestral de la media es:

$$\sigma_x^2 = \frac{\sum_{i=1}^m (x_i - \mu)^2 n_i}{n}$$

4.3.2 Teorema del límite central

Sea $X_1, X_2, \dots, X_n$, una muestra aleatoria de tamaño n , tomada de una población (finita o infinita) con media muestral $\bar{x}$, entonces la forma límite de la distribución de Z es normal estándar, si $n \rightarrow \infty$.

$$Z = \frac{\bar{x} - \mu_x}{\frac{\sigma_x}{\sqrt{n}}}$$

Es decir, a mayor tamaño de muestra, la distribución de probabilidad del estimador tiende a ser una normal estándar, sujeta a condiciones.

4.3.3 Distribución muestral de la proporción

La distribución de probabilidad de $\hat{p}$ se halla a partir de la variable aleatoria binomial X : número de éxitos. Sin embargo, la proporción también se puede hallar como un promedio, cuando 1 se asocia al éxito y 0 al fracaso. Entonces, se toma una muestra aleatoria de tamaño n y se halla la distribución de probabilidad de las proporciones. Luego, la esperanza matemática de las proporciones muestrales es la proporción poblacional y la varianza de la distribución muestral de proporciones se reparte en las n muestras, como se observa a continuación.

$$E(\hat{p}) = p$$

$$Var(\hat{p}) = \sigma_{\hat{p}}^2 = \frac{\sigma_p^2}{n}$$

El error estándar de estimación de la proporción es:

$$S_e(\hat{p}) = \sqrt{\frac{\sigma_p^2}{n}}$$

Donde la varianza de la distribución muestral de la proporción es:

$$\sigma_p^2 = pq$$

Luego, por el teorema del limite central aplicado a este caso se tiene:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{\sigma_p^2}{n}}}$$

Ejemplo 4-1. A partir de cada variable cuantitativa, describa el parámetro: media

- a. Variable: Peso corporal (kg).

Tipo: Continua. Parámetro: Peso corporal promedio de niños (kg)

- b. Variable: Tiempo de sueño de un deportista (horas) Tipo: Continua.

Parámetro: Tiempo promedio de sueño de un deportista (horas)

- c. Variable: Perímetro abdominal en hombres (cm) Tipo: Continua.

Parámetro: Perímetro abdominal promedio en hombres (cm)

Ejemplo 4-2. A partir de cada variable cualitativa, elegir un atributo para determinar la variable binomial y el el parámetro: proporción de éxitos.

- a. Variable: Tipo de alimentación del niño (Lactancia materna, fórmula, mixta).

Tipo: Discreta (binomial, por elegir un atributo de interés), por lo tanto:

X : Número de niños alimentados solamente con fórmula.

Parámetro: Proporción de niños alimentados solamente con fórmula.

- b. Variable: Estado emocional predominante (Feliz, tranquilo, irritable, ansioso)

Tipo: Discreta (binomial, por elegir un atributo de interés), por lo tanto:

X : Número de personas con estado emocional predominante: Ansioso.

Parámetro: Proporción de personas con estado emocional predominante: Ansioso.

- c. Variable: Categoría del deportista (Juvenil, Senior, Máster, Súper Máster)

Tipo: Discreta (binomial, por elegir un atributo de interés), por lo tanto:

X : Número de deportistas de la categoría Juvenil.

Parámetro: Proporción de deportistas de la categoría Juvenil.

En la Tabla 4.1 se presenta el resumen de resultados obtenidos.

Tabla 4.1

Cuadro comparativo de distribuciones muestrales

Característica	Media	Proporción
Tipo de variable aleatoria	Continua	Discreta
Tipo de distribución	Normal	Normal, con $np > 5$
Parámetro	μ	p
Error de estimación	$\frac{\sigma}{\sqrt{n}}$	$\sqrt{\frac{pq}{n}}$
Estadístico	$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$	$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$

4.4 Estimación por intervalo

Como se mencionó anteriormente, un estimador de parámetros por intervalos proporciona un rango de valores en el cual oscila el parámetro. Dicho intervalo se construye con un nivel de confianza que determina puntos críticos en una distribución de probabilidad. Los intervalos de confianza pueden ser de tipo bilateral, unilateral superior y unilateral inferior. A continuación, solo se desarrolla el caso bilateral. Los elementos del intervalo de confianza son:

- El Nivel de confianza representa la probabilidad de que el parámetro pertenezca al intervalo de confianza. Se denota por $1 - \alpha$. La confianza y la significancia son complementarias. La confianza suele tomar valores entre 0.90 y 0.99.
- El Nivel de significancia representa la probabilidad de que el parámetro no pertenezca al intervalo de confianza. Se denota por α . La significancia toma valores entre 0.01 y 0.10.
- El Nivel de confianza y el nivel de significancia son complementarios entre sí, por lo tanto, su suma es 1, ya que cubren toda el área bajo la curva de la distribución de probabilidad: $1 - \alpha + \alpha = 1$
- Puntos críticos que delimitan las zonas de confianza y significancia hallados en la distribución de probabilidad correspondiente.
- Tamaño de la muestra n que influye en el tamaño del intervalo.

4.5 Intervalo de confianza bilateral para la media poblacional μ

Si X es una variable aleatoria con distribución de media μ y varianza poblacional σ^2 y se toma una muestra de tamaño n , entonces le corresponde el estadístico de la distribución muestral de la media:

$$Z = \frac{\bar{x} - \mu_x}{\frac{\sigma_x}{\sqrt{n}}}$$

Pero si la variable aleatoria es aproximadamente normal y no se conoce la varianza poblacional, se debe aplicar la variable aleatoria estandarizada T y la distribución t-Student con $n-1$ grados de libertad:

$$T = \frac{\bar{x} - \mu_x}{\frac{s_x}{\sqrt{n}}}$$

Por lo tanto, un intervalo de confianza bilateral para la media con una confianza $1 - \alpha$ presenta dos casos.

- Caso 1. Intervalo de confianza bilateral para la media con varianza poblacional conocida.
- Caso 2. Intervalo de confianza bilateral para la media con varianza poblacional desconocida.

Ejemplo 4-3. Variables aleatorias continuas del área del desarrollo infantil en las que se podrían realizar estimaciones por intervalo para el promedio.

- Puntaje en una Escala de Desarrollo (Escala Bayley, entre otras.)
- Tiempo de Atención Sostenida (segundos o minutos)

Ejemplo 4-4. Variables aleatorias continuas del área de la actividad física en las que se podrían realizar estimaciones por intervalo para el promedio.

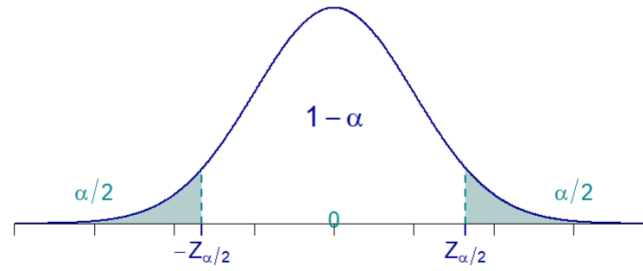
- Consumo Máximo de Oxígeno (VO_2 máx) en ml/kg/min.
- Umbral de Lactato en sangre en mmol/L.

4.5.1 Intervalo de confianza bilateral para la media poblacional μ , con varianza poblacional σ^2 conocida.

Sea X una variable aleatoria con distribución normal y varianza poblacional σ^2 conocida. Se toma una muestra aleatoria de tamaño n con media $\bar{x}$, entonces un intervalo de confianza bilateral para μ con una confianza $1 - \alpha$ es:

$$\bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

En la figura 4.2 se representan los puntos críticos en la distribución normal estándar Z correspondientes a los límites del intervalo de confianza bilateral para μ , que delimitan las zonas de significancia α distribuidas en las colas y de confianza $1 - \alpha$ en el centro.

Figura 4.2*Puntos críticos en la distribución normal estándar*

Ejemplo 4-5. Algunos maratonistas organizan el tiempo de sueño en dos partes, en la noche un promedio de 8 horas y en el día un promedio de 2 horas. Se conoce que la distribución del tiempo de sueño en el día sigue una distribución normal con una desviación estándar de 15 minutos. Se toma una muestra de 24 maratonistas y se determina que duermen en promedio 1.75 horas. Use un nivel de confianza del 95 % para hallar el intervalo de confianza bilateral para el tiempo promedio de sueño en el día.

Considerando que la variable tiempo de este ejemplo esta definida en horas, el tiempo de 15 minutos se transforma a horas, por lo tanto,

$$\sigma = 0.25$$

El nivel de confianza en porcentaje es 95 %, pero en probabilidad es: $0.95 = 1 - \alpha$. Luego, el nivel de significancia es:

$$\alpha = 0.05$$

Los puntos críticos correspondientes se hallan mediante la tabla de la distribución normal estándar $F(z) = p(Z < z)$ del Apéndice B. Es suficiente hallar el punto crítico en la cola izquierda y debido a la simetría de la distribución normal se halla el otro punto crítico:

$$Z_{\alpha/2} = Z_{0.025} = \pm 1.96$$

Se reemplazan estos resultados en la fórmula del intervalo de confianza:

$$\begin{aligned} \bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} &\leq \mu \leq \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \\ 1.75 - 1.96 \frac{0.25}{\sqrt{18}} &\leq \mu \leq 1.75 + 1.96 \frac{0.25}{\sqrt{18}} \end{aligned}$$

$$[1.6345, 1.8655]$$

Se concluye que, con una confianza del 95 % el tiempo promedio de sueño durante el día para maratonistas de élite oscila entre 1.6345 y 1.8655 horas.

4.5.2 Intervalo de confianza bilateral para la media poblacional μ , con varianza poblacional σ^2 desconocida.

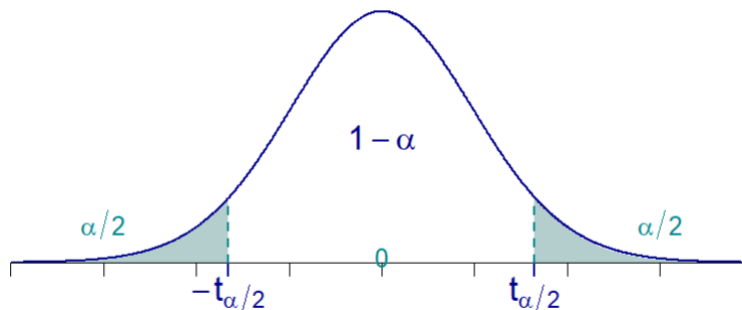
Sea X una variable aleatoria con distribución aproximadamente normal con varianza poblacional σ^2 desconocida. En este caso se toma la distribución t-Student con $n-1$ grados de libertad. Luego, el intervalo de confianza bilateral para la media es:

$$\bar{x} - t_{\frac{\alpha}{2}, g.l} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2}, g.l} \frac{s}{\sqrt{n}}$$

La figura 4.3 es similar al caso de la figura 4.2, donde se observan los puntos críticos correspondientes a los límites del intervalo de confianza bilateral para la media.

Figura 4.3

Puntos críticos en la distribución t-Student



Ejemplo 4-6. Un psicólogo realiza una prueba estandarizada por computador para estimar el tiempo de reacción a un estímulo visual en niños de 5 años. Esta variable sigue una distribución normal. Se toma una muestra aleatoria de 18 niños y los resultados de la prueba en mili segundos son: 450, 520, 480, 610, 390, 550, 440, 590, 510, 470, 430, 570, 500, 410, 530, 460, 495, 585. Use un nivel de confianza del 95 % para hallar un intervalo de confianza bilateral para el tiempo promedio de reacción.

Primero, se deben hallar la media y la varianza muestrales.

$$\bar{x} = 499.44$$

$$s = 64.33$$

El nivel de confianza en porcentaje es 95 %, pero en probabilidad es:

$$0.95 = 1 - \alpha$$

Luego, el nivel de significancia es:

$$\alpha = 0.05$$

Los grados de libertad son $gl = n - 1 = 18 - 1 = 17$

Los puntos críticos son:

$$t_{\alpha/2} = t_{0.025,17} = \pm 2.11$$

Se tienen dos puntos críticos por ser un intervalo de confianza bilateral.

Se reemplazan estos resultados en la fórmula del intervalo de confianza:

$$\begin{aligned} \bar{x} - t_{\alpha/2, gl} \frac{s}{\sqrt{n}} &\leq \mu \leq \bar{x} + t_{\alpha/2, gl} \frac{s}{\sqrt{n}} \\ 499.44 - 2.11 \frac{64.33}{\sqrt{18}} &\leq \mu \leq 499.44 + 2.11 \frac{64.33}{\sqrt{18}} \\ &[467.45, 531.43] \end{aligned}$$

Con una confianza del 95 %, el tiempo de reacción promedio ante el estímulo en niños de 5 años oscila entre 467.45 y 531.43 mili segundos.

4.6 Intervalo de confianza bilateral para la proporción poblacional p

Dada la variable aleatoria binomial X : número de éxitos, se obtiene una muestra aleatoria de tamaño n , donde $\hat{p} = x/n$ es el estimador puntual de la proporción poblacional p . Se conoce que la distribución de muestreo de $\hat{p}$ es aproximadamente normal (siempre que $np > 5$) con media np y varianza $\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}$ entonces, el intervalo de confianza bilateral al $100(1 - \alpha) \%$ para estimar la proporción poblacional p es:

$$\hat{p} - Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

El gráfico que representa los puntos críticos en la distribución normal estándar es el mismo que se presentó en la sección 4.5.1.

Ejemplo 4-7. Variables aleatorias binomiales del desarrollo infantil en las que se podrían realizar estimaciones por intervalo para la proporción

- Número de menores que viven en contextos familiares no funcionales.
- Número de niños que alcanzaron el hito de desarrollo a los 12 meses (caminar).
- Número de niños con bajo peso para su edad.

Ejemplo 4-8. Variables aleatorias binomiales de la actividad física en las que se podrían realizar estimaciones por intervalo para la proporción

- Número de deportistas lesionados durante la etapa de entrenamiento.
- Número de deportistas que usan suplementos deportivos sin receta de un profesional.
- Número de deportistas con bajo umbral de percepción del dolor.

Ejemplo 4-9. Se realiza un experimento controlado con niños de 3 años, para determinar la estrategia de negación ante el ofrecimiento de un objeto no deseado. Se presentan dos posibles reacciones o estrategias de negación: mover la cabeza o empujar el objeto. Se ha tomado una muestra de 32 niños de 3 años, y se determina que 7 niños empujan el objeto y el resto mueve la cabeza. Halle el intervalo de confianza bilateral para estimar la proporción p de niños que empujan el objeto. Use una significancia $\alpha = 0.01$.

Proporción muestral de niños que empujan el objeto:

$$\hat{p} = \frac{7}{32}$$

Proporción muestral de niños que no empujan el objeto:

$$\hat{q} = 1 - \hat{p} = 1 - \frac{7}{32} = \frac{25}{32}$$

Para $\alpha = 0.01$, se tienen dos puntos críticos $Z_{\alpha/2} = Z_{0.005} = \pm 2.575$. Se reemplazan estos resultados en la fórmula del intervalo de confianza:

$$IC = \hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

$$\frac{7}{32} \pm 2.575 \sqrt{\frac{\frac{7}{32} \frac{25}{32}}{32}}$$

$$[0.0306, 0.4069]$$

Con una confianza del 99 % se estima que la proporción de niños de 3 años que empujan el objeto no deseado oscila entre 0.0306 y 0.4069.

4.7 Tamaño de la muestra para poblaciones infinitas

Es importante resaltar que una población se considera infinita para poblaciones de más de 6000 individuos o elementos. Y que el cálculo del tamaño de la muestra por sí solo no es suficiente, se debe estudiar toda una teoría sobre el muestreo para desarrollar adecuadamente un estudio.

La base construida en la estimación por intervalo permite deducir el tamaño de la muestra. Dependiendo del parámetro a estimar, el tamaño de la muestra se deduce a partir del intervalo de confianza correspondiente. En este libro solo se han considerado dos estimaciones por intervalo; sin embargo, existen muchas posibilidades de estimación por intervalo, ya sea para un grupo, dos o más grupos y otro tipo de estimación. En lo que corresponde a las dos estimaciones de parámetros estudiadas en este libro, se tiene que:

- El tamaño de la muestra para estimar la media poblacional μ se despeja n del intervalo de confianza bilateral para estimar la media μ .
- El tamaño de la muestra para estimar la proporción poblacional p se determina n del intervalo de confianza bilateral para estimar la proporción p .

Las fórmulas para calcular el tamaño de la muestra que se presentan a continuación son válidas para poblaciones infinitas, y en general son aplicables para los dos casos mencionados. Su fórmula es:

$$n_{\infty} = \left(\frac{z_{\alpha/2} \cdot \sigma}{e} \right)^2$$

No obstante, mediante la aplicación de la corrección por población finita, también puede utilizarse para el caso de poblaciones finitas. Esta corrección por población finita se aplica cuando el tamaño de la muestra n_{∞} es mayor o

igual que el 5 % del tamaño de la población total N . La fórmula de corrección por población finita es:

$$n = \frac{n_{\infty}}{1 + \frac{n_{\infty}}{N}}$$

Para evitar el proceso de corrección por población finita, se usa esta fórmula general utilizada tanto para poblaciones finitas como infinitas:

$$n = \frac{Z_{\alpha/2}^2 \sigma^2 N}{e^2 (N - 1) + Z_{\alpha/2}^2 \sigma^2}$$

Observación. Si un estudio que tiene como instrumento una encuesta y tiene más de un parámetro a estimar, se debe elegir el parámetro de mayor importancia para el cálculo del tamaño de la muestra.

4.7.1 Consideraciones generales sobre el tamaño de la muestra

- A mayor varianza, mayor tamaño de muestra. Es decir, entre la varianza y el tamaño de muestra existe una relación de proporcionalidad directa.
- A menor error, el tamaño de muestra aumenta. Es decir, entre el error y el tamaño de muestra existe una relación de proporcionalidad inversa.
- Resulta conveniente usar una muestra cuando el tamaño de la población es grande o infinita.

4.7.2 Tamaño de la muestra para estimar la media

Si X es variable continua, el parámetro de interés es la media poblacional μ entonces, el tamaño de la muestra para estimar la media es:

$$n = \frac{Z_{\alpha/2}^2 \sigma^2 N}{e^2 (N - 1) + Z_{\alpha/2}^2 \sigma^2}$$

Donde:

N : Tamaño de la población

n : Tamaño de la muestra

$Z_{\alpha/2}$: Punto crítico en la distribución normal

e : $|\bar{x} - \mu|$

σ^2 : Varianza poblacional

Consideraciones importantes

- El error de estimación debe tener las mismas unidades de la variable de estudio, pero no debe superar el 5 % en términos de porcentaje. Por ejemplo, para la variable peso (kg), la media poblacional de un proceso similar o teórico es $\mu = 120kg$, para tomar un error de 5 %, se debe calcular: $0.05 * 120kg = 6kg = e$.
- La varianza poblacional toma el valor teórico si este existe; caso contrario, se aproxima con la varianza de estudios anteriores o de una prueba piloto (estudio en un grupo reducido de elementos o individuos); en todo caso, para dos casos se usa la fórmula de la varianza muestral:

$$s_x^2 = \frac{\sum_{i=1}^m (x_i - \bar{x})^2 \cdot n_i}{n - 1}$$

Ejemplo 4-10. Calcular el tamaño de la muestra para estimar el peso promedio de bebés al nacer y otros factores asociados al peso del recién nacido. Dicha población está formada por 1100 bebés de una data de un hospital. Usar un nivel de confianza del 95 %. De acuerdo a datos oficiales el peso promedio debe ser 3250 gramos, con una desviación estándar de 300 gramos. Tome un error equivalente al 5 %.

Datos:

Para un nivel de confianza $1 - \alpha = 0.95$, se tiene una significancia de $\alpha = 0.05$; luego, el punto crítico es $Z_{\alpha/2} = Z_{0.025} = 1.96$.

$$\sigma = 300 \text{ gramos}$$

$$e = 3250 \text{ gramos} * 0.02 = 65 \text{ gramos}$$

$$n = \frac{Z_{\alpha/2}^2 \sigma^2 N}{e^2 (N - 1) + Z_{\alpha/2}^2 \sigma^2}$$

$$n = \frac{1.96^2 * 300^2 * 1100}{65^2 (1100 - 1) + 1.96^2 * 300^2}$$

$$n = 76.23 = 77$$

Se redondea al entero siguiente para no cometer un error mayor al máximo admisible. Se debe tomar la información primaria de 77 bebés.

4.7.3 Tamaño de la muestra para estimar la proporción

Si X es variable binomial, el parámetro de interés es la proporción poblacional p entonces, el tamaño de la muestra para estimar la proporción es:

$$n = \frac{Z_{\alpha/2}^2 \sigma^2 N}{e^2 (N - 1) + Z_{\alpha/2}^2 \sigma^2}$$

Donde:

N : Tamaño de la población

n : Tamaño de la muestra

$Z_{\alpha/2}$: Punto crítico en la distribución normal

e : $|\hat{p} - p|$

σ^2 : $p * q$

En este caso particular n se puede expresar como sigue:

$$n = \frac{Z_{\alpha/2}^2 pqN}{e^2 (N - 1) + Z_{\alpha/2}^2 pq}$$

Consideraciones importantes

- El error de estimación se lo toma directamente con proporciones entre 0.01 y 0.05. Por ejemplo, $e = 0.01$.
- La varianza poblacional se toma del valor teórico si este se conoce; caso contrario, se aproxima con la varianza de estudios anteriores, o realizar una prueba piloto (estudio en un grupo reducido de elementos o individuos); además se puede aplicar la regla empírica que considera la misma proporción entre éxitos y fracasos, es decir: $p = q = 0.5$, luego $s^2 = p * q = 0.5 * 0.5$, y con esta combinación se obtiene el máximo tamaño de muestra.

Ejemplo 4-11. Calcular el tamaño de la muestra para estimar la proporción de deportistas que no desayunan antes de una competencia que inicia a las 07:00 am. La población está formada por 200 deportistas. Tomar un nivel de confianza del 95 %. Un error del 2 %. De un estudio anterior se conoce que la proporción de deportistas que no desayunan éxitos es 0.01. Utilice esta información para aproximar la varianza.

Para un nivel de confianza $1 - \alpha = 0.95$, se tiene una significancia de $\alpha = 0.05$; luego, el punto crítico es $Z_{\alpha/2} = Z_{0.025} = 1.96$.

$$p = 0.01$$

$$q = 1 - 0.01 = 0.99$$

$$\sigma^2 = p * q = 0.01 * 0.99$$

Luego:

$$n = \frac{Z_{\alpha/2}^2 pq N}{e^2 (N - 1) + Z_{\alpha/2}^2 pq}$$

$$n = \frac{1.96^2 * 0.01 * 0.99 * 200}{0.02^2 (200 - 1) + 1.96^2 * 0.01 * 0.99}$$

$$n = 64.66 = 65$$

Se debe tomar la información primaria a 65 deportistas.

Ejemplo 4-12. Calcular nuevamente el tamaño de la muestra anterior, pero ahora considere que no se dispone de información alguna sobre la varianza.

Al no disponer de información sobre la varianza, resulta conveniente aplicar la regla empírica, por lo tanto:

$$p = q = 0.5$$

$$\sigma^2 = p * q = 0.5 * 0.5$$

$$n = \frac{1.96^2 * 0.5 * 0.5 * 200}{0.02^2 (200 - 1) + 1.96^2 * 0.5 * 0.5}$$

$$n = 184.69 = 185$$

Se debe tomar la información primaria a 185 deportistas.

Ejemplo 4-13. Discuta los resultados obtenidos en los dos ejercicios anteriores

La varianza y el tamaño de la muestra tienen una relación de proporcionalidad directa, puesto que en el primer caso, la varianza es: $0.01 * 0.99 = 0.0099$ es menor que en el segundo caso la varianza es $0.5 * 0.5 = 0.25$ por lo que les corresponden tamaños de muestra el primero más pequeño que en el segundo caso. Esta comparación tiene sentido si se mantiene el mismo valor en el error en los dos casos, esto es: $e = 0.02$.

Ejemplo 4-14. Calcular nuevamente el tamaño de la muestra anterior, pero ahora considere otra vez la regla empírica, pero tome un error de estimación máximo de 0.05. ¿Cuál es su conclusión?

$$n = \frac{1.96^2 * 0.5 * 0.5 * 200}{0.05^2 (200 - 1) + 1.96^2 * 0.5 * 0.5}$$

$$n = 131.75 = 136$$

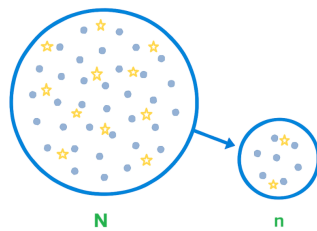
Con respecto al ejercicio anterior, el tamaño de la muestra pasa de 185 a 136 deportistas. Es decir, al incrementar el error, el tamaño de muestra disminuye.

4.8 Introducción a las técnicas de muestreo

El muestreo es esencial en el diseño de la investigación, ya que permite analizar solo una parte representativa de la población. El muestreo se clasifica en técnicas de muestreo probabilístico y no probabilístico. La elección de la técnica responde a la naturaleza del problema, los recursos, el tiempo y el nivel de precisión requerido en los resultados. En todo caso, la muestra debe tener las mismas características que la población, como se representa en la figura 4.4

Figura 4.4

Población y muestra



A continuación, se presentan conceptos básicos.

La **Unidad de Análisis** es la unidad mínima de la cual se toman las observaciones.

La **Unidad de Muestreo** es el conjunto de unidades de análisis.

El **Marco Muestral** es el conjunto de unidades de muestreo.

4.8.1 Muestreo no probabilístico

El muestreo no probabilístico se basa en criterios subjetivos o de accesibilidad, sin que todas las unidades tengan la misma oportunidad de ser incluidas en la muestra. Solo es de utilidad en estudios exploratorios cualitativos, restringiendo la generalización de los resultados hacia la población. Los tipos de muestreo no probabilístico son el muestreo por conveniencia, por juicio, bola de nieve y cuotas.

Ejemplo 4-15. Una docente realiza un estudio exploratorio sobre hábitos de lectura en niños de 4 a 5 años. Para ello, selecciona solamente a los estudiantes de su propio grupo de clase, ya que tiene acceso directo a ellos y puede observarlos durante las actividades diarias. ¿Qué técnica de muestreo se aplicaría en este caso?

Está aplicando muestreo no probabilístico de conveniencia al seleccionar a los estudiantes de su aula. No existe aleatoriedad al observar a todos sus estudiantes, sin que ello implique generalizar los resultados obtenidos a todos los niños de características similares, incluso de la misma unidad educativa.

Ejemplo 4-16. ¿Cuáles son las unidades de análisis, de muestreo y marco muestral del ejemplo anterior?

La unidad de análisis es un niño, la unidad de muestreo es un niño y el marco muestral es la nómina de niños de esa aula de la docente.

4.8.2 Muestreo probabilístico

El muestreo se denomina probabilístico cuando cada unidad de la población tiene una probabilidad de ser elegida, y con ello se garantiza mayor rigor en el análisis y la posibilidad de aplicar métodos inferenciales con mayor precisión. Los tipos de muestreo probabilístico son muestreo aleatorio simple, estratificado, sistemático y por conglomerados.

4.8.2.1 MUESTREO ALEATORIO SIMPLE

En esta técnica de muestreo todos los elementos de la población tienen la misma probabilidad de ser elegidos. Los elementos de la muestra se eligen de forma aleatoria (al azar). En este caso, la unidad de análisis es igual a la unidad de muestreo.

4.8.2.2 MUESTREO SISTEMÁTICO

En esta técnica de muestreo, los elementos se eligen de 1 en k , eso significa que cada intervalo de tamaño k se elige un elemento. En este caso, la unidad de análisis es igual a la unidad de muestreo.

4.8.2.3 MUESTREO ESTRATIFICADO

En esta técnica de muestreo, la población está formada por L grupos o estratos, los cuales no deben solaparse. Los estratos se agrupan por homogeneidad, pero entre estratos son heterogéneos respecto de la variable a estudiar. La muestra tiene representantes de cada estrato obtenidos de forma aleatoria simple. En este caso, la unidad de análisis es igual a la unidad de muestreo.

4.8.2.4 MUESTREO POR CONGLOMERADOS

En esta técnica de muestreo, la población está formada por M unidades de muestreo o conglomerados, estos grupos no deben solaparse. Con respecto de la variable a estudiar, los conglomerados se agrupan por heterogeneidad, pero entre estratos son homogéneos. Se hace un muestreo aleatorio simple de conglomerados, pero en cada conglomerado seleccionado se realiza un censo.

Ejemplo 4-17. Un investigador desea conocer el nivel de condición física de los estudiantes universitarios. Utiliza un muestreo aleatorio estratificado, seleccionando proporcionalmente estudiantes de tres carreras de una facultad para asegurar una representación equilibrada de toda la población estudiantil, con respecto a la carga de trabajo y por ende el tiempo libre para actividad física. Determine las unidad de análisis, muestreo y marco muestral.

- La unidad de análisis es un estudiante.
- La unidad de muestreo es estudiante.
- El marco muestral es la nómina de estudiantes clasificados por carreras.

La elección de dichas unidades y la técnica de muestreo está sujeta al nivel de acceso de la información y a los individuos a través de diferentes recursos que faciliten el proceso.

4.9 Laboratorio de Estimación

4.9.1 Intervalo de confianza para la media en Jamovi

Este procedimiento se aplica para variables con distribución t-Student con $n - 1$ grados de libertad o distribución normal.

Proceso: Menú Exploración, Descriptivas, ingresar la variable, marcar **Intervalo de confianza para la media**, donde puede modificar el nivel de confianza que por defecto es 95 %, según se observa en la figura 4.5. Marque o desmarque otros resultados según su preferencia. Luego, dar click en la flecha para generar los resultados.

Figura 4.5

Menu Intervalo de confianza

Dispersión de Medias

☐ Error est. de la Media

☒ Intervalo de confianza para la Media 95 %

☐ Shapiro-Wilk

Valores atípicos

☐ Más extremo 5

Ejemplo 4-18. Obtenga el intervalo de confianza bilateral para el peso promedio. Use el archivo pesos. El peso sigue una distribución normal.

Siguiendo el proceso indicado anteriormente, pero ingresando la variable peso, se obtienen los resultados que se observan en la figura 4.6, donde se observan tres descriptivos adicionales al intervalo de confianza.

Figura 4.6

IC. del peso promedio

Descriptivas	
	Peso_lb
N	90
Media	163.2
IC 95% de la media límite inferior	158.9
IC 95% de la media límite superior	167.5
Desviación estándar	20.59

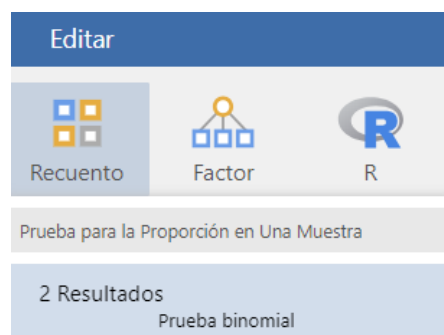
Nota. El CI de la media supone que las medias muestrales siguen una distribución t con $N - 1$ grados de libertad

4.9.2 Intervalo de confianza para la proporción en Jamovi

Este procedimiento se aplica para variables que provienen de muestras grandes $np \geq 5$. Proceso: Menú Recuento, Descriptivas, Prueba binomial (2 resultados), según se observa en la figura 4.7.

Figura 4.7

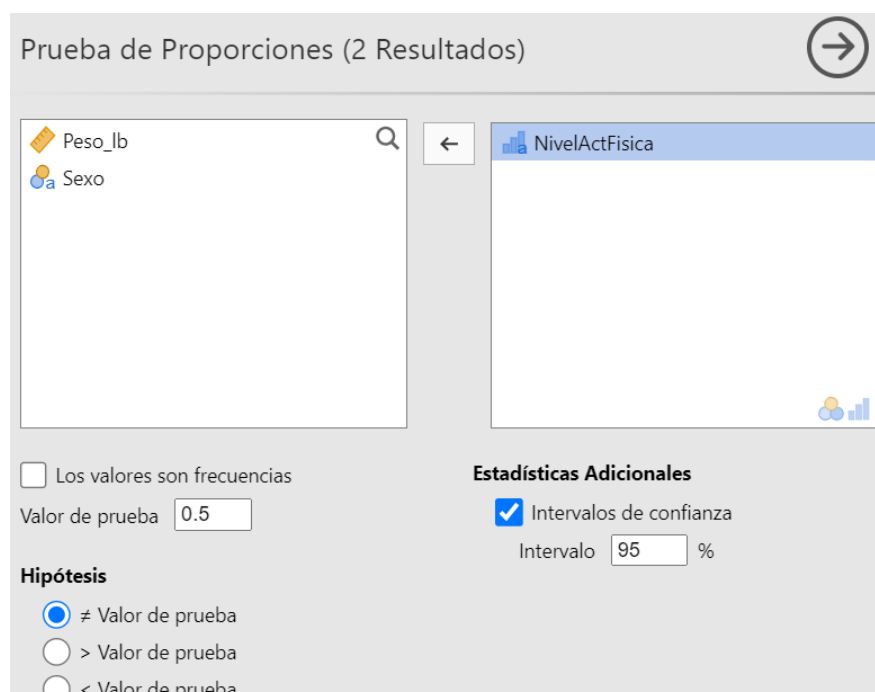
Menu Recuento



Ingresa la variable cualitativa, marcar **Intervalos de confianza**, donde puede modificar el nivel de confianza que por defecto es 95 %, como se indica en la figura 4.8

Figura 4.8

Menu Intervalo de confianza para p



Ejemplo 4-19. Obtenga el intervalo de confianza bilateral para estimar la proporción de personas que realizan actividad física a un nivel Alto. Use el archivo pesos.

Siguiendo el proceso indicado anteriormente, pero ingresando la variable NivelActFisica, se obtienen los resultados de la figura 4.9, donde incluso se puede hallar un intervalo de confianza para cada atributo de nivel de actividad física, así como probar hipótesis que se estudiarán más adelante.

Figura 4.9

IC para la proporción

Prueba Binomial

						Intervalo de Confianza al 95%	
	Nivel	Frecuencia	Total	Proporción	p	Inferior	Superior
NivelActFisica	Alto	10	90	0.1111	1.052e-14	0.05459	0.1949
	Bajo	41	90	0.4556	0.4608	0.35015	0.5640
	Medio	39	90	0.4333	0.2461	0.32918	0.5420

Nota. H_0 es proporción $\neq 0.5$

4.9.3 Muestreo en Jamovi

En esta sección se realiza la generación de una muestra aleatoria simple, para lo cual se requieren los siguientes ítems.

- Tamaño de la población N y el tamaño de la muestra n .
- Marco muestral que es la lista numerada de individuos o elementos de la población de tamaño N . Esta lista puede crearse en Jamovi o importarse desde Excel.
- Tamaño de la muestra n hallada con la técnica de muestreo correspondiente. En este caso, muestreo aleatorio simple.

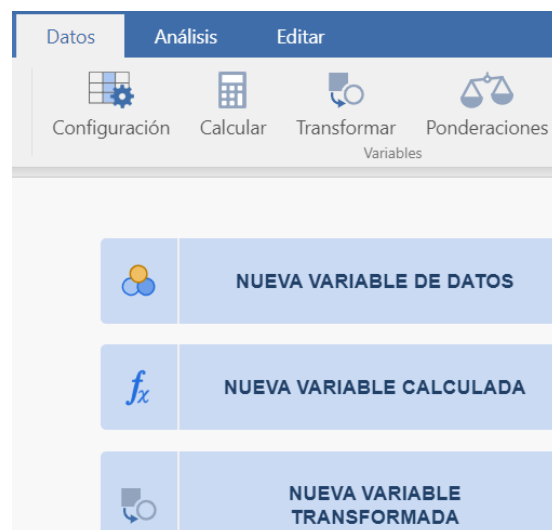
Ejemplo 4-20. Obtenga una muestra aleatoria simple de tamaño $n = 65$ a partir de una población de tamaño $N = 200$ en Jamovi

- a. Importar el marco muestral desde Excel. En este caso solo se dispone la lista de ID desde ID1 al ID200 (MarcoMuestral.xlsx).
- b. En el menú Datos, elegir Calcular, o también se coloca en una columna en blanco, y elegir NUEVA VARIABLE CALCULADA, tal como se muestra

en la figura 4.10. Sin embargo, con la opción del filtro se puede realizar el mismo proceso. Elegir el nombre, por ejemplo Muestra.

Figura 4.10

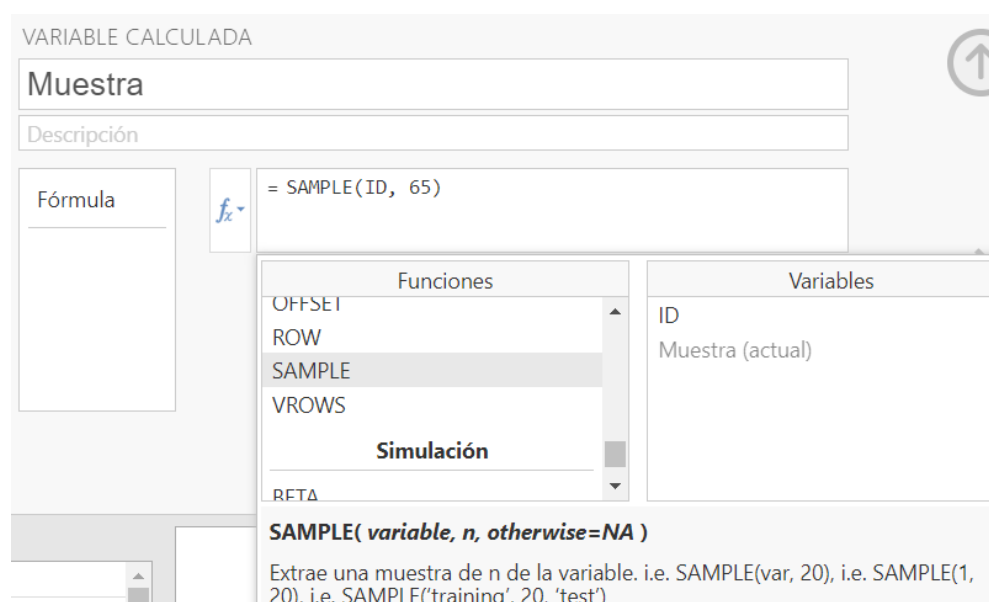
Nueva variable calculada



- c. Ahora se dispone del menú del cálculos y funciones, se elige SAMPLE, para elegir la variable objetivo y el tamaño respectivo: SAMPLE(ID, 65), como se muestra en la figura 4.11.

Figura 4.11

Función a calcular



- d. Luego de lo cual, se genera la muestra, como se observa en la figura 4.12 los espacios en blanco son las filas o individuos no seleccionados.

Figura 4.12

Marco muestral y muestra

	ID	Muestra
1	ID1	ID1
2	ID2	
3	ID3	
4	ID4	
5	ID5	
6	ID6	ID6
7	ID7	
8	ID8	
9	ID9	
10	ID10	
11	ID11	ID11
12	ID12	
13	ID13	
14	ID14	ID14
15	ID15	

- e. Puede filtrar los elementos no seleccionados en Jamovi y descargar el archivo. Aunque también puede descargar el archivo sin filtrar y hacerlo en Excel.

Ejemplo 4-21. Obtenga una muestra aleatoria simple de tamaño $n = 65$ a partir de una población de tamaño $N = 200$ en Excel

- Utilizando el mismo archivo del marco muestral, se puede crear una nueva columna llamada aleatorio para escribir la función `=aleatorio.entre(1;200)`, copiar esa función en 65 filas de la columna aleatorio.
- Copiar la columna aleatorio en una tercera columna llamada muestra, pero realizando un pegado especial solo con valores para dejar fija la muestra.
- Ordenar esta última columna de menor a mayor para verificar si hay repeticiones. En caso de existir repeticiones generar nuevos números aleatorios hasta obtener una muestra de 65 observaciones sin repetición.

4.10 Ejercicio propuestos del capítulo

Ejercicios para Educación inicial

4.1. Hallar el intervalo de confianza bilateral al 95 % para el tiempo promedio de atención en niños de 3 años. Una maestra elige como actividad la lectura y registra el tiempo de atención sostenida en una muestra de 30 niños de 3 años; en esta muestra se obtiene un tiempo promedio de atención de 6.2 minutos, con una desviación estándar de 1.5 minutos.

4.2. A una muestra de 25 niños de 5 años se les pidió que contaran una historia libre y se encontró un promedio de 56 palabras usadas, con una desviación estándar de 7 palabras. Hallar un intervalo de confianza bilateral al 95 % para el número de palabras al contar la historia.

4.3. Con una confianza del 0.95, hallar un intervalo de confianza bilateral para la proporción de niños de 3 años que reconocen colores básicos. En una muestra de 50 niños de 3 años se determina que 36 logran identificar correctamente al menos tres colores básicos (rojo, azul, amarillo).

4.4. Con una confianza del 0.95 hallar el intervalo de confianza bilateral para la proporción de niños que escriben correctamente su nombre. En una muestra de 36 niños de 5 años, 28 escriben correctamente su nombre completo. ¿Se puede afirmar que al menos el 80 % de los niños de 5 años escriben correctamente su nombre completo?

Ejercicios para Actividad Física

4.5. En una muestra de 20 partidos de tenis, se obtuvo una media de 8.1 aces con una desviación estándar de 1.1. Hallar el intervalo de confianza bilateral con una significancia de 0.05

4.6. En una competencia de salto ecuestre, se toma una muestra de 25 jinetes y se obtiene una puntuación promedio de 83.2 puntos, con una desviación estándar de 4.5. Utilice un nivel de significancia de 0.01 para hallar un intervalo de confianza bilateral sobre la puntuación en la competencia.

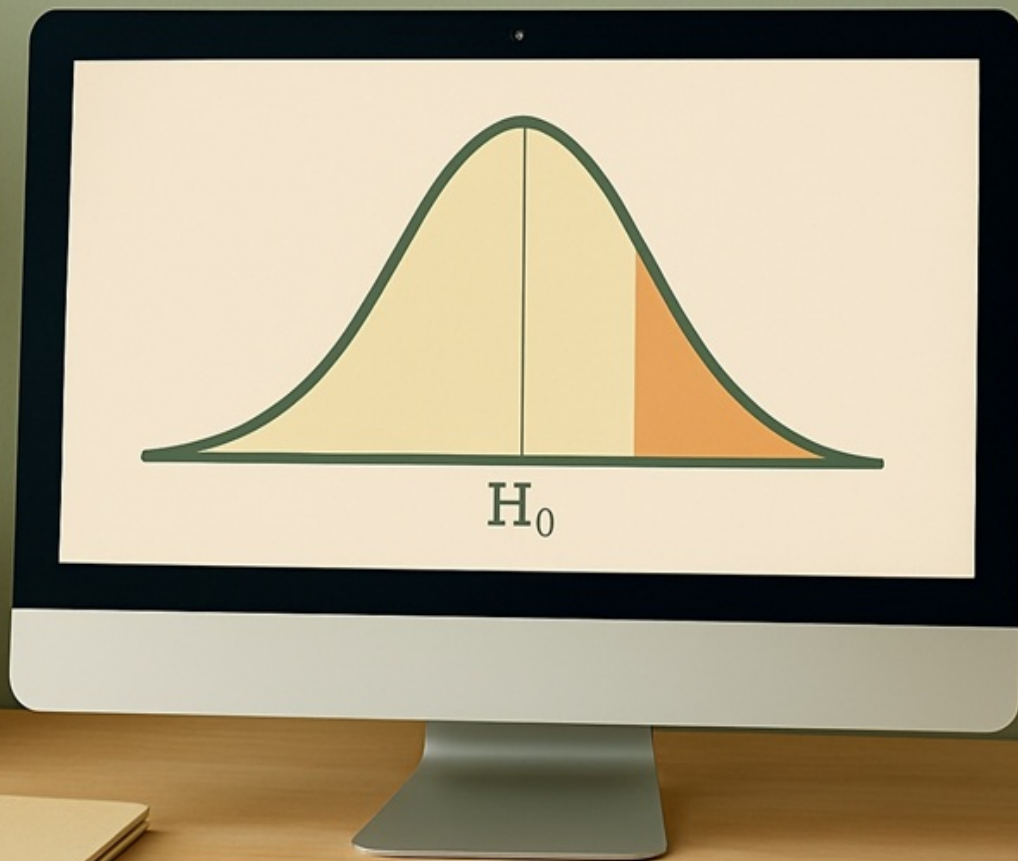
4.7. Hallar el intervalo de confianza bilateral para determinar la proporción de competidores que completaron la maratón en menos de 4 horas. De una

muestra de 200 participantes, 138 completaron la carrera en menos de 4 horas.

4.8. Hallar el intervalo de confianza bilateral al 95 % para estimar la proporción de gimnastas que logran la puntuación perfecta en el salto del potro. De una muestra de 50 gimnastas, 12 lograron una puntuación perfecta en el salto de potro.

4.9. En una prueba de orientación militar, 72 de 100 cadetes completaron el recorrido sin errores. Hallar el intervalo de confianza bilateral para estimar la proporción de cadetes que tienen éxito en la navegación?

4.10. Hallar el intervalo de confianza bilateral al 95 % para estimar el consumo Máximo de Oxígeno (VO_2 máx) en corredores élite durante una carrera de 5 kilómetros. Se toma una muestra de 36 corredores en una prueba de 5 kilómetros y se determina que el (VO_2 máx) es 85 ml/kg/min. Se conoce también que el (VO_2 máx) sigue una distribución normal con una desviación estándar $\sigma = 6$ ml/kg/min.



Fuente: Elaboración propia mediante IA (Copilot)

CAPÍTULO 5

Pruebas de hipótesis para una muestra

5.1 Definiciones básicas

Una vez que se han estudiado todos los conceptos básicos de la estadística descriptiva, probabilidades y estimación de parámetros, es momento de desarrollar una herramienta muy importante dentro de la estadística inferencial como son las pruebas de hipótesis, que en este libro se estudian solamente las pruebas para una muestra. Se empieza definiendo una **hipótesis de investigación** como una proposición planteada por el investigador con la intención de responder a un problema de estudio y se redacta en lenguaje natural. A continuación, se presentan los conceptos básicos de este capítulo.

Hipótesis estadística es una proposición o afirmación, la misma que está sujeta a verificación. Se redacta en términos en lenguaje estadístico formal, siendo la expresión operativa de la hipótesis de investigación.

En términos estadísticos, estas hipótesis se pueden plantear sobre los parámetros, las relaciones entre variables y sobre el tipo de distribución de probabilidad que tiene una variable.

Se tienen **dos tipos de hipótesis estadísticas**, la hipótesis nula H_0 y la hipótesis alternativa H_1 , las mismas que son opuestas entre sí. La hipótesis nula H_0 representa el estado inicial, la igualdad o ausencia de efecto.

Antes de formular estas hipótesis, es importante identificar correctamente las variables de interés de forma que sean operativas.

En el proceso de verificación estadística de la hipótesis se llama **Prueba de hipótesis**.

5.1.1 Clasificación de las Pruebas de Hipótesis

5.1.1.1 TIPOS DE PRUEBAS POR LOS SUPUESTOS SOBRE LA POBLACIÓN

Dependiendo del cumplimiento de supuestos sobre la distribución de probabilidad y sobre la variabilidad, entre otros supuestos, las pruebas de hipótesis se clasifican de la siguiente manera:

- Paramétricas
- No Paramétricas

Las pruebas paramétricas se pueden plantear sobre parámetros como la media, la varianza, la proporción, entre otros. Estas pruebas requieren el cumplimiento de supuestos como la normalidad, entre otros, para su aplicación.

Ante el incumplimiento de estos supuestos, se desarrollan las pruebas no paramétricas para verificar los supuestos sobre medidas descriptivas, así como sobre las relaciones entre variables categóricas; sobre la distribución de probabilidad.

Estas dos últimas pruebas no paramétricas se estudiarán en el laboratorio al final de este capítulo, por ser importantes en las pruebas sobre una muestra.

Ejemplo 5-1. Hipótesis paramétrica

El peso promedio de una mujer con una estatura de 1.50 metros es de 110 libras.

Ejemplo 5-2. Hipótesis no paramétricas

- La actividad preferida en el tiempo libre depende del género del encuestado.
- El peso de los bebés recién nacidos sigue una distribución normal.

5.1.1.2 TIPOS DE PRUEBAS POR LA DIRECCIÓN DE LA HIPÓTESIS ALTERNA

Para el caso de las pruebas de hipótesis donde el interés recae en comprobar un supuesto sobre un parámetro, las hipótesis se plantean usando relaciones de igualdad o desigualdad, y también se utilizan relaciones de orden. En este sentido, se tienen tres tipos de pruebas. El nombre o tipo de prueba está dado por la hipótesis alternativa, como se indica a continuación.

- Bilateral, por cuanto la hipótesis alternativa se expresa $\neq$ lo que implica dos posibilidades menor que: $<$ y mayor que: $>$.
- Unilateral superior, la hipótesis alternativa se expresa con mayor que: $>$.
- Unilateral inferior, cuando, la hipótesis alternativa se expresa con menor que: $<$

A continuación, se describen las hipótesis nula y alternativa para cada tipo de prueba, considerando el parámetro θ y un valor empírico θ_0

- Bilateral (Dos colas)

$$H_0 : \theta = \theta_0$$

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

- Unilateral superior (Cola superior)

$$H_0 : \theta \leq \theta_0$$

$$H_1 : \theta > \theta_0$$

- Unilateral inferior (Cola inferior)

$$H_0 : \theta \geq \theta_0$$

$$H_1 : \theta < \theta_0$$

El valor empírico θ_0 tomado para el parámetro θ es dado por expertos, es el resultado de investigaciones similares o es un valor teórico.

Ejemplo 5-3. Identifique el tipo de prueba, la variable, parámetro y escriba las hipótesis estadísticas del ejemplo 5.1

Prueba: bilateral, Variable: peso (kg), Parámetro: peso promedio (kg)

$$H_0 : \mu = 110$$

$$H_1 : \mu \neq 110$$

Ejemplo 5-4. De acuerdo a cifras del ministerio de educación (2015-2016), más del 50 % de los niños menores de 3 años no acceden al subnivel 1 de la educación inicial. Use este dato para plantear una hipótesis, identificando el tipo de prueba, la variable, el parámetro para plantear hipótesis estadísticas.

Prueba: unilateral superior, Variable: Número de niños que no acceden al subnivel 1, Parámetro: proporción de niños que no acceden al subnivel 1.

$$H_0 : p \leq 0.5$$

$$H_1 : p > 0.5$$

Ejemplo 5-5. Identifique el tipo de prueba del ejemplo 5.2 y escriba las hipótesis estadísticas para el género (X) y el pasatiempo (Y)

. No tiene dirección determinada.

$$\begin{cases} H_0 : X \text{ e } Y \text{ son independientes} \\ H_1 : X \text{ e } Y \text{ son dependientes} \end{cases}$$

5.1.2 Errores de Tipo I y tipo II

Al realizar una prueba de hipótesis se pueden cometer dos tipos de errores:

- Error de tipo I: $\alpha = P(\text{Rechazar } H_0 \mid H_0 \text{ es verdadera})$
- Error de tipo II: $\beta = P(\text{Aceptar } H_0 \mid H_0 \text{ es falsa})$

Estos errores se presentan por el estado de la hipótesis y la decisión tomada, como se observa en la Tabla 5.2

Tabla 5.1

Decisión vs Estado de la hipótesis

Estado	Rechazar H_0 : R	Aceptar H_0 : A
H_0 verdadera: V	Error de tipo I $\alpha = P(R V)$	No hay error $P(A V)$
H_0 falsa: F	No hay error $P(R F)$	Error de tipo II $\beta = P(A F)$

Figura 5.1

Curvas y errores para H_0 y H_1



5.1.3 Pasos para realizar una prueba de hipótesis

Previo al proceso de verificación de las hipótesis estadísticas, se debe contar con hipótesis de investigación, identificar las variables y hacerlas operativas, así como establecer los valores de los parámetros a comprobar, o las relaciones con otras variables; tomar los datos de la muestra. Verificar las condiciones para seguir un proceso paramétrico o no paramétrico. En todo caso, el proceso paramétrico o no paramétrico sigue estos pasos generales:

1. Establecer la hipótesis nula H_0 y la hipótesis alternativa H_1 .
2. Seleccionar el nivel de significancia α y el o los puntos críticos correspondientes en la distribución de probabilidad.
3. Hallar las medidas descriptivas de la muestra y el estadístico de prueba apropiado.
4. Establecer la regla de decisión, o región de rechazo de la hipótesis nula.
5. Interpretar los resultados, que estará implícito en los pasos.

Esquema de los pasos en la figura 5.2

Figura 5.2

Pasos de la prueba de hipótesis



Imagen editada con IA

5.2 Prueba de hipótesis sobre la media poblacional μ

De igual forma que se introdujeron las condiciones en los intervalos de confianza para la media, se tiene una variable aleatoria X con distribución normal o aproximadamente normal, con estadísticos estandarizados respectivamente Z y T :

$$Z = \frac{\bar{x} - \mu_x}{\frac{\sigma_x}{\sqrt{n}}}$$
$$T = \frac{\bar{x} - \mu_x}{\frac{s_x}{\sqrt{n}}}$$

Para comprobar una hipótesis sobre la media poblacional μ se presentan dos casos:

- Caso 1. Prueba de hipótesis sobre la media con varianza poblacional conocida.
- Caso 2. Prueba de hipótesis sobre la media con varianza poblacional desconocida.

En esta sección se estudian las pruebas de hipótesis tanto bilaterales como unilaterales.

5.2.1 Prueba de hipótesis sobre la media μ con varianza σ^2 conocida

Si X es una variable aleatoria con distribución normal con varianza poblacional σ^2 conocida, de la cual se toma una muestra aleatoria de tamaño n : $X_1, X_2, \dots, X_n$ con media muestral $\bar{x}$, entonces una hipótesis sobre la media μ se verifica mediante los cinco pasos mencionados anteriormente y de acuerdo con uno de los tres tipos de pruebas.

5.2.1.1 PRUEBA BILATERAL SOBRE μ

1. Establecer las hipótesis nula y alternativa. Se toma μ_0 como valor empírico.

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

2. Seleccionar el nivel de significancia α y los puntos críticos $Z_{\alpha/2} =$

3. Hallar el estadístico de prueba:

$$Z_0 = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

4. Regla de decisión

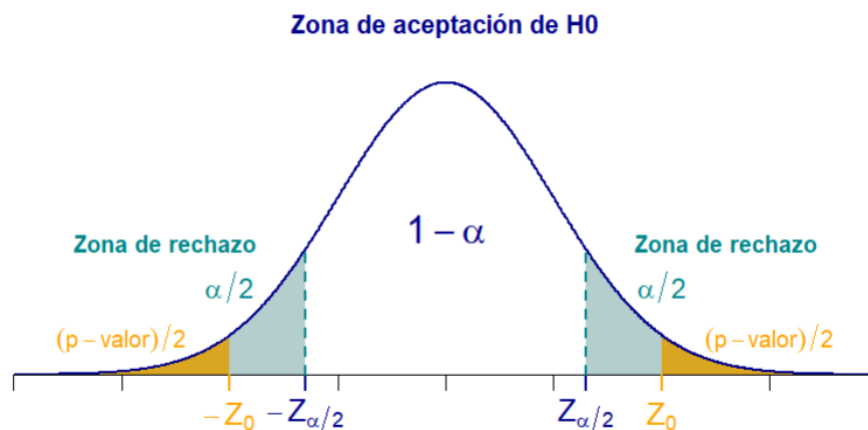
H_0 se rechaza, si $Z_0 > Z_{\alpha/2}$ o si $Z_0 < -Z_{\alpha/2}$, o mas bien,

H_0 se rechaza, si $|Z_0| > +Z_{\alpha/2}$

En la figura 5.3 están representados los puntos críticos $\pm Z_{\alpha/2}$ que delimitan la zona de aceptación de la hipótesis nula con probabilidad $1 - \alpha$, así como las zonas de rechazo de la hipótesis nula con probabilidad o significancia $\alpha/2$; además, del estadístico de prueba Z_0 ubicado en dos zonas: con probabilidad a la izquierda de su valor negativo y probabilidad a la derecha del valor positivo, y la suma de probabilidades se denominada p-valor.

Figura 5.3

Puntos, estadísticos y probabilidades en la prueba bilateral



5.2.1.2 PRUEBA UNILATERAL SUPERIOR SOBRE μ

1. Hipótesis estadísticas

$$\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

2. Significancia α y punto críticos $Z_{\alpha} = +$

3. Estadístico de prueba:

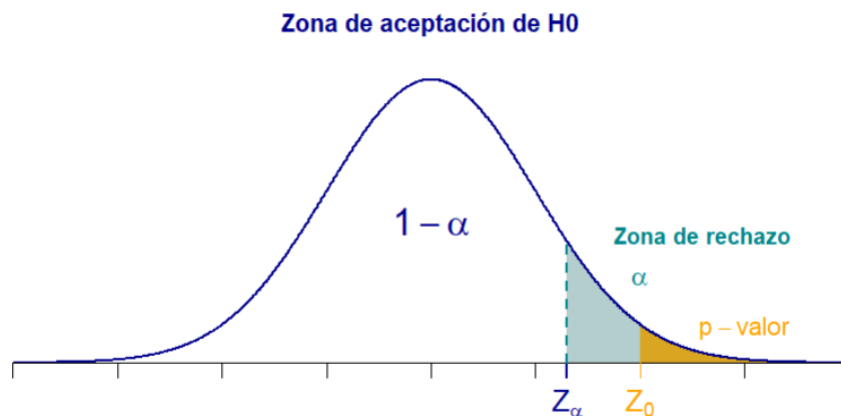
$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

4. Regla de decisión: H_0 se rechaza, si $Z_0 > +Z_\alpha$

La figura 5.4 representa el punto crítico Z_α que delimita la zona de aceptación de la hipótesis nula con probabilidad $1 - \alpha$, así como la zona de rechazo de la hipótesis nula con probabilidad o significancia α ; el posible estadístico de prueba Z_0 con probabilidad a la derecha denominado p-valor.

Figura 5.4

Puntos, estadísticos y probabilidades, prueba unilateral superior



5.2.1.3 PRUEBA UNILATERAL INFERIOR SOBRE μ

1. Hipótesis estadísticas

$$\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

2. Significancia α y punto críticos $Z_\alpha = -$

3. Estadístico de prueba:

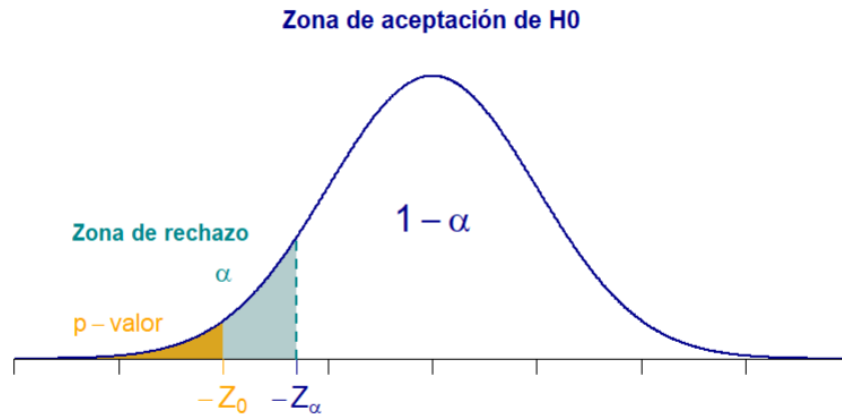
$$Z_0 = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

4. Regla de decisión: H_0 se rechaza, si $Z_0 < -Z_\alpha$

La figura 5.5 representa el punto crítico $-Z_\alpha$ que delimita la zona de aceptación de la hipótesis nula con probabilidad $1 - \alpha$, así como la zona de rechazo de la hipótesis nula con probabilidad o significancia α ; el posible estadístico de prueba Z_0 con probabilidad a la izquierda denominado p-valor.

Figura 5.5

Puntos, estadísticos y probabilidades, prueba unilateral superior



5.2.2 Prueba de hipótesis sobre la media μ con varianza σ^2 desconocida

Si X es una variable aleatoria con distribución normal y varianza poblacional σ^2 desconocida, de la que se toma una muestra aleatoria de tamaño n : $X_1, X_2, \dots, X_n$, con media muestral $\bar{x}$, entonces una hipótesis sobre la media μ se verifica de forma similar a lo realizado en la sección anterior, pero ahora se usa la distribución t-Student con grados de libertad $gl = n - 1$ y en el estadístico se usa la desviación estándar muestral. A continuación, se exponen los tres tipos de pruebas.

5.2.2.1 PRUEBA BILATERAL SOBRE μ

1. Establecer las hipótesis nula y alternativa. Se toma μ_0 como valor empírico.

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

2. Seleccionar el nivel de significancia α y los puntos críticos $t_{\alpha/2, gl} =$
3. Hallar el estadístico de prueba:

$$T_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

4. Regla de decisión

H_0 se rechaza, si $T_0 > T_{\alpha/2, gl}$ o si $T_0 < -T_{\alpha/2, gl}$, o más bien,

H_0 se rechaza, si $|T_0| > T_{\alpha/2, gl}$

La representación gráfica es similar a la expuesta en la figura 5.3 pero ahora se debe considerar la distribución t-Student.

5.2.2.2 PRUEBA UNILATERAL SUPERIOR SOBRE μ

1. Establecer las hipótesis nula y alternativa. Se toma μ_0 como valor empírico.

$$\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

2. Seleccionar el nivel de significancia α y los puntos críticos $t_{\alpha, gl} = +$
3. Hallar el estadístico de prueba:

$$T_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

4. Regla de decisión: H_0 se rechaza, si $T_0 > +T_\alpha$

La representación gráfica es similar a la expuesta en la figura 5.4, pero ahora se debe considerar la distribución t-Student.

5.2.2.3 PRUEBA UNILATERAL INFERIOR SOBRE μ

1. Establecer las hipótesis nula y alternativa. Se toma μ_0 como valor empírico.

$$\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

2. Seleccionar el nivel de significancia α y los puntos críticos $t_{\alpha, gl} = -$
3. Hallar el estadístico de prueba:

$$T_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

4. Regla de decisión: H_0 se rechaza, si $T_0 < -T_\alpha$

La representación gráfica es similar a la figura 5.5, pero ahora se debe considerar la distribución t-Student.

A continuación se expone otra forma realizar el paso 4 en las pruebas de hipótesis que se han revisado.

5.3 Uso del p-valor

El p-valor, se define como el mínimo valor de probabilidad con el que la hipótesis nula se rechaza.

La regla de decisión para cualquier tipo de prueba usando el p-valor es:
 Ho se rechaza si $p - valor < \alpha$

El cálculo del p-valor depende de la distribución de probabilidad y del tipo de prueba. La representación gráfica del p-valor se encuentra en los tres casos de pruebas de hipótesis sobre la media tales como figura 5.3, figura 5.4, figura 5.5.

Para el caso de la distribución normal estándar, se tiene:

$$\begin{aligned} \text{Prueba bilateral:} & \quad p - valor = 2 * P(Z > |Z_0|) \\ \text{Prueba unilateral superior:} & \quad p - valor = P(Z > Z_0) \\ \text{Prueba unilateral inferior:} & \quad p - valor = P(Z < Z_0) \end{aligned}$$

Para el caso de la distribución normal t-Student se tiene:

$$\begin{aligned} \text{Prueba bilateral:} & \quad p - valor = 2 * P(T_{gl} > |T_0|) \\ \text{Prueba unilateral superior:} & \quad p - valor = P(T_{gl} > T_0) \\ \text{Prueba unilateral inferior:} & \quad p - valor = P(T_{gl} < T_0) \end{aligned}$$

Como se observa, el p-valor de la prueba bilateral es el doble de la probabilidad de una de las colas por la propiedad de simetría de la distribución normal y también de la distribución t-Student. En estos dos casos se pueden aplicar las propiedades de evento contrario o de simetría según sea conveniente, ya que las tablas de las dos distribuciones están construidas solamente con probabilidades acumulativas en cola izquierda. En el caso de usar software estadístico, esto podría no ser necesario dependiendo de las opciones que brinde dicho software.

Ejemplo 5-6. Un fabricante de juguetes educativos afirma que el tiempo que un niño de 6 años tarda en armar cierto rompecabezas sigue una distribución normal con una media $\mu = 12$ minutos, y una desviación estándar conocida de $\sigma = 2.4$ minutos. Para verificar esta afirmación una psicóloga especializada en desarrollo cognitivo toma una muestra de 30 niños de su locali-

dad, donde el tiempo promedio del armado del rompecabezas es 11.2 minutos. Plantear y verificar las hipótesis apropiadas al nivel de significancia de $\alpha = 0.05$.

La hipótesis de investigación para esta prueba bilateral es:

El tiempo promedio de armado de cierto rompecabezas es 12 minutos. Este caso corresponde a una prueba bilateral sobre la media con varianza conocida.

1. Hipótesis estadísticas:

$$\begin{cases} H_0 : \mu = 12 \\ H_1 : \mu \neq 12 \end{cases}$$

2. Para $\alpha = 0.05$ y los puntos críticos son $Z_{\alpha/2} = Z_{0.025} = \pm 1.96$

3. Estadístico de prueba:

$$Z_0 = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{11.2 - 12.0}{\frac{2.4}{\sqrt{30}}} \approx -1.826$$

4. Regla de decisión: H_0 se rechaza, si $|Z_0| > +Z_{\alpha/2}$

$$1.826 > +1.96 \quad \text{Falso}$$

Por lo tanto, no hay evidencia estadística suficiente para determinar que el tiempo promedio de armado de cierto rompecabezas en niños de 6 años sea diferente del tiempo promedio de armado indicado por el fabricante: 12 minutos.

Ejemplo 5-7. Tome la decisión sobre el ejemplo anterior usando el p-valor.

Regla de decisión H_0 se rechaza si $p - \text{valor} < \alpha$

Cálculo del p-valor

$$p - \text{valor} = 2 * P(Z > |Z_0|)$$

$$p - \text{valor} = 2 * P(Z > 1.83)$$

$$p - \text{valor} = 2 * P(Z < -1.83)$$

$$p - \text{valor} = 2 * 0.03362 = 0.06725$$

$$0.06725 < 0.05 \quad \text{Falso}$$

Por lo tanto, no hay evidencia estadística suficiente para rechazar H_0 , o se podría escribir la misma conclusión anterior.

Ejemplo 5-8. El Índice de Masa Corporal (IMC) de maratonistas hombres de élite sigue una distribución normal con una media $\mu = 19.5 \text{ kg/m}^2$ y una desviación estándar σ desconocida. Determine si el IMC promedio es mayor o igual al valor referencial. Utilice un nivel de significancia de 0.05 y una muestra de 36 maratonistas donde el IMC promedio es $\bar{x} = 20 \text{ kg/m}^2$ con una desviación estándar $s = 0.92 \text{ kg/m}^2$

Hipótesis de investigación: **El Índice de Masa Corporal de maratonistas hombres élite es mayor o igual al valor referencial que 19.5 kg/m^2 .**

Este caso corresponde a una prueba unilateral inferior sobre la media con varianza desconocida.

1. Hipótesis estadísticas:

$$\begin{cases} H_0 : \mu \leq 19.5 \\ H_1 : \mu < 19.5 \end{cases}$$

2. Para $\alpha = 0.05$ y $gl = n - 1 = 36 - 1 = 35$, se tiene el punto crítico:

$$t_{\alpha, gl} = t_{0.05, 35} = -1.690$$

Obtenido de la tabla de la distribución t-Student, tal como se observa en la figura 5.6

Figura 5.6

Puntos críticos negativos y probabilidades en cola izquierda

g.l	α						
	0.01	0.015	0.02	0.025	0.04	0.05	0.1
34	-2.441	-2.265	-2.136	-2.032	-1.805	-1.691	-1.307
35	-2.438	-2.262	-2.133	-2.030	-1.803	-1.690	-1.306
40	-2.423	-2.250	-2.123	-2.021	-1.796	-1.684	-1.303
45	-2.412	-2.241	-2.115	-2.014	-1.791	-1.679	-1.301

3. Estadístico de prueba:

$$T_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{20 - 19.5}{\frac{0.92}{\sqrt{36}}} \approx 3.26$$

4. Regla de decisión: H_0 se rechaza, si $T_0 < -t_\alpha$

$$3.26 < -1.690 \quad \text{Verdadero}$$

Como se observa, el estadístico de prueba T_0 no cae en la región de rechazo de la hipótesis nula H_0 , por lo tanto, no hay evidencia suficiente para rechazar la hipótesis nula. En este caso, se mantiene la hipótesis nula: el IMC promedio es mayor o igual a 19.5.

Ejemplo 5-9. Tome la decisión sobre el ejemplo anterior usando el p-valor.

Regla de decisión y cálculo del p-valor:

$$H_0 \text{ se rechaza si } p - \text{valor} < \alpha$$

$$p - \text{valor} = P(T_{gl} < |T_0|)$$

$$p - \text{valor} = P(T_{35} < 3.26)$$

Se busca la probabilidad correspondiente al valor de $T=3.26$ en la distribución t-Student en cola izquierda, la cual no existe, por lo que se procede a hallar un rango de valores de probabilidad correspondientes a los valores que más se aproximan por la izquierda y por la derecha al valor de $T=3.26$ en la fila de 35 grados de libertad. En la figura 5.7 se observan dichos valores.

Figura 5.7

Puntos críticos positivos y probabilidades en cola izquierda

g.l	α										
	0.9	0.95	0.96	0.975	0.98	0.985	0.99	0.995	0.9975	0.999	0.9995
34	1.307	1.691	1.805	2.032	2.136	2.265	2.441	2.728	3.002	3.348	3.601
35	1.306	1.690	1.803	2.030	2.133	2.262	2.438	2.724	2.996	3.340	3.591
40	1.303	1.684	1.796	2.021	2.123	2.250	2.423	2.704	2.971	3.307	3.551

De acuerdo a la figura, T se encuentra en el siguiente intervalo:

$$2.996 < t_{35} < 3.340$$

Por ende, sus probabilidades se encuentran en el intervalo:

$$0.9975 < P(T_{35}) < 0.9990$$

$$0.9975 < p - \text{valor} < 0.9990$$

Finalmente, al comparar con la significancia del problema que es $\alpha = 0.05$, se llega a la misma conclusión que en el ejemplo anterior.

$$0.9975 < p - \text{valor} < 0.9990 < \alpha = 0.05 \quad (\text{Falso})$$

5.4 Prueba de hipótesis sobre la proporción poblacional p

Para la variable aleatoria binomial X : número de éxitos, se obtiene una muestra aleatoria de tamaño n , donde $\hat{p} = x/n$ es el estimador puntual de la proporción poblacional p . Se conoce que la distribución de muestreo de $\hat{p}$ es aproximadamente normal (siempre que $np > 5$) con media np y varianza $\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}$. Entonces, una hipótesis sobre la proporción poblacional se verifica con los siguientes pasos y tipos de pruebas según corresponda:

Tabla 5.2

Pasos de la prueba de hipótesis

N.	Bilateral	Unilateral Inferior	Unilateral Superior
1.	$H_0 : p = p_0$ $H_1 : p \neq p_0$	$H_0 : p \leq p_0$ $H_1 : p > p_0$	$H_0 : p \geq p_0$ $H_1 : p < p_0$
2.	α $Z_{\alpha/2} = \pm$	α $Z_{\alpha} = +$	α $Z_{\alpha} = -$
3.	$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}, \quad \hat{p} = \frac{x}{n}, \quad x : \text{número de éxitos en } n$		
4.	H_0 se rechaza si: $ Z_0 > +Z_{\alpha/2}$	H_0 se rechaza si: $Z_0 > +Z_{\alpha}$	H_0 se rechaza si: $Z_0 > +Z_{\alpha}$

Ejemplo 5-10. Hipótesis sobre la proporción

- Si la variable aleatoria es el número de menores que viven en contextos familiares no funcionales. Se puede plantear la hipótesis: La proporción de menores que viven en contextos familiares no funcionales es inferior a 0.08.
- En actividad física se podría plantear la hipótesis: La proporción de deportistas lesionados durante el calentamiento es 0.25.

Ejemplo 5-11. De acuerdo al Censo 2020 publicados por INEGI de México, en albergues para mujeres o víctimas de violencia, 77.1% de los niños de 3 a 5 años no asistían a la escuela (<https://www.inegi.org.mx/programas/ccpv/2020/>). Actualmente se quiere determinar si esta cifra se ha incrementado en la población, por lo que, se toma una muestra de $n = 480$ niños viviendo en albergues, y se determina que 391 niños no asisten a la escuela. Plantear y comprobar hipótesis apropiadas con un nivel de significancia de 0.05.

Hipótesis de investigación: **En un porcentaje inferior a 77.1 % de niños de 3 a 5 años no asisten a la escuela.**

Este caso corresponde a una prueba unilateral superior sobre la proporción.

1. Hipótesis estadísticas:

$$\begin{cases} H_0 : p \leq 0.771 \\ H_1 : p > 0.771 \end{cases}$$

2. Para $\alpha = 0.05$ el punto crítico es: $Z_\alpha = Z_{0.05} = +1.645$

3. Estadístico de prueba Z_0 :

Proporción de niños que no asisten a la escuela en la muestra:

$$\hat{p} = \frac{x}{n}$$

$$\hat{p} = \frac{391}{480} = 0.815$$

Estadístico de prueba:

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$$Z_0 = \frac{0.815 - 0.771}{\sqrt{\frac{0.771(1-0.771)}{480}}} = 2.483$$

4. Regla de decisión:

H_0 se rechaza, si $Z_0 > -Z_\alpha$

$$2.483 > 1.645 \quad \text{Verdadero}$$

Se rechaza la hipótesis nula $H_0 : p \leq 0.771$ al nivel de significancia de 0.05, es decir, existe evidencia estadísticamente significativa para aceptar la hipótesis alternativa $H_1 : p > 0.771$.

Dicho de otra forma, se puede afirmar que la proporción de niños de 3 a 5 años que no asisten a la escuela es mayor que 0.771 con una confianza del 95 %.

Ejemplo 5-12. Tome la decisión sobre el ejemplo anterior usando el p-valor.

Regla de decisión H_0 se rechaza si $p - \text{valor} < \alpha$

Cálculo del p-valor

$$p - \text{valor} = P(Z > Z_0)$$

$$p - \text{valor} = P(Z > 2.48)$$

$$p - \text{valor} = P(Z < -2.48)$$

$$p - \text{valor} = 0.023$$

$$0.00656 < 0.05 \quad \text{Verdadero}$$

Con un nivel de significancia de 0.05, existe evidencia estadística suficiente para afirmar que la proporción de niños de 3 a 5 años que no asisten a la escuela es mayor a 0.771.

El resultado de la prueba es estadísticamente significativo al nivel de significancia de 0.05, ya que el p-valor 0.00656 obtenido es menor que $\alpha = 0.05$. Por lo tanto, existe evidencia suficiente para concluir que la proporción de niños de 3 a 5 años que no asisten a la escuela es mayor que 0.771.

Ejemplo 5-13. De acuerdo al artículo ¿Cuánto se lesiona un jugador profesional del fútbol?, disponible en el sitio web Barça Innovation Hub, se ha determinado que tasa de lesión de un jugador profesional del fútbol oscila entre 0.36 y 0.39. Utilice esta información para comprobar que la tasa de lesión es mayor o igual que 0.36 aplicado a otro equipo competitivo, para lo cual se toma una muestra de $n = 84$ futbolistas, de quienes 22 han sufrido alguna lesión con la práctica del fútbol profesional. Tome la decisión usando el p-valor.

Hipótesis de investigación: La tasa de lesión de un jugador de fútbol profesional es mayor o igual que 0.36.

En este caso, la hipótesis de investigación es nula, y como el nombre de la prueba está dado por la hipótesis alternativa, entonces la prueba es de tipo unilateral inferior.

Figura 5.8

Lesiones en fútbol profesional



<https://imfisfisioterapia.com/>

1. Hipótesis estadísticas:

$$\begin{cases} H_0 : p \geq 0.36 \\ H_1 : p < 0.36 \end{cases}$$

2. $\alpha = 0.05$, pero no es necesario hallar el punto crítico, cuando se toma la decisión usando del p-valor.

3. Estadístico de prueba Z_0 :

Proporción de futbolistas lesionados:

$$\hat{p} = \frac{x}{n} = \frac{19}{84} = 0.226$$

Estadístico de prueba:

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$
$$Z_0 = \frac{0.226 - 0.36}{\sqrt{\frac{0.226*0.774}{84}}} = -3.51$$

4. Regla de decisión: H_0 se rechaza, si $p - valor < -\alpha$

$$p - valor = P(Z < -3.51)$$

$$p - valor = 0.00022$$

$$0.00022 < 0.05 \quad \text{Verdadero}$$

Se rechaza la hipótesis nula $H_0 : p \geq 0.35$ al nivel de significancia de 0.05, es decir, existe evidencia significativa para aceptar la hipótesis alternativa $H_1 : p < 0.35$. O también, se puede concluir que, en la población estudiada, se puede afirmar que la proporción de futbolistas que sufren algún tipo de lesión es menor que 0.35 al nivel de confianza de 0.95.

5.4.1 Laboratorio de Pruebas de hipótesis

5.4.1.1 PRUEBA DE HIPÓTESIS SOBRE LA MEDIA CON JAMOVİ

Ejemplo 5-14. Con una significancia del 5 % comprobar que el Peso promedio es 145 para el grupo de personas del conjunto de 90 datos del Apéndice A

1. Hipótesis estadísticas

$$\begin{cases} H_0 : \mu = 145 \\ H_1 : \mu \neq 145 \end{cases}$$

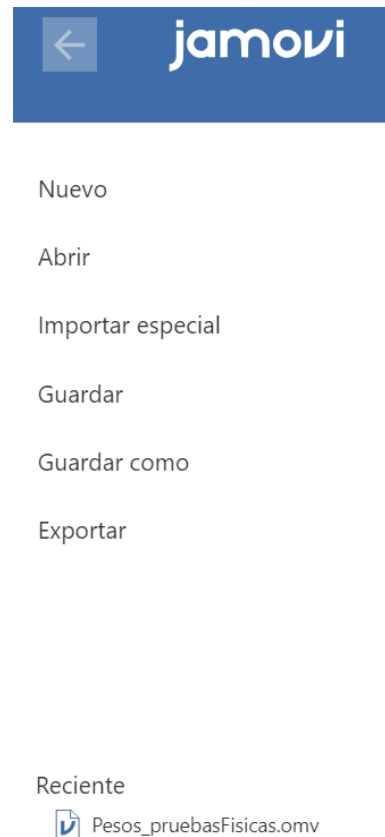
2. Significancia $\alpha = 0.05$

3. Proceso en Jamovi

Previamente debe crear los datos (Nuevo), abrir datos existentes (Abrir, Reciente) o importar los datos como se observa en la Figura 5.9. Los datos y resultados se guardan con la extensión **omv**.

Figura 5.9

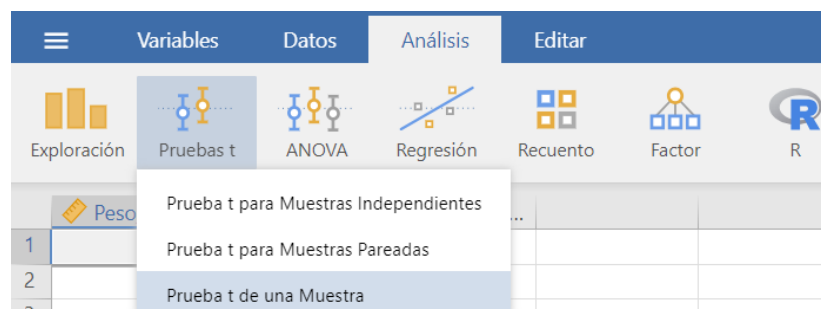
Menú Abrir datos



Luego, del Menu **Análisis** seleccionar el menú pruebas t según la Figura 5.10.

Figura 5.10

Menú pruebas t



Completar los campos que se observan en la Figura 5.11, para el caso bilateral en el campo de Hipótesis se elige el valor de la prueba 145 y se marca $\neq$ que corresponde a la hipótesis alternativa. Finalmente, ingresar la variable *Peso* al campo de las variables dependientes.

Figura 5.11

Opciones de el Menú pruebas t para una muestra

Los resultados de la prueba t se observan en la Figura 5.12, donde el p-valor es $7.11e - 13 < 0.05$, por lo tanto, H_0 se rechaza y se acepta.

Ejemplo 5-15. Tomando los datos del ejemplo anterior, comprobar que, el *Peso* promedio es menor o igual que 145 libras.

Figura 5.12*Salida en Jamovi***Prueba t de una Muestra**

Prueba t de una Muestra

		Estadístico	gl	p
Peso_lb	T de Student	8.382	89.00	7.115e-13

Nota. $H_a: \mu \neq 145$ **1. Hipótesis estadísticas**

$$\begin{cases} H_0 : \mu \leq 145 \\ H_1 : \mu > 145 \end{cases}$$

2. Significancia $\alpha = 0.05$ **3. Proceso en Jamovi**

Siguiendo el proceso anterior, elegir **mayor que** del menú mostrado en la Figura 5.11, marcando la opción específica de la Figura 5.13 que corresponde a la hipótesis alternativa. Esta selección da como resultado los estadísticos de la Figura 5.14

Figura 5.13*Salida en Jamovi*

Hipótesis

Valor de prueba

☐ $\neq$ Valor de prueba
☒ $>$ Valor de prueba
☐ $<$ Valor de prueba

Figura 5.14*Salida en Jamovi*

Prueba t de una Muestra

		Estadístico	gl	p
Peso_lb	T de Student	8.382	89.00	3.557e-13

Nota. $H_a: \mu > 145$

El $p\text{-valor} = 3.55e^{-13}$ es inferior a la significancia dada $\alpha = 0.05$ por lo que, se tiene evidencia estadísticamente significativa para rechazar la hipótesis nula H_0 . Es decir, hay evidencia suficiente para concluir que la media poblacional μ es significativamente mayor que 145.

5.4.1.2 PRUEBA DE HIPÓTESIS SOBRE LA PROPORCIÓN EN JAMOVİ

Ejemplo 5-16. Con una significancia del 5 % comprobar que la proporción de personas que realizan actividad física a un nivel alto es mayor o igual que 0.20. Tomar el conjunto A1 Apéndice A

1. Hipótesis estadísticas

$$\begin{cases} H_0 : p = 0.2 \\ H_1 : p \neq 0.2 \end{cases}$$

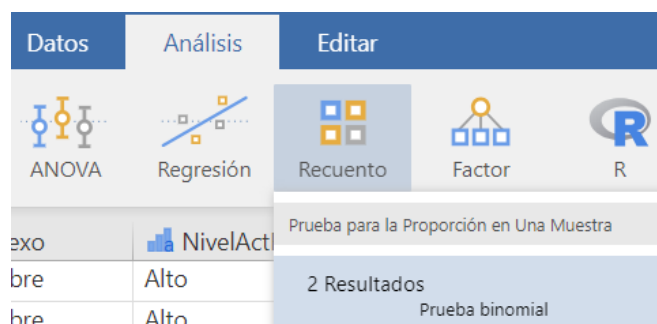
2. Significancia $\alpha = 0.05$

3. Proceso en Jamovi Previamente debe crear los datos (Nuevo), abrir datos existentes o importar los datos como se observa en la Figura 5.9. Los datos y resultados se guardan con la extensión **omv**.

Luego seleccionar el Menú **Análisis, Recuento, 2 Resultados, Prueba Binomial**, como se observa en la Figura 5.15

Figura 5.15

Menú Prueba Binomial



Selección de opciones en el menú, cuyos campos a llenar se observan en la Figura 5.16

- Ingresar la variable nivel de actividad física
- Ingresar el valor empírico 0.2
- En hipótesis, marcar la tercera opción correspondiente a prueba unilateral inferior

Figura 5.16

Opciones de la Prueba Binomial

Los resultados se observan en la Figura 5.17. Representados por una tabla con valores observados, y las dos últimas columnas contienen proporciones de éxito para tres pruebas simultáneas para los atributos Alto, Bajo y Medio con los respectivos p – *valor*.

Figura 5.17

Opciones de la Prueba Binomial

Prueba de Proporciones (2 Resultados)

Prueba Binomial					
	Nivel	Frecuencia	Total	Proporción	p
NivelActFisica	Alto	10	90	0.1111	0.01896
	Bajo	41	90	0.4556	1.00000
	Medio	39	90	0.4333	1.00000

Nota. H_a es proporción < 0.2

En este caso se toma la primera fila que corresponde a la hipótesis planteada, donde p – *valor* = 0.01896 es menor que la significancia dada de 0.05, por lo tanto, existe evidencia estadísticamente significativa para rechazar la hipótesis nula H_0 . Es decir, hay evidencia suficiente para concluir que la proporción de personas que realizan actividad física a un nivel alto es menor que 0.2.

5.5 Ejercicio propuestos del capítulo

Ejercicios para Educación inicial

5.1. El tiempo promedio de atención en niños de 3 años es de 7 minutos. Para comprobar el supuesto, una maestra elige como actividad la lectura y registra el tiempo de atención sostenida con una muestra de 30 niños de 3 años. En la muestra, el tiempo promedio de atención es de 6.2 minutos, con una desviación estándar de 1.5 minutos. ¿Existe evidencia estadística para afirmar que el tiempo promedio de atención es diferente del valor establecido de 7 minutos?. Tome la decisión con una significancia de 0.05.

5.2. Verificar que el número promedio de palabras que usan los niños de 5 años en una narración libre supera las 50 palabras. A una muestra de 25 niños de 5 años se les pidió que contaran una historia libre y se encontró que, en promedio, usan 56 palabras, con una desviación estándar de 7 palabras. ¿Se puede afirmar que los niños de 5 años usan más de 50 palabras en promedio?. Use $\alpha = 0.05$

5.3. Con una significancia de 0.05 comprobar que la proporción de niños de 3 años que reconocen colores básicos es mayor a 0.6. Para comprobar el supuesto, se toma una muestra de 50 niños de 3 años y se determina que 36 logran identificar correctamente al menos tres colores básicos (rojo, azul, amarillo). ¿Se puede afirmar que más del 60 % de los niños de 3 años reconocen colores básicos?

5.4. Con una significancia de 0.01, comprobar que al menos el 80 % de los niños de 5 años pueden escribir su nombre completo. En una muestra de 36 niños de 5 años, 28 escriben correctamente su nombre completo. ¿Se puede afirmar que al menos el 80 % de los niños de 5 años escriben correctamente su nombre completo?

Ejercicios para Actividad Física

5.5. Comprobar si los jugadores profesionales de tenis realizan en promedio más de 10 aces por partido. En una muestra de 20 partidos, se obtuvo una media de 8.1 aces con una desviación estándar de 1.1. Como dato informativo, se conoce que, en el circuito ATP el número promedio de aces ronda entre 8 y

12 aces por partido por jugador, sin embargo, esto puede variar según el tipo de superficie. Use $\alpha = 0.05$

5.6. Se quiere saber si la puntuación promedio en salto ecuestre en una competencia regional es inferior a 85 puntos. En una muestra de 25 jinetes, la media fue de 83.2 puntos, con una desviación estándar de 4.5. Utilice un nivel de significancia de 0.01.

5.7. En una maratón con 200 participantes, 138 completaron la carrera en menos de 4 horas. ¿Con una significancia de $\alpha = 0.05$ se puede afirmar que al menos el 65 % de los corredores logran ese tiempo?

5.8. En una muestra de 50 gimnastas, 12 lograron una puntuación perfecta en el salto de potro. ¿Con una significancia de $\alpha = 0.05$ se puede afirmar que menos del 30 % de los gimnastas alcanzan puntuación perfecta?

5.9. En una prueba de orientación militar, 72 de 100 cadetes completaron el recorrido sin errores. ¿Con una significancia de $\alpha = 0.05$ se puede afirmar que más del 70 % de los cadetes tienen éxito en la navegación?

5.10. El consumo Máximo de Oxígeno (VO_2 máx) en corredores élite durante una carrera de 5 kilómetros puede ser 80 ml/kg/min. Se quiere comprobar si este valor referencial lo alcanzan en promedio un grupo de corredores élite, para lo cual se toma una muestra de 36 corredores en una prueba de 5 kilómetros y se determina que el (VO_2 máx) es 85 ml/kg/min. Se conoce también que el (VO_2 máx) sigue una distribución normal con una desviación estándar $\sigma = 6$ ml/kg/min

5.11. Utilizando los datos del ejercicio 5.1, tome la decisión utilizando el p-valor.

5.12. Utilizando los datos del ejercicio 5.2, tome la decisión utilizando el p-valor.

5.13. Utilizando los datos del ejercicio 5.3, ¿cuál sería la decisión tomando un nivel de significancia de 0.10?

5.14. Utilizando los datos del ejercicio 5.4, ¿cuál sería la decisión tomando un nivel de significancia de 0.10?

5.15. Utilizando los datos del ejercicio 5.6, tome la decisión utilizando el p-valor.

5.16. Utilizando los datos del ejercicio 5.7, tome la decisión utilizando el p-valor.

5.17. Utilizando los datos del ejercicio 5.8, ¿cuál sería la decisión, tomando un nivel de significancia de 0.10?

5.18. Utilizando los datos del ejercicio 5.9, ¿cuál sería la decisión tomando un nivel de significancia de 0.10?



CAPÍTULO 6

Variables en la práctica

6.1 Introducción

El objetivo de este capítulo es desarrollar un esquema de propuesta de análisis estadístico sin ejecutar el proceso descriptivo ni inferencial. El esquema está compuesto por elementos que sugieren el planteamiento de hipótesis, análisis descriptivos e inferenciales estudiados en este libro. Un aspecto importante que se incorpora en este esquema es la Operacionalización de una variable, lo que implica su definición en términos operativos que permitan su observación y medición [1].

6.1.1 Esquema de análisis

- Definición de la variable: Describe que se estudia y mide.
- Tipo de variable y nivel de medida: El conocimiento del tipo de variable determina los posibles parámetros que se pueden estimar.
- Operacionalización: Se determina como medir las variables, para las cuales se podrían identificar la dimensión, el indicador e incluso se puede seleccionar el instrumento para obtener la información.
- Descriptivas sugeridas: Se sugieren descriptivos para resumir los datos, tales como tablas, gráficos y medidas descriptivas.
- Hipótesis de investigación sugerida e hipótesis estadísticas: Las hipótesis se plantean en función de la naturaleza de las variables y/o parámetros.

6.2 Variables en el desarrollo infantil

6.2.1 Contexto motricidad

La motricidad es el conjunto de funciones neuromusculares que permiten al individuo realizar movimientos [2], [3]. La motricidad se clasifica en motricidad gruesa y motricidad fina.

- Motricidad gruesa: Movimientos de grandes grupos musculares (correr, saltar, trepar, entre otros). En la figura 6.1, se representa una actividad para evaluar la motricidad gruesa.
- Motricidad fina: Movimientos precisos que requieren coordinación de músculos pequeños (dibujar, abotonar, manipular objetos pequeños). En la figura 6.2, se aprecia una actividad para evaluar la motricidad fina.

Figura 6.1

Niños saltando



<https://www.freepik.com/>

Figura 6.2

Niño pasando cuentas



Imagen generada con IA

Observación: El estudio de la motricidad es de utilidad tanto en la educación inicial como en la actividad física.

6.2.1.1 MOTRICIDAD GRUESA

La motricidad gruesa se evalúa mediante actividades y el test de equilibrio estático y dinámico, la coordinación de piernas, la coordinación de brazos y piernas, el desplazamiento y la motricidad combinada.

6.2.1.2 TEST DE OZER

El Test Motor Ozeretki - Guillmain desarrollado para evaluar la edad motriz del niño de 3 a 8 años en 1936 en Rusia [4], con algunas adaptaciones, entre ellas la versión latinoamericana [5]. El Test utiliza una escala entre 0 y 3 puntos,

descrita en la tabla 6.1.

Tabla 6.1

Valoración de la actividad (acción)

Valor	Significado / Descripción
0	No realiza: No muestra la habilidad o no intenta
1	Lo intenta: Intenta pero no la completa correctamente.
2	Realiza parcialmente: Realiza con errores o incompleta.
3	Realiza correctamente: Ejecución adecuada y completa.

El test consta de 20 pruebas organizadas en 4 áreas: motricidad gruesa, Motricidad fina, Coordinación global, Rapidez y precisión. La motricidad gruesa se evalúa desde la pregunta 7 hasta la 12 descritas en la tabla 6.2 mediante el área y la actividad.

Tabla 6.2

Áreas y actividades para evaluar la motricidad gruesa

Area	Actividad
Equilibrio estático	Mantenerse de pie sobre un pie durante 5 segundos
Equilibrio dinámico	Caminar sobre una línea recta (dibujada en el suelo)
Coordinación de piernas	Saltar con ambos pies juntos hacia adelante
Coordinación de brazos y piernas	Lanzar una pelota grande hacia adelante
Desplazamiento	Correr de un punto a otro (5 metros)
Motricidad combinada	Subir y bajar escaleras sin ayuda

Ejemplo 6-1. Motricidad gruesa

- Variable: Motricidad gruesa en niños de 5 años.

Para evaluar la motricidad gruesa, se usa el Test Escala psicomotriz de Ozer en las 6 áreas correspondientes, cada una representada por una actividad.

- Tipo de variable: Cuantitativa
- Nivel de medida: Razón
- Operacionalización

La variable motricidad gruesa es el resultado de la suma de valores asignados a cada actividad en una escala de 0 a 3, como se observa en la tabla 6.4.

X : Puntuación obtenida en Test Escala psicomotriz de Ozer.

La variable aleatoria X toma valores en el rango de 0 a 18, siendo 18 el valor óptimo para que su edad motriz coincida con su edad biológica, para el caso de niños de 8 años. Para niños de 5 años, se tomará un valor referencial de 12.

- Descriptivas sugeridas

Histograma de frecuencias, gráfico de cajas, entre otros.

Medidas descriptivas de la puntuación tales como: media, desviación estándar.

- Hipótesis de investigación

Este caso corresponde al **parámetro media**, por ejemplo:

La puntuación de la motricidad gruesa media para niños de 5 años es mayor o igual que 12 puntos.

- Hipótesis estadísticas. En este caso prueba unilateral superior.

$$\begin{cases} H_0 : \mu \leq 12 \\ H_1 : \mu > 12 \end{cases}$$

Ejemplo 6-2. Motricidad gruesa, otra perspectiva

El ejemplo anterior consideró la evaluación de la motricidad gruesa mediante la suma de puntuaciones en las 7 actividades. Pero, ahora se toma una

sola actividad y simplemente se evalúa el éxito de dicha actividad, asignando el 0 para el No cumplimiento y el 1 para el cumplimiento relativo o completo.

- Variable: Motricidad gruesa, atributo 6.

De los 6 atributos de la motricidad gruesa, se ha elegido el atributo 6 para transformarlo en variable aleatoria.

- Tipo de variable: Cualitativa
- Nivel de medida: Nominal (Si: 3 puntos, No: 0-2)
- Operacionalización

El atributo elegido se convierte en una variable aleatoria que tiene solo dos posibles resultados: Si (éxito), No (fracaso), que se define de la siguiente forma:

X : número de niños de 5 años que suben y bajan escaleras sin ayuda

- Descriptivas sugeridas

Con la variable que recoge la evaluación de la motricidad gruesa y sus 6 atributos se puede realizar una tabla de frecuencias y gráfico de barras con las 6 opciones, destacando el resultado correspondiente a la actividad elegida.

Pero, con la nueva variable aleatoria definida X se puede realizar un gráfico de barras con el 100 % de las observaciones repartidas entre sí y no del cumplimiento de la actividad. En este último caso, se puede determinar el estadístico de la muestra: proporción de niños que suben y bajan escaleras sin ayuda.

- Hipótesis de investigación

Este caso corresponde al parámetro proporción, por ejemplo:

El 25 % de niños de 5 años suben y bajan escaleras sin ayuda (valor que debe ser justificado para su planteamiento).

- Hipótesis estadísticas

$$\begin{cases} H_0 : p = 0.25 \\ H_1 : p \neq 0.25 \end{cases}$$

6.2.2 Contexto: Autonomía de niños

La autonomía en niños es la capacidad para realizar actividades por sí mismos, tomar decisiones simples y desenvolverse con independencia en su entorno cotidiano [6].

El fomento de la autonomía contribuye al fortalecimiento de la autoestima, la autorregulación y la responsabilidad. En la figura 6.3 se muestra una rutina de autonomía promovida dentro del proceso educativo.

Figura 6.3

Actividades de autonomía



Imagen generada con Open AI chatgpt

La influencia de las rutinas diarias juega un papel importante en el desarrollo de la autonomía en niños de 4 a 5 años. En este contexto, se identifican los diferentes elementos hasta plantear la hipótesis estadística.

- Definición de las Variables y sus roles:

La variable independiente es **Aplicación de rutinas diarias en el aula**: saludar, guardar materiales, lavarse las manos, organizar el espacio, entre otras. La variable dependiente es **Nivel de autonomía de los niños**: tomar decisiones simples, seguir instrucciones, cuidar de sí mismo y desenvolverse con independencia en el entorno escolar. Incluye dimensiones como autocuidado, iniciativa, autorregulación y responsabilidad.

- Tipo de variable y nivel de medida

- Rutinas diarias: Variable categórica nominal (sí / no).

- Autonomía infantil: Variable categórica nominal (sí, no).
- Operacionalización

Tabla 6.3*Operacionalización de variables cualitativas*

Item	Rutinas diarias	Autonomía infantil
Dimensión	Organización pedagógica	Desarrollo de la autonomía
Indicador	1. Saluda al ingresar 2. Guarda sus materiales 3. Se lava las manos 4. Organiza su espacio	1. Se viste solo 2. Sigue instrucciones 3. Participa sin ayuda 4. Cuida sus pertenencias
Unidad	Categórica nominal	Categórica ordinal
Instrumento	Registro docente de rutinas	Lista de cotejo de comportamientos observables

- Descriptivas sugeridas: Tabla de contingencia y gráficos de barras entre Rutinas diarias y Autonomía.
- Hipótesis de investigación sugerida Se estudia cómo la implementación sistemática de rutinas influye en el desarrollo de comportamientos autónomos.
- Hipótesis estadística

$$\begin{cases} H_0 : \text{Las rutinas diarias de los niños no influyen en su autonomía} \\ H_1 : \text{Las rutinas diarias de los niños sí influyen en su autonomía} \end{cases}$$

Observación

Se deja al lector investigar otro tipo de operacionalización de las variables en el que se establezcan conteos de Rutinas diarias y puntuaciones para las actividades de autonomía.

6.3 Variables en la actividad física

6.3.1 Contexto Medidas antropométricas

Las medidas antropométricas son dimensiones del cuerpo humano, se hallan en magnitudes de longitud, perímetro, diámetro y pliegues cutáneos. Las mediciones se realizan bajo protocolos estandarizados como los de la ISAK (International Society for the Advancement of Kinanthropometry) [7].

Estas medidas permiten tomar decisiones para seleccionar deportistas y elegir planes de entrenamiento para el cumplimiento de objetivos.

6.3.1.1 EL PERÍMETRO ABDOMINAL

El perímetro abdominal es la medida de la circunferencia alrededor del abdomen, generalmente tomada a la altura del ombligo o entre la última costilla y la cresta ilíaca. Observe la figura 6.4

Ejemplo 6-3. Perímetro Abdominal en corredores de 100 m

Figura 6.4

Atleta de carrera



Competencia de Atletismo Inter Escuelas 2023, ESMIL-Ecuador)

En atletas de alto rendimiento, el perímetro abdominal es relevante para evaluar el rendimiento. En atletas de velocidad de 100 m, un perímetro abdo-

mental bajo suele asociarse a menor masa grasa y mayor eficiencia biomecánica en la fase de aceleración.

- Definición de la variable : Perímetro abdominal: circunferencia medida en el plano horizontal a nivel del ombligo, con el sujeto de pie, en espiración normal, siguiendo el protocolo ISAK.
- Tipo de variable: Cuantitativa continua. Nivel de medida: razón
- Operacionalización

Tabla 6.4

Operacionalizar

Item	Descripción
Dimensión	Composición corporal
Indicador	Circunferencia abdominal
Unidad	cm
Instrumento	Cinta métrica flexible (precisión 0,1 cm)
Procedimiento	Medir en posición erguida, pies juntos, brazos relajados, cinta paralela al suelo, en espiración normal

- Descriptivas sugeridas y gráficos

Media, desviación estándar, mínimo, máximo. Histogramas y gráfico de cajas (boxplots) para distribución

- Hipótesis de investigación

El perímetro abdominal promedio de los atletas de velocidad de 100 m es igual a 78 cm.

- Hipótesis estadísticas

$$\begin{cases} H_0 : \mu = 78 \\ H_1 : \mu \neq 78 \end{cases}$$

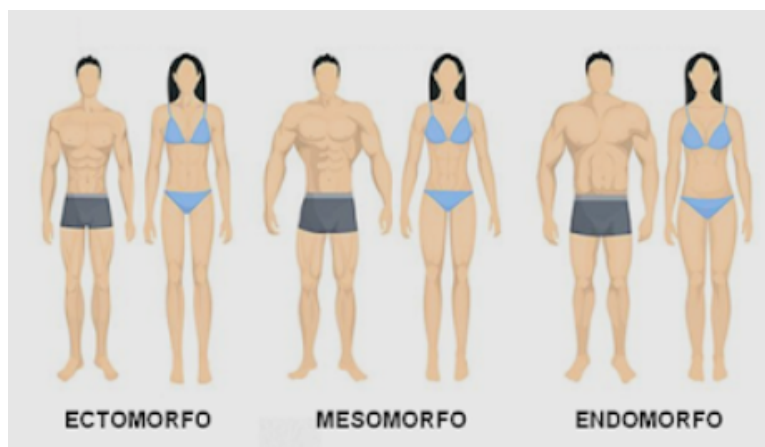
El valor empírico de 78 cm se justifica en estudios realizados con atletas élite de velocidad, en los que se reporta un perímetro abdominal promedio oscilan entre 75 y 80 cm en hombres adultos.

6.3.2 Cotexto: Somatotipo

El Somatotipo es un sistema de clasificación de los cuerpos según su forma de la constitución física de una persona (morfología) en relación a tres componentes: ectomorfia (delgadez), mesomorfia (musculatura) y endomorfia (grasa corporal). Para clasificar los cuerpos se usan las formulas Carter y medidas antropometricas. En la figura 6.5 observe el somatotipo.

Figura 6.5

Somatotipo



<https://www.optifutura.com//>

6.3.2.1 SOMATOTIPO EN CORREDORES DE 100 METROS

En disciplinas de velocidad, la proporción de atletas con predominio ectomorfo-mesomorfo puede influir en la economía de carrera y la potencia relativa, y en distancias más largas predomina el tipo ectomorfo.

Ejemplo 6-4. Somatotipo en Atletas de 100 m

- Definición de la variable: Tipo de cuerpo o somatotipo.
- Tipo de variable: Cualitativa nominal
- Nivel de medida: Nominal (categorías sin orden jerárquico)
- Operacionalización
- Descriptivas sugeridas: Tablas de Frecuencias y gráficos de barras para el tipo de cuerpo.

Tabla 6.5

Operacionalizar

Item	Descripción
Dimensión	Morfología corporal
Indicador	Somatotipo predominante: Ectomorfo / Mesomorfo / Endomorfo
Unidad	Antropometría ISAK + fórmula Heath-Carter
Instrumento	Cinta métrica flexible (precisión 0,1 cm)
Procedimiento	Medición de pliegues, perímetros y diámetros óseos; cálculo de componentes y clasificación según valor más alto

- Hipótesis de investigación

El porcentaje de velocistas de 100 m con tipo de cuerpo ectomorfo es menor que 40 %

- Hipótesis estadísticas

$$\begin{cases} H_0 : p \geq 0.4 \\ H_1 : p < 0.4 \end{cases}$$

El valor empírico se ha tomado de investigaciones en atletas de velocidad que muestran que entre el 38 % y 45 % presentan predominio ectomorfo, lo que se asocia a extremidades largas y bajo peso relativo, favoreciendo la aceleración.

La hipótesis planteada es de tipo unilateral inferior, pero también existe la posibilidad de plantear hipótesis bilaterales; todo depende de la amplitud con que deben cumplirse los resultados. En el contexto anterior, una hipótesis de tipo bilateral a corroborar es:

$$\begin{cases} H_0 : p = 0.4 \\ H_1 : p \neq 0.4 \end{cases}$$

6.3.3 Contexto: Ciclismo de ruta

El ciclismo de ruta consiste en recorrer largas distancias sobre pavimento, requiere el uso de bicicletas diseñadas para velocidad y eficiencia aerodinámica. Esta disciplina se practica en competencias individuales o por equipos, en circuitos urbanos, rurales o montañosos. Su entrenamiento exige resistencia cardiovascular, estrategia y técnica avanzada.

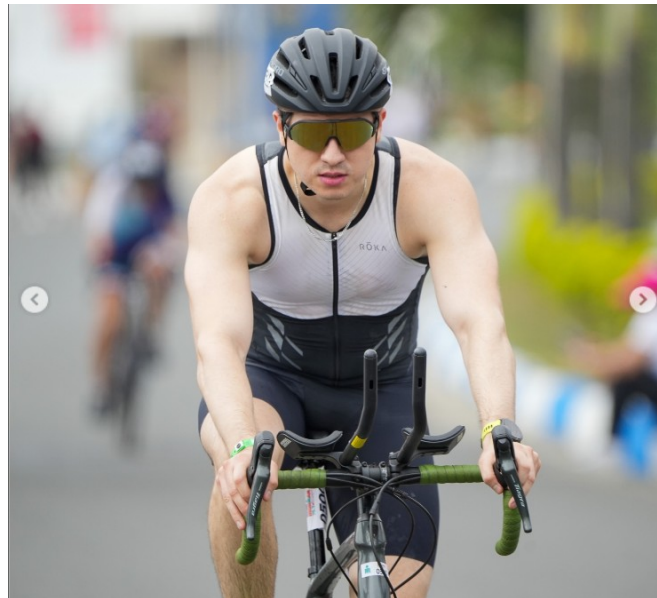
Para el ciclismo de ruta, se presentan algunas variables tanto cuantitativas como cualitativas que pueden ser utilizadas para evaluar el rendimiento del ciclista:

- Variables cuantitativas: Frecuencia cardíaca, Relación peso/potencia (W/kg), VO_2 máximo (mlg/dL), Umbral de lactato (LT o OBLA), Tiempo total de carrera (minutos).
- Variables cualitativas: Técnica de alimentación e hidratación, Marca y configuración de la bicicleta, Estilo de entrenamiento previo, Tipo de terreno predominante. Lesiones por mala hidratación.

En la Figura 6.6 se observa un ciclista en competencia con indumentaria adecuada.

Figura 6.6

Ciclista de ruta



Iroman 70.3 Manta

6.3.3.1 CORRELACIÓN ENTRE VARIABLES CUANTITATIVAS EN EL CICLISMO DE RUTA

El VO_2 y el umbral de lactato reflejan la capacidad aeróbica y la intensidad sostenible, mientras que la relación peso/potencia determina la eficiencia en ascensos. La frecuencia cardíaca se utiliza para monitorear la carga interna y el tiempo de carrera permite modelar el impacto de estas variables sobre el desempeño competitivo reflejado finalmente en el tiempo en la carrera (minutos). En la figura 6.7 se observa un ciclista en transición.

Figura 6.7

Transición en competencia de Iroman



Iroman 70.3 Manta

Ejemplo 6-5. Tiempo en la carrera

Este ejemplo difiere de los temas expuestos anteriormente, por cuanto, además de la variable tiempo, se toman otras variables cuantitativas que influyen sobre el tiempo en la carrera.

- Las variables tipo y nivel de medida. Se han elegido las 5 variables cuantitativas de la sección anterior, cuyo nivel de medida es escala o razón. Se deja al lector investigar la definición de cada variable.
- Operacionalización de las variables independientes y dependiente. Para este ejemplo, se elige una variable independiente: Relación peso/poten-

cia, y como variable dependiente: Tiempo en la carrera. La herramienta estadística apropiada sería un análisis de regresión simple, pero si se toma más de una variable independiente, el estudio sería de regresión múltiple. Observe la Tabla 6.6

Tabla 6.6

Operacionalización

Item	Tiempo carrera	Peso/potencia
Dimensión	Rendimiento	Fisiológica
Indicador	Tiempo final	Mecánica
Unidad	[Minutos]	[W/kg]
Instrumento	Cronómetro oficial o chip de carrera	Potenciómetro + báscula digital

■ Descriptivas sugeridas

- Tablas de frecuencias según valores estandarizados.
- Histogramas y boxplots para la distribución.
- Medidas descriptivas.

■ Hipótesis de investigación.

La relación entre el peso / potencia influye sobre el tiempo ejecutado en la carrera.

■ Hipótesis estadísticas

Para comprobar esta relación se utiliza la prueba estadística sobre el coeficiente de correlación ρ .

$$\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{cases}$$

Se deja al lector profundizar en este tema, que no forma parte de este libro.

6.3.4 Contexto: Lesiones deportivas

Una lesión deportiva es un daño físico que se presenta por la práctica de actividad física o deporte, afectando músculos, ligamentos, tendones, huesos o estructuras nerviosas [8].

Clasificación de las lesiones:

- Según el momento de aparición
- Según el tejido afectado
- Según la gravedad
- Según el mecanismo de lesión

6.3.4.1 FACTORES ASOCIADOS A LAS LESIONES

Además de la práctica deportiva, existen otros factores que determinan las lesiones deportivas, tales como:

- Factores individuales
- Factores técnicos y deportivos. Ejemplo en la figura 6.8
- Factores psicológicos
- Factores biomecánicos y anatómicos.

Figura 6.8

Futbolista lesionado



<https://saquedemeta.co//>

Ejemplo 6-6. Asociación ente el género y la gravedad de la lesión

- Según la gravedad de la lesión, estas se clasifican en: Leve, Moderada y Grave.

- Entre los factores individuales que pueden influir en una lesión están la edad, el sexo, la condición física, el historial de lesiones, la fatiga y el somatotipo.
- Existen estudios que afirman que las lesiones pueden presentarse en mayor proporción entre un género y otro, como consecuencia de las diferencias entre sus características anatómicas [9].

Este caso no corresponde a una prueba sobre medidas descriptivas; más bien, el interés recae en realizar una prueba explicativa.

- Variables y operacionalización en la Tabla 6.7.

Tabla 6.7

Operacionalización de variables cualitativas

Item	Sexo	Tipo de lesión
Dimensión	Biológica	Clínica
Indicador	1. Hombre	1. Leve
	2. Mujer	2. Moderada
		3. Aguda
Unidad	Categórica nominal	Categórica ordinal
Instrumento	Ficha personal o encuesta	Diagnóstico Médico de la gravedad

- Descriptivas sugeridas
 - Tablas de contingencia entre sexo y gravedad de la lesión.
 - Gráficos entre Sexo y gravedad de la lesión con porcentajes respecto a las filas, tomando filas como sexo, para establecer comparaciones.
- Hipótesis estadísticas

$$\begin{cases} H_0 : \text{El sexo no influye en la gravedad de la lesión} \\ H_1 : \text{El sexo influye en la gravedad de la lesión} \end{cases}$$

6.4 Ejercicios del capítulo

En los siguientes ejercicios propuestos, debe realizar las actividades de acuerdo con el esquema desarrollado en el capítulo, es decir:

- Tipo de variable y nivel de medida.
- Descriptivas sugeridas.
- Operacionalización de las variables. Esta no se realiza cuando el enunciado del ejercicio ya la proporciona.
- Hipótesis de investigación sugerida.
- Hipótesis estadísticas.

Datos para los ejercicios del 6.1 al 6.3: Lista de rutinas y de comportamientos autónomos

Rutinas diarias: Actividades inducidas u organizadas por el adulto.

- Saludar al ingresar al aula
- Colgar la mochila en su lugar
- Lavarse las manos antes de comer
- Guardar los materiales después de usarlos
- Formarse para salir al recreo
- Ordenar los juguetes al finalizar la actividad
- Colaborar en la limpieza del espacio
- Despedirse al finalizar la jornada

Comportamientos autónomos: El niño actúa por sí mismo, sin depender de instrucciones constantes.

- Se viste solo (chaqueta, zapatos)
- Guarda sus pertenencias sin que se lo pidan
- Sigue instrucciones simples sin supervisión constante

- Elige materiales para trabajar según la consigna
- Cuida sus objetos personales (mochila, fiambarrera, útiles)
- Solicita ayuda cuando la necesita (sin esperar que lo noten)
- Reconoce cuándo debe ir al baño y lo hace por sí mismo
- Toma decisiones simples (elige con quién jugar, qué actividad hacer)
- Regula su conducta en situaciones sociales (espera su turno, respeta normas)

6.1. Utilice la variable Autonomía y establezca una escala de medición para valorar los resultados de forma que obtenga una variable cuantitativa, de modo que pueda plantear una hipótesis sobre la valoración promedio.

6.2. Utilice la variable rutinas y, asumiendo que se valora con el cumplimiento (sí, no), elija una de las rutinas para plantear una hipótesis sobre la proporción de niños que cumplen cierta rutina.

6.3. Utilice la rutina: Saludar al ingresar al aula (sí, no) y la actividad autónoma Regula su conducta en situaciones sociales (espera su turno, respeta normas) (sí, no) para plantear una hipótesis que relacione las dos variables.

Datos para los ejercicios del 6.4 al 6.7 en la Tabla 6.8

6.4. Con la variable de frecuencia cardíaca, plantear una hipótesis sobre la frecuencia cardíaca media.

6.5. Con la variable umbral de lactato, plantear una hipótesis sobre el umbral de lactato medio.

6.6. Utilice las variables VO_2 máx (independiente) y tiempo en la carrera (dependiente) que se encuentran en la Tabla 6.8 y en la Tabla 6.6 para plantear una hipótesis de correlación entre las variables.

6.7. Plantear una hipótesis sobre el umbral de lactato en deportistas de carreras de fondo.

Tabla 6.8

Variables de interés en el desempeño deportivo

Variable	Dimensión	Indicador / Unidad	Instrumento
Frecuencia cardíaca	Fisiológica	Latidos por minuto promedio [lpm]	Pulsómetro o banda cardíaca
Relación peso/potencia	Fisiológica Mecánica	Potencia media dividida por peso corporal [W/kg]	Potenciómetro + báscula digital
VO_2 máximo	Aeróbica	Consumo máximo de oxígeno [ml/kg/min]	Cicloergómetro + analizador de gases
Umbral de lactato (LT/OBLA)	Metabólica	Concentración de lactato en sangre [mmol/L]	Analizador portátil de lactato

6.8. En el deporte del voleibol, utilice la variable puntos por partido para plantear una hipótesis sobre la puntuación promedio.

6.9. En el deporte del voleibol, utilice la variable Porcentaje de aciertos en saques para plantear una hipótesis sobre el porcentaje de aciertos promedio.

6.10. En el deporte del voleibol, utilice la variable Número de Bloqueos exitosos para plantear una hipótesis sobre la proporción de bloqueos exitosos por partido.

6.11. Una de las pruebas físicas de ingreso a una Escuela Militar es realizar un cierto número de abdominales por minuto. Un resultado mínimo para cumplir con la prueba es realizar 50 abdominales por minuto. Utilizando la variable número de abdominales por minuto y la información indicada, plantea una hipótesis adecuada sobre el número promedio de abdominales por minuto.

6.12. Tomando en cuenta la información del ejercicio anterior, tome como variable X : número de aspirantes que cumplen con la prueba de abdominales (aquellos que realizan 50 o más abdominales), se plantea una hipótesis adecuada sobre la proporción de aspirantes que superan dicha prueba.

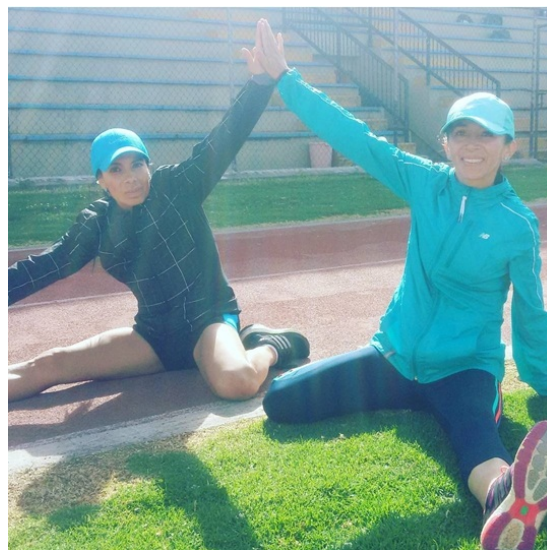
6.13. Tomando como referencia que una de las pruebas militares de control es subir el cabo, plantee una hipótesis sobre la proporción de miembros de la fuerza terrestre que cumplen con esta prueba.

6.14. En carreras pedestres, los competidores suelen realizar una rutina de estiramientos al finalizar la carrera. En la figura 6.8 se observa un ejercicio de estiramiento. Plantear una hipótesis sobre la relación entre la lesión deportiva (sí, no) y el estiramiento después de una carrera (sí, no).

En las carreras pedestres, los competidores suelen realizar rutinas de estiramiento al finalizar la prueba, como por ejemplo, en la figura 6.9 se observa el estiramiento de aductores y flexores de cadera en posición sentada. Se propone plantear una hipótesis acerca de la relación entre la presencia de lesión deportiva (sí/no) y la práctica de estiramiento posterior a la carrera (sí/no).

Figura 6.9

Estiramiento de aductores y flexores





Las ciencias aplicadas no existen; sólo las aplicaciones de la ciencia.
Louis Pasteur

Fuente: Elaboración propia mediante IA (Canva)

Apéndices

Apéndice A. Datos

A.1 Datos de pesos

Peso_lb	Sexo	NivelActFisica	Peso_lb	Sexo	NivelActFisica
140	Hombre	Alto	149	Hombre	Medio
153	Hombre	Alto	151	Hombre	Medio
153	Hombre	Bajo	151	Hombre	Medio
154	Hombre	Bajo	151	Hombre	Medio
154	Hombre	Bajo	152	Hombre	Medio
155	Hombre	Bajo	153	Hombre	Medio
155	Hombre	Bajo	158	Hombre	Medio
155	Hombre	Bajo	160	Hombre	Medio
156	Hombre	Bajo	163	Hombre	Medio
156	Hombre	Bajo	166	Hombre	Medio
156	Hombre	Bajo	180	Hombre	Medio
156	Hombre	Bajo	181	Hombre	Medio
157	Hombre	Bajo	186	Hombre	Medio
158	Hombre	Bajo	192	Hombre	Medio
159	Hombre	Bajo	201	Hombre	Medio
159	Hombre	Bajo	206	Hombre	Medio
160	Hombre	Bajo	208	Hombre	Medio
161	Hombre	Bajo	210	Hombre	Medio
162	Hombre	Bajo	211	Hombre	Medio
164	Hombre	Bajo	135	Mujer	Alto
168	Hombre	Bajo	136	Mujer	Alto
170	Hombre	Bajo	138	Mujer	Alto
172	Hombre	Bajo	140	Mujer	Alto
175	Hombre	Bajo	144	Mujer	Alto
176	Hombre	Bajo	146	Mujer	Alto
178	Hombre	Bajo	148	Mujer	Alto
178	Hombre	Bajo	149	Mujer	Alto
180	Hombre	Bajo	133	Mujer	Bajo
181	Hombre	Bajo	155	Mujer	Bajo
185	Hombre	Bajo	158	Mujer	Bajo
187	Hombre	Bajo	133	Mujer	Medio
189	Hombre	Bajo	141	Mujer	Medio
190	Hombre	Bajo	142	Mujer	Medio
191	Hombre	Bajo	143	Mujer	Medio
192	Hombre	Bajo	144	Mujer	Medio
192	Hombre	Bajo	144	Mujer	Medio
199	Hombre	Bajo	146	Mujer	Medio
200	Hombre	Bajo	147	Mujer	Medio
204	Hombre	Bajo	148	Mujer	Medio
205	Hombre	Bajo	148	Mujer	Medio
137	Hombre	Medio	149	Mujer	Medio
143	Hombre	Medio	150	Mujer	Medio
143	Hombre	Medio	153	Mujer	Medio
148	Hombre	Medio	166	Mujer	Medio
149	Hombre	Medio	167	Mujer	Medio

<https://drive.google.com/drive/folders/1Usn4dEzvvLrggNY3c18JgD0pLlxXiIXU?usp=sharing>

A.2 Corpus del habla de niños bilingües

Descripción de las variables

- X1: Primer idioma
- X2: Genero
- X3: Edad de llegada a Canada (meses)
- X4: Edad de exposición al inglés (meses)
- X5: Edad al momento de la grabación (meses)
- Y: Meses de inglés
- A10: Periodo alrededor de 10 meses

ID	X1	X2	X3	X4	X5	Y	A10
1	Farsi	F	58	63	73	10	Hasta 10 meses
2	Spanish	M	53	56	66	10	Hasta 10 meses
3	Romanian	F	69	69	74	5	Hasta 10 meses
4	Mandarin	F	73	73	81	8	Hasta 10 meses
5	Cantonese	M	1	48	62	14	Más 10 meses
6	Mandarin	M	53	55	64	9	Hasta 10 meses
7	Mandarin	M	42	47	54	7	Hasta 10 meses
8	Spanish	M	67	67	75	8	Hasta 10 meses
9	Spanish	M	45	58	68	10	Hasta 10 meses
10	Korean	F	48	60	62	2	Hasta 10 meses
11	Mandarin	F	53	53	71	18	Más 10 meses
12	Ukrainian	M	66	66	79	13	Más 10 meses
13	Arabic	F	1	47	58	11	Más 10 meses
14	Mandarin	F	56	56	60	4	Hasta 10 meses
15	Japanese	F	26	42	51	9	Hasta 10 meses
16	Spanish	M	88	89	94	5	Hasta 10 meses
17	Cantonese	M	1	40	56	16	Más 10 meses
18	Spanish	M	45	46	61	15	Más 10 meses
19	Farsi	M	64	66	78	12	Más 10 meses
20	Spanish	M	60	60	66	6	Hasta 10 meses
21	Farsi	M	37	39	50	11	Más 10 meses
22	Mandarin	M	68	70	77	7	Hasta 10 meses
23	Arabic	M	1	42	50	8	Hasta 10 meses
24	Arabic	M	1	57	59	9	Hasta 10 meses

www.kaggle.com/datasets/ratatman/corpus-of-bilingual-childrens-speech

Apéndice B. Tablas de probabilidad

B.1 Distribución normal

z	-0.09	-0.08	-0.07	-0.06	-0.05	-0.04	-0.03	-0.02	-0.01	0
-3.6	0.00011	0.00012	0.00012	0.00013	0.00013	0.00014	0.00014	0.00015	0.00015	0.00016
-3.5	0.00017	0.00017	0.00018	0.00019	0.00019	0.00020	0.00021	0.00022	0.00022	0.00023
-3.4	0.00024	0.00025	0.00026	0.00027	0.00028	0.00029	0.00030	0.00031	0.00032	0.00034
-3.3	0.00035	0.00036	0.00038	0.00039	0.00040	0.00042	0.00043	0.00045	0.00047	0.00048
-3.2	0.00050	0.00052	0.00054	0.00056	0.00058	0.00060	0.00062	0.00064	0.00066	0.00069
-3.1	0.00071	0.00074	0.00076	0.00079	0.00082	0.00084	0.00087	0.00090	0.00094	0.00097
-3.0	0.00100	0.00104	0.00107	0.00111	0.00114	0.00118	0.00122	0.00126	0.00131	0.00135
-2.9	0.00139	0.00144	0.00149	0.00154	0.00159	0.00164	0.00169	0.00175	0.00181	0.00187
-2.8	0.00193	0.00199	0.00205	0.00212	0.00219	0.00226	0.00233	0.00240	0.00248	0.00256
-2.7	0.00264	0.00272	0.00280	0.00289	0.00298	0.00307	0.00317	0.00326	0.00336	0.00347
-2.6	0.00357	0.00368	0.00379	0.00391	0.00402	0.00415	0.00427	0.00440	0.00453	0.00466
-2.5	0.00480	0.00494	0.00508	0.00523	0.00539	0.00554	0.00570	0.00587	0.00604	0.00621
-2.4	0.00639	0.00657	0.00676	0.00695	0.00714	0.00734	0.00755	0.00776	0.00798	0.00820
-2.3	0.00842	0.00866	0.00889	0.00914	0.00939	0.00964	0.00990	0.01017	0.01044	0.01072
-2.2	0.01101	0.01130	0.01160	0.01191	0.01222	0.01255	0.01287	0.01321	0.01355	0.01390
-2.1	0.01426	0.01463	0.01500	0.01539	0.01578	0.01618	0.01659	0.01700	0.01743	0.01786
-2.0	0.01831	0.01876	0.01923	0.01970	0.02018	0.02068	0.02118	0.02169	0.02222	0.02275
-1.9	0.02330	0.02385	0.02442	0.02500	0.02559	0.02619	0.02680	0.02743	0.02807	0.02872
-1.8	0.02938	0.03005	0.03074	0.03144	0.03216	0.03288	0.03362	0.03438	0.03515	0.03593
-1.7	0.03673	0.03754	0.03836	0.03920	0.04006	0.04093	0.04182	0.04272	0.04363	0.04457
-1.6	0.04551	0.04648	0.04746	0.04846	0.04947	0.05050	0.05155	0.05262	0.05370	0.05480
-1.5	0.05592	0.05705	0.05821	0.05938	0.06057	0.06178	0.06301	0.06426	0.06552	0.06681
-1.4	0.06811	0.06944	0.07078	0.07215	0.07353	0.07493	0.07636	0.07780	0.07927	0.08076
-1.3	0.08226	0.08379	0.08534	0.08691	0.08851	0.09012	0.09176	0.09342	0.09510	0.09680
-1.2	0.09853	0.10027	0.10204	0.10383	0.10565	0.10749	0.10935	0.11123	0.11314	0.11507
-1.1	0.11702	0.11900	0.12100	0.12302	0.12507	0.12714	0.12924	0.13136	0.13350	0.13567
-1.0	0.13786	0.14007	0.14231	0.14457	0.14686	0.14917	0.15151	0.15386	0.15625	0.15866
-0.9	0.16109	0.16354	0.16602	0.16853	0.17106	0.17361	0.17619	0.17879	0.18141	0.18406
-0.8	0.18673	0.18943	0.19215	0.19489	0.19766	0.20045	0.20327	0.20611	0.20897	0.21186
-0.7	0.21476	0.21770	0.22065	0.22363	0.22663	0.22965	0.23270	0.23576	0.23885	0.24196
-0.6	0.24510	0.24825	0.25143	0.25463	0.25785	0.26109	0.26435	0.26763	0.27093	0.27425
-0.5	0.27760	0.28096	0.28434	0.28774	0.29116	0.29460	0.29806	0.30153	0.30503	0.30854
-0.4	0.31207	0.31561	0.31918	0.32276	0.32636	0.32997	0.33360	0.33724	0.34090	0.34458
-0.3	0.34827	0.35197	0.35569	0.35942	0.36317	0.36693	0.37070	0.37448	0.37828	0.38209
-0.2	0.38591	0.38974	0.39358	0.39743	0.40129	0.40517	0.40905	0.41294	0.41683	0.42074
-0.1	0.42465	0.42858	0.43251	0.43644	0.44038	0.44433	0.44828	0.45224	0.45620	0.46017
0.0	0.46414	0.46812	0.47210	0.47608	0.48006	0.48405	0.48803	0.49202	0.49601	0.50000

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989

B.2 Distribución t-Student

- Esta tabla contiene los puntos críticos correspondientes a una probabilidad en cola izquierda.
- Se deben aplicar las propiedades para obtener probabilidades en cola izquierda.
- Las probabilidades para valores de t que no están en la tabla, pero que están dentro del rango, solo quedaran expresadas en una relación de orden.

t_0 tal que: $p(T_{gl} < t_0) = \alpha$

g.l. = Grados de libertad



g.l	0.0005	0.001	0.003	0.005	0.01	0.015	0.02	0.025	0.04	0.05	0.1	0.25	0.4	0.5
1	-636.619	-318.309	-127.321	-63.657	-31.821	-21.205	-15.895	-12.706	-7.916	-6.314	-3.078	-1.000	-0.325	0
2	-31.599	-22.327	-14.089	-9.925	-6.965	-5.643	-4.849	-4.303	-3.320	-2.920	-1.886	-0.816	-0.289	0
3	-12.924	-10.215	-7.453	-5.841	-4.541	-3.896	-3.482	-3.182	-2.605	-2.353	-1.638	-0.765	-0.277	0
4	-8.610	-7.173	-5.598	-4.604	-3.747	-3.298	-2.999	-2.776	-2.333	-2.132	-1.533	-0.741	-0.271	0
5	-6.869	-5.893	-4.773	-4.032	-3.365	-3.003	-2.757	-2.571	-2.191	-2.015	-1.476	-0.727	-0.267	0
6	-5.959	-5.208	-4.317	-3.707	-3.143	-2.829	-2.612	-2.447	-2.104	-1.943	-1.440	-0.718	-0.265	0
7	-5.408	-4.785	-4.029	-3.499	-2.998	-2.715	-2.517	-2.365	-2.046	-1.895	-1.415	-0.711	-0.263	0
8	-5.041	-4.501	-3.833	-3.355	-2.896	-2.634	-2.449	-2.306	-2.004	-1.860	-1.397	-0.706	-0.262	0
9	-4.781	-4.297	-3.690	-3.250	-2.821	-2.574	-2.398	-2.262	-1.973	-1.833	-1.383	-0.703	-0.261	0
10	-4.587	-4.144	-3.581	-3.169	-2.764	-2.527	-2.359	-2.228	-1.948	-1.812	-1.372	-0.700	-0.260	0
11	-4.437	-4.025	-3.497	-3.106	-2.718	-2.491	-2.328	-2.201	-1.928	-1.796	-1.363	-0.697	-0.260	0
12	-4.318	-3.930	-3.428	-3.055	-2.681	-2.461	-2.303	-2.179	-1.912	-1.782	-1.356	-0.695	-0.259	0
13	-4.221	-3.852	-3.372	-3.012	-2.650	-2.436	-2.282	-2.160	-1.899	-1.771	-1.350	-0.694	-0.259	0
14	-4.140	-3.787	-3.326	-2.977	-2.624	-2.415	-2.264	-2.145	-1.887	-1.761	-1.345	-0.692	-0.258	0
15	-4.073	-3.733	-3.286	-2.947	-2.602	-2.397	-2.249	-2.131	-1.878	-1.753	-1.341	-0.691	-0.258	0
16	-4.015	-3.686	-3.252	-2.921	-2.583	-2.382	-2.235	-2.120	-1.869	-1.746	-1.337	-0.690	-0.258	0
17	-3.965	-3.646	-3.222	-2.898	-2.567	-2.368	-2.224	-2.110	-1.862	-1.740	-1.333	-0.689	-0.257	0
18	-3.922	-3.610	-3.197	-2.878	-2.552	-2.356	-2.214	-2.101	-1.855	-1.734	-1.330	-0.688	-0.257	0
19	-3.883	-3.579	-3.174	-2.861	-2.539	-2.346	-2.205	-2.093	-1.850	-1.729	-1.328	-0.688	-0.257	0
20	-3.850	-3.552	-3.153	-2.845	-2.528	-2.336	-2.197	-2.086	-1.844	-1.725	-1.325	-0.687	-0.257	0
21	-3.819	-3.527	-3.135	-2.831	-2.518	-2.328	-2.189	-2.080	-1.840	-1.721	-1.323	-0.686	-0.257	0
22	-3.792	-3.505	-3.119	-2.819	-2.508	-2.320	-2.183	-2.074	-1.835	-1.717	-1.321	-0.686	-0.256	0
23	-3.768	-3.485	-3.104	-2.807	-2.500	-2.313	-2.177	-2.069	-1.832	-1.714	-1.319	-0.685	-0.256	0
24	-3.745	-3.467	-3.091	-2.797	-2.492	-2.307	-2.172	-2.064	-1.828	-1.711	-1.318	-0.685	-0.256	0
25	-3.725	-3.450	-3.078	-2.787	-2.485	-2.301	-2.167	-2.060	-1.825	-1.708	-1.316	-0.684	-0.256	0
26	-3.707	-3.435	-3.067	-2.779	-2.479	-2.296	-2.162	-2.056	-1.822	-1.706	-1.315	-0.684	-0.256	0
27	-3.690	-3.421	-3.057	-2.771	-2.473	-2.291	-2.158	-2.052	-1.819	-1.703	-1.314	-0.684	-0.256	0
28	-3.674	-3.408	-3.047	-2.763	-2.467	-2.286	-2.154	-2.048	-1.817	-1.701	-1.313	-0.683	-0.256	0
29	-3.659	-3.396	-3.038	-2.756	-2.462	-2.282	-2.150	-2.045	-1.814	-1.699	-1.311	-0.683	-0.256	0
30	-3.646	-3.385	-3.030	-2.750	-2.457	-2.278	-2.147	-2.042	-1.812	-1.697	-1.310	-0.683	-0.256	0
31	-3.633	-3.375	-3.022	-2.744	-2.453	-2.275	-2.144	-2.040	-1.810	-1.696	-1.309	-0.682	-0.256	0
32	-3.622	-3.365	-3.015	-2.738	-2.449	-2.271	-2.141	-2.037	-1.808	-1.694	-1.309	-0.682	-0.255	0
33	-3.611	-3.356	-3.008	-2.733	-2.445	-2.268	-2.138	-2.035	-1.806	-1.692	-1.308	-0.682	-0.255	0
34	-3.601	-3.348	-3.002	-2.728	-2.441	-2.265	-2.136	-2.032	-1.805	-1.691	-1.307	-0.682	-0.255	0
35	-3.591	-3.340	-2.996	-2.724	-2.438	-2.262	-2.133	-2.030	-1.803	-1.690	-1.306	-0.682	-0.255	0
40	-3.551	-3.307	-2.971	-2.704	-2.423	-2.250	-2.123	-2.021	-1.796	-1.684	-1.303	-0.681	-0.255	0
100	-3.390	-3.174	-2.871	-2.626	-2.364	-2.201	-2.081	-1.984	-1.769	-1.660	-1.290	-0.677	-0.254	0
500	-3.310	-3.107	-2.820	-2.586	-2.334	-2.176	-2.059	-1.965	-1.754	-1.648	-1.283	-0.675	-0.253	0
1000	-3.300	-3.098	-2.813	-2.581	-2.330	-2.173	-2.056	-1.962	-1.752	-1.646	-1.282	-0.675	-0.253	0



		α													
g.l		0.5	0.6	0.75	0.9	1	1	0.98	0.98	0.99	0.99	0.995	0.998	0.999	0.9995
1	0	0.325	1.000	3.078	6.314	7.916	12.706	15.895	21.205	31.821	63.657	127.321	318.309	636.619	
2	0	0.289	0.816	1.886	2.920	3.320	4.303	4.849	5.643	6.965	9.925	14.089	22.327	31.599	
3	0	0.277	0.765	1.638	2.353	2.605	3.182	3.482	3.896	4.541	5.841	7.453	10.215	12.924	
4	0	0.271	0.741	1.533	2.132	2.333	2.776	2.999	3.298	3.747	4.604	5.598	7.173	8.610	
5	0	0.267	0.727	1.476	2.015	2.191	2.571	2.757	3.003	3.365	4.032	4.773	5.893	6.869	
6	0	0.265	0.718	1.440	1.943	2.104	2.447	2.612	2.829	3.143	3.707	4.317	5.208	5.959	
7	0	0.263	0.711	1.415	1.895	2.046	2.365	2.517	2.715	2.998	3.499	4.029	4.785	5.408	
8	0	0.262	0.706	1.397	1.860	2.004	2.306	2.449	2.634	2.896	3.355	3.833	4.501	5.041	
9	0	0.261	0.703	1.383	1.833	1.973	2.262	2.398	2.574	2.821	3.250	3.690	4.297	4.781	
10	0	0.260	0.700	1.372	1.812	1.948	2.228	2.359	2.527	2.764	3.169	3.581	4.144	4.587	
11	0	0.260	0.697	1.363	1.796	1.928	2.201	2.328	2.491	2.718	3.106	3.497	4.025	4.437	
12	0	0.259	0.695	1.356	1.782	1.912	2.179	2.303	2.461	2.681	3.055	3.428	3.930	4.318	
13	0	0.259	0.694	1.350	1.771	1.899	2.160	2.282	2.436	2.650	3.012	3.372	3.852	4.221	
14	0	0.258	0.692	1.345	1.761	1.887	2.145	2.264	2.415	2.624	2.977	3.326	3.787	4.140	
15	0	0.258	0.691	1.341	1.753	1.878	2.131	2.249	2.397	2.602	2.947	3.286	3.733	4.073	
16	0	0.258	0.690	1.337	1.746	1.869	2.120	2.235	2.382	2.583	2.921	3.252	3.686	4.015	
17	0	0.257	0.689	1.333	1.740	1.862	2.110	2.224	2.368	2.567	2.898	3.222	3.646	3.965	
18	0	0.257	0.688	1.330	1.734	1.855	2.101	2.214	2.356	2.552	2.878	3.197	3.610	3.922	
19	0	0.257	0.688	1.328	1.729	1.850	2.093	2.205	2.346	2.539	2.861	3.174	3.579	3.883	
20	0	0.257	0.687	1.325	1.725	1.844	2.086	2.197	2.336	2.528	2.845	3.153	3.552	3.850	
21	0	0.257	0.686	1.323	1.721	1.840	2.080	2.189	2.328	2.518	2.831	3.135	3.527	3.819	
22	0	0.256	0.686	1.321	1.717	1.835	2.074	2.183	2.320	2.508	2.819	3.119	3.505	3.792	
23	0	0.256	0.685	1.319	1.714	1.832	2.069	2.177	2.313	2.500	2.807	3.104	3.485	3.768	
24	0	0.256	0.685	1.318	1.711	1.828	2.064	2.172	2.307	2.492	2.797	3.091	3.467	3.745	
25	0	0.256	0.684	1.316	1.708	1.825	2.060	2.167	2.301	2.485	2.787	3.078	3.450	3.725	
26	0	0.256	0.684	1.315	1.706	1.822	2.056	2.162	2.296	2.479	2.779	3.067	3.435	3.707	
27	0	0.256	0.684	1.314	1.703	1.819	2.052	2.158	2.291	2.473	2.771	3.057	3.421	3.690	
28	0	0.256	0.683	1.313	1.701	1.817	2.048	2.154	2.286	2.467	2.763	3.047	3.408	3.674	
29	0	0.256	0.683	1.311	1.699	1.814	2.045	2.150	2.282	2.462	2.756	3.038	3.396	3.659	
30	0	0.256	0.683	1.310	1.697	1.812	2.042	2.147	2.278	2.457	2.750	3.030	3.385	3.646	
31	0	0.256	0.682	1.309	1.696	1.810	2.040	2.144	2.275	2.453	2.744	3.022	3.375	3.633	
32	0	0.255	0.682	1.309	1.694	1.808	2.037	2.141	2.271	2.449	2.738	3.015	3.365	3.622	
33	0	0.255	0.682	1.308	1.692	1.806	2.035	2.138	2.268	2.445	2.733	3.008	3.356	3.611	
34	0	0.255	0.682	1.307	1.691	1.805	2.032	2.136	2.265	2.441	2.728	3.002	3.348	3.601	
35	0	0.255	0.682	1.306	1.690	1.803	2.030	2.133	2.262	2.438	2.724	2.996	3.340	3.591	
40	0	0.255	0.681	1.303	1.684	1.796	2.021	2.123	2.250	2.423	2.704	2.971	3.307	3.551	
100	0	0.254	0.677	1.290	1.660	1.763	1.984	2.081	2.201	2.364	2.626	2.871	3.174	3.390	
500	0	0.253	0.675	1.283	1.648	1.754	1.965	2.059	2.176	2.334	2.586	2.820	3.107	3.310	
1000	0	0.253	0.675	1.282	1.646	1.752	1.962	2.056	2.173	2.330	2.581	2.813	3.098	3.300	

Apéndice C.

Respuestas de ejercicios seleccionados

Capítulo 1

Los estadísticos de posición pueden variar dependiendo del algoritmo usado en cada software estadístico.

1.3.

- a. Media: 15.9474, Mediana: 16, Moda: 16
- b. Rango: 4, Varianza: 1.9415, Desviación estándar: 1.3934, Cv: 0.0873

1.5.

- a. Media: 102.21, Mediana: 99.5, Moda: 99
- b. Rango: 47, Varianza: 162.43, Desviación estándar: 12.74, cv: 0.125

1.9.

- a. Q1: 93, Mediana (Q2): 99, Q3: 115
- b. Porcentaje mayor a 100: 45.83
- c. Porcentaje mayor o igual que 100: 50

1.9.

- a. Frecuencias: Diestros = 270, Zurdos = 30
- b. Porcentajes: 90

1.11.

- a. Media: 22.38, Mediana: 22.25, Moda: no hay
- b. Rango: 3.8, Varianza: 1.46, Desviación estándar: 1.21, cv: 0.054

1.13.

- a. Grupo A → Media: 1.90, Mediana: 1.90, Moda: 1.88
- b. Grupo B → Media: 1.90, Mediana: 1.92, Moda: 1.92
- c. Comparación: Coinciden en el valor de la media, pero el grupo B tiene la mediana y la moda con mayor valor.

Capítulo 2

2.1

- a. $13/52$
- b. $13/52$

2.3

- a. $12/51$
- b. $13/51$

2.5

- a. 0.62
- b. 0.53

2.7

- a. 0.40625
- b. 0.59375

2.9

- a. 0.07
- b. 0.41

2.11

- a. 0.696
- b. 0.226

2.13

- a. 0.3125
- b. 0.5200

2.15

- a. 0.3360
- b. 0.1575

2.16

- a. 0.6735
- b. 0.2339

Capítulo 3

3.3

- a. Binomial, $n=24$, $p=0.70$
- b. 0.01989
- c. 0.03139
- d. 0.43169

3.4

- a. Hipergeométrica, $N=40$, $K=18$, $n=10$
- b. 0.215
- c. 0.957
- d. 0.847

3.5

- a. 0.0907
- b. 0.5967
- c. 4.8
- d. 0.0948

3.7

- a. 14
- b. 4.2
- c. 2.094

3.10

- a. 2.4
- b. 2.4
- c. 1.5492

3.11

- a. 0.5
- b. 0.2514
- c. 0.4971
- d. 0.1587

3.13

a. 0.3932

b. 0.9011

3.15

a. 0.3936

b. 0.4067

3.17

a. 0.1008

b. 0.1992

3.19

a. 0.0918

b. 0.8176

Capítulo 4

4.1 (5.66, 6.74) minutos

4.2 (53.16, 58.84) palabras

4.3 (0.59, 0.85)

4.5 (7.59, 8.61) accesos

4.6 (80.78, 85.62) puntos

4.7 (0.62, 0.76)

4.9 (0.63, 0.81)

Capítulo 5

5.1 Se rechaza H_0 , existe evidencia de diferencia

5.2 Se rechaza H_0 , usan más de 50 palabras

5.3 No se rechaza H_0 , no hay evidencia de un valor superior al 60 %

5.4 No se rechaza H_0 , se cumple al menos 80 %

5.5 No se rechaza H_0 , no superan 10 accesos

5.6 Se rechaza H_0 , es inferior a 85 puntos

5.7 Se rechaza H_0 , más del 65 % logran valores menores que 4 horas

5.8 Se rechaza H_0 , menos del 30 % logran perfecto

5.9 No se rechaza H_0 , no hay evidencia de valores mayores que 70 %

5.10 Se rechaza H_0 , superan 80 ml/kg/min

Referencias

- [1] R. Hernández Sampieri, C. Fernández Collado, and P. Baptista Lucio, *Metodología de la investigación*. México: McGraw-Hill, 6 ed., 2014.
- [2] E. Andreu-Cabrera and F. J. Romero-Naranjo, “Neuromotricidad, psicomotricidad y motricidad. nuevas aproximaciones metodológicas,” *Retos: Nuevas Tendencias en Educación Física, Deporte y Recreación*, vol. 42, pp. 924–938, 2021.
- [3] S. Collado Vázquez, C. Pérez García, and J. Carrillo Esteban, *Motricidad: Fundamentos y aplicaciones*. Madrid, España: Dykinson, 2004.
- [4] M. N. Ozer, *The Ozeretsky Scale of Motor Development: A translation from the Russian adaptation*. Seattle, WA: University of Washington Press, 1987.
- [5] M. N. Ozer, “Escala de desarrollo motor de ozeretsky: Adaptación y validación para población infantil latinoamericana,” *Revista Latinoamericana de Psicología*, vol. 24, no. 2, pp. 245–260, 1992.
- [6] T. Ke, “The development of children’s autonomy and reasonable paternalistic intervention,” *Humanities and Social Sciences Communications*, vol. 10, no. 1, p. 874, 2023.
- [7] K. Norton and T. Olds, *Standards for Anthropometry Assessment*. London / New York: Informa – Taylor Francis, 4 ed., 2025.
- [8] F. Gimigliano, G. Resmini, A. Moretti, M. Aulicino, F. Gargiulo, A. Gimigliano, S. Liguori, M. Paoletta, and G. Iolascon, “Epidemiology of musculoskeletal injuries in adult athletes: A scoping review,” *Medicina*, vol. 57, no. 10, p. 1118, 2021.
- [9] L. Sun, L. Luo, Y. Yang, Q. Wangchuan, and J. Luo, “Gender differences in the incidence and types of sports injuries among female athletes: A scoping review,” *International Journal of Sports Medicine and Related Sciences*, vol. –, no. –, p. –, 2025.

Bibliografía

- [10] R. Pérez Juste, J. L. García Llamas, and J. A. Gil Pascual, *Estadística aplicada a la educación*. Madrid: Pearson Educación, 2012.
- [11] H. Martínez Puertas, E. M. Artés Rodríguez, S. Martínez Puertas, and M. Casimiro Artés, *Métodos estadísticos aplicados a las Ciencias de la Actividad Física y el Deporte*. Almería: Editorial Universidad de Almería, 2020.
- [12] R. Selker, J. Love, and D. Dropmann, “jamovi: An easy to use statistical software that works,” *Journal of Statistical Software*, vol. 88, no. 1, pp. 1–7, 2019.
- [13] The jamovi Project, *jamovi Documentation*, 2024. Documentación oficial.
- [14] C. Pérez, *Estadística Aplicada a Través de Excel*. Madrid, España: Ediciones Paraninfo, 2ª ed., 2022.
- [15] P. A. López, *Probabilidad y estadística: Conceptos, modelos y aplicaciones en Excel*. Prentice Hall, México: Pearson Educación, 2009.
- [16] C. Salas-Eljatib, *Análisis de datos con el programa estadístico R: Una introducción aplicada*. Santiago de Chile, Chile: Universidad Mayor, 2021.
- [17] W. Vásquez and J. Ortiz, *Análisis Estadístico con Excel: Aplicaciones en Investigación*. Guayaquil, Ecuador: Editorial Universidad de Guayaquil, 2ª ed., 2022.
- [18] D. C. Montgomery and G. C. Runger, *Probabilidades y estadística para ingeniería*. México: Limusa Wiley, 5 ed., 2010.
- [19] D. A. Lind, W. G. Marchal, and S. A. Wathen, *Estadística aplicada a los negocios y la economía*. México: McGraw-Hill, 15 ed., 2012.
- [20] D. M. Levine, D. F. Stephan, and K. A. Szabat, *Estadística para Administración con Excel*. México: Pearson Educación, 8ª ed., 2023.

Estadística en Acción, Educación y Movimiento

La estadística es una herramienta fundamental para comprender fenómenos, interpretar información y sustentar decisiones basadas en evidencia. Su aplicación permite transformar observaciones y datos en conocimiento útil para el análisis y la práctica profesional.

Estadística en Acción: Educación y Movimiento ha sido concebido como una guía de aprendizaje para los campos de la Educación Inicial y la Actividad Física integrando los fundamentos estadísticos con aplicaciones contextualizadas y herramientas tecnológicas orientadas a fortalecer el pensamiento analítico. Los contenidos han sido organizados de tal forma que el lector encuentre conceptos fundamentales, ejemplos desarrollados, ejercicios propuestos y actividades prácticas que facilitan la comprensión y aplicación de los métodos estadísticos.

El libro integra el uso de software especializado como Jamovi y Microsoft Excel mediante laboratorios guiados sin perder de vista la interpretación y el significado de los resultados.

Dirigido a estudiantes, docentes y profesionales vinculados a la Educación en la primera infancia y a la Actividad Física, este libro busca convertir el aprendizaje estadístico en una experiencia dinámica.

ISBN: 978-9942-652-53-9



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA