Esta obra presenta, de una manera sencilla, para estudiantes y docentes, la utilidad de las Ciencias Estadísticas en el campo deportivo; es trascendental el monitoreo de un atleta o grupo deportivo en este espacio. En las secciones preliminares de este texto muestra inicialmente el análisis de la información mediante la Estadística Descriptiva y la Teoría de la probabilidad, con ayuda de estadísticos sencillos de evaluar, como también la visualización de esta información con la aplicación de gráficas descriptivas. Una vez realizado el estudio exploratorio, el siguiente paso presentado en el documento, es aplicar la Estadística Inferencial como un método básico para el análisis final. En las últimas secciones es relacionada una nueva propuesta de indicador de desempeño para la evaluación de la capacidad de un atleta en referencia a unos estándares de trabajo.

El libro, además, muestra el procedimiento teórico de cada una de las técnicas estadísticas aplicadas en la solución de los ejemplos, como también el uso del programa R-project https://www.r-project.org/, como medio, aplicando comandos o una estructura de programación, en la solución de dichos problemas. Es importante resaltar en este medio, la escasa o nula temática recabada en escritos, en la importancia de los métodos estadísticos en las innumerables actividades deportivas. El contenido de este libro aporta en forma preliminar hacia esta dirección, aceptando con agrado las sugerencias de los diversos autores o responsables en este ámbito.

Escanée el código QR para conocer más títulos publicados por el Sello Editorial Universidad del Atlántico

Universidad del Atlántico

ISBN 978-958-5525-20-7

ESTADÍSTICA DEPORTIVA

Roberto José Herrera Acosta

Sello Editorial Universidad del Atlántico

Roberto José Herrera Acosta

ESTADÍSTICA DEPORTIVA

Roberto José Herrera Acosta

ESTADÍSTICA DEPORTIVA

Universidad del Atlántico

Roberto José Herrera Acosta

ESTADÍSTICA DEPORTIVA

Sello Editorial
UNIVERSIDAD DEL ATLÁNTICO

Catalogación en la publicación. Universidad del Atlántico. Departamento de Bibliotecas

Herrera Acosta, Roberto José

Estadística Deportiva / Roberto José Herrera Acosta -- Barranquilla: Sello Editorial Universidad del Atlántico, 2018. 152 páginas. 21x27 centímetros. Ilustraciones. Incluye bibliografía.

ISBN 978-958-5525-20-7 (Libro descargable PDF)

1. Estadística 2. Estadística -- Educación física. I. Herrera Acosta Roberto José.

CDD 519.5 H565

Estadística Deportiva
Autoría: Roberto José Herrera Acosta

Universidad del Atlántico, 2018

Edición:
Sello Editorial Universidad del Atlántico
Km 7 Vía Puerto Colombia (Atlántico)
www.uniatlantico.edu.co
publicaciones@mail.uniatlantico.edu.co

Producción Editorial:
Calidad Gráfica S.A.
Av. Circunvalar Calle 110 No. 6QSN-522
PBX: 336 8000
info@calidadgrafica.com.co
Barranquilla, Colombia

Publicación Electrónica

Cómo citar este libro:
Herrera Acosta R. J. (2018) *Estadística Deportiva*. Barranquilla. Editorial Universidad del Atlántico.

A mis hijos Luis Roberto y Luz Karime

AGRADECIMIENTOS

Agradecimiento a la Universidad del Atlántico y en especial a los compañeros de los programas de Matemáticas, Educación Física e Ingeniería Industrial por su constante apoyo.

Contenido

Agradecimientos 7

Apéndice: Tabla de parámetros 21

Prólogo 23

Introducción 24

Unidad 1
Análisis preliminar de la información 26

Unidad 2
Medidas de localización, variabilidad y forma 46

Unidad 3
Las medidas de probabilidad 63

Unidad 4
Distribuciones: discretas - continuas de probabilidad 72

Unidad 5
Estadística de algunos deportes 84

Unidad 6
Estimación puntual e intervalo de confianza.
Metodología del contraste de hipótesis 89

Unidad 7
Metodología del contraste de hipótesis 98

Unidad 9
El análisis de varianza anova 106

Unidad 8
Modelo de regresión lineal simple 114

Unidad 10
Tablas de contingencia-los modelos logísticos-capacidad de rendimiento deportivo 119

Actividades evaluativas 136

Referencias Bibliográficas 143

Anexos 146

Glosario 152

Lista de Tablas

Tabla 1
Esquema de una Tabla de frecuencia. 28

Tabla 2
Tiempo desarrollado por un nadador en competencia 28

Tabla 3
Tabla de frecuencia del tiempo desarrollado por un nadador en competencia. 29

Tabla 4
Información de los pesos en libras de boxeadores de la categoría Pesada 30

Tabla 5
Información de los pesos en libras de boxeadores de la categoría Pesada 31

Tabla 6
Tiempos de un corredor de 40 mts. planos 32

Tabla 7
Frecuencia de los tiempos de un corredor de 40 mts planos... 32

Tabla 8
Información de los tiempos, en minutos, obtenidos de las diferentes
actividades de un nadador.. 32

Tabla 9
Frecuencia de los tiempos del nadador 33

Tabla 10
Frecuencia de ocurrencia de las modalidades.. 35

Tabla 11
Número de juegos clasificados de un equipo en una temporada 35

Tabla 12
Número de errores durante una temporada de tres posiciones en el campo de juego 36

Tabla 13
Información de los tiempos de un velódromo.. 38

Tabla 14I
nformación de la distancia recorrida por el jugador de fútbol en la Liga Colombiana 38

Tabla 15I
nformación del número de no conformidades o fallas presente en un juego de basquetbol 40

Tabla 16
Peso y el tiempo registrado en un encuentro deportivo 43

Tabla 17
Tiempo en segundos de un nadador 47

Tabla 18
La tabla de frecuencia o distribución empírica del contenido de colesterolde un equipo de futbol 48

Tabla 19
Contenido glicémico de los aspirantes a formar parte del equipo ciclístico universitario 49

Tabla 20
Tiempo en horas de entrenamiento de los ciclistas.. 50

Tabla 21
Peso en gramos de la proteína ofrecida a un atleta de alto rendimiento 50

Tabla 22
Conteo del peso de alumnos de un equipo de futbol americano...................................... 51

Tabla 23
Tiempo de un circuito de fórmula uno 52

Tabla 24
Número de dianas de un equipo de fútbol 53

Tabla 25
Número de inscritos en las actividades deportivas 54

Tabla 26
Fallas de un automóvil por kilómetros recorridos en competencias de Rally 56

Tabla 27
Porcentaje de grasa de un grupo de atletas organizada en una Tabla de frecuencia 58

Tabla 28
Peso levantado y el tiempo establecido 60

Tabla 29
Cálculo de sumas de cuadrado del peso y tiempo de levantamiento 61

Tabla 30
Contingencia en función de los años de experiencia deportiva y su edad 66

Tabla 31
Contingencia en función de los años de experiencia deportiva y su edad 67

Tabla 32
Distribución Binomial y Poisson 73

Tabla 33
Estandarización de una variable con distribución normal y parámetros conocidos 80

Tabla 34
Distribuciones continuas de probabilidad 82

Tabla 35
Distribuciones continuas de probabilidad (continuación) 83

Tabla 36
Factor de fuerza de la confederación en las tres últimas copas mundiales y los valores equivalentes 85

Tabla 37
Criterio de puntajes FIFA 85

Tabla 38
Ejemplo de los criterios de puntajes FIFA 86

Tabla 39
Kilómetros completados en cada equipo 87

Tabla 40
Ranking de kilómetros promedio por marca de motor 88

Tabla 41
Posición de los pilotos de carrera en la temporada 2017 88

Tabla 42
Estandarización de los modelos de distribución de muestreo 90

Tabla 43
Valores estandarizados de Z basado en el nivel de significancia confiabilidad 92

Tabla 44
Peso levantado por dos atletas en intervalos de tiempo equidistantes 94

Tabla 45
Errores en los contrastes de hipótesis 99

Tabla 46
Estadísticos de prueba en contraste de hipótesis.. 100

Tabla 53
Descomposición de la suma de cuadrados y el contraste de la hipótesis.......................... 107

Tabla 54
Carga aerodinámica en porcentaje, de cada uno de los alerones de prueba...................... 108

Tabla 49
Descomposición de la suma de cuadrados y el contraste de hipótesis 108

Tabla 56
Tiempo recabado de cada uno de los paracaídas, utilizado en el experimento................. 111

Tabla 57
Descomposición de la suma de cuadrados y el contraste de hipótesis 112

Tabla 52
Prueba LSD de los tipos de materiales para equipo de parapentes.................................. 113

Tabla 47
La suma de cuadrados del modelo de regresión simple.. 115

Tabla 48
Variables relacionadas: conteo de la disminución de la velocidad y
porcentaje de oxígeno del entorno 116

Tabla 49
Modelo de regresión simple del conteo en la disminución de la velocidad y
la variable independiente el porcentaje de concentración de oxígeno en el entorno 117

Tabla 50
Variable peso del Atleta y tiempo de recorrido en segundos 117

Tabla 51
Modelo de regresión simple del conteo en la disminución de la velocidad y
la variable independiente el porcentaje de concentración de oxígeno en el entorno 118

Tabla 52
Coeficientes del modelo de regresión simple.......... 118

Tabla 59
Ejemplo de Tabla de contingencia de doble entrada.......... 119

Tabla 60
Contingencia para las variables Equipos x Resultado.......... 120

Tabla 61
Valores esperados de la Tabla de contingencia
para las variables Equipos x Resultado.......... 121

Tabla 62
Contingencia para las variables Equipos x Resultado
cálculo del estadístico de prueba.......... 121

Tabla 63
Valores calculados del Estadístico.......... 121

Tabla 64
Contingencia para las variables Equipos x Resultado,
cálculo razón de verosimilitud.......... 122

Tabla 65
Medidas simétricas.......... 122

Tabla 66
Frecuencias en una Tabla de contingencia 2×2.......... 123

Tabla 67
Nivel de colesterol y rendimiento del atleta.......... 127

Tabla 68
Variable peso del Atleta y tiempo de recorrido en segundos.......... 133

Tabla 69
Intervalos de confianza, variable pH
y concentración de un producto alimenticio.......... 134

Tabla 70
Tabla de frecuencia.......... 136

Tabla 71
Ángulo de conversión.......... 137

Tabla 72
Información de las características de calidad de una unas zapatillas deportivas.......... 137

Tabla 73
Kilómetros recorridos por un ciclista en un circuito contrarreloj 138

Tabla 74
Tabla de frecuencia de errores en un equipo de patinaje 139

Tabla 75
Categorías de carreras de automóvil 140

Tabla 76
Inventario de los productos alimenticios ofrecidos a una delegación
deportiva departamental 140

Tabla 77
El recorrido en kilómetros de los jugadores de un equipo de fútbol 141

Tabla 78
Información de la velocidad de las máquinas de bateo 141

Tabla 79
Costos asociados a la contratación de jugadores de béisbolen cuatro ciudades
de la costa norte de Colombia 142

Tabla 80
Contusiones de los estudiantes universitarios en Fútbol y Baloncesto 142

Lista de Gráficas

Gráfica 1
Histograma de los pesos en libras de boxeadores de la categoría Pesada 31

Gráfica 2
Polígono de frecuencia.. 33

Gráfica 3
Histograma de frecuencia con diferente tamaño Sturges... 34

Gráfica 4
Diagrama de barras del tipo de juego del equipo.. 36

Gráfica 5
Diagrama de barras de los jugadores de béisbol... 37

Gráfica 6
Polígono de frecuencia de los tiempos de espera.. 38

Gráfica 7
Polígono de frecuencia de la distancia recorrida por el jugadorde fútbol en la
Liga Colombiana, salida en R 39

Gráfica 8
Diagrama de torta de las fallas presentadas en un juego de basquetbol 41

Gráfica 9
Diagrama radial del número de fallas presentadas en un juego de basquetbol................. 41

Gráfica 10
Diagrama de torta de las fallas salida en R 42

Gráfica 11
Diagrama de seguimiento o serie de tiempo 42

Gráfica 12
Correlación muestra la asociación entre la variable X y la variable Y 43

Gráfica 13
Diagrama de barras del tipo de juego del equipo... 44

Gráfica 14
Diagrama de correlación . Salida en R... 44

Gráfica 15
Diagrama de barras del porcentaje de grasa de un grupo de atletas 58

Gráfica 16
Diagramas distribución es Mesocúrtica, Leptocúrtica y Platicúrtica............................... 59

Gráfica 17
Diagrama de Venn de la medida de probabilidad de la unión e intercepción
de dos eventos. 68

Gráfica 18
Obtención de la Regla General de la Multiplicación del Evento B (color) 69

Gráfica 19
Distribución Binomial con parámetro $p = 0.1, n = 10$ 72

Gráfica 20
Distribución Poisson con parámetro $\lambda = 0.5$. 75

Gráfica 21
Distribución Geométrica con parámetro $p = 0.43$.. 77

Gráfica 22
Modelo de la distribución Normal 79

Gráfica 23
Diagrama box plots de los diferentes tipos de alerones 109

Gráfica 24
El parapente, materiales utilizados para aumentar el tiempo de vuelo 111

Gráfica 25
Diagrama de box plots para los cuatro tipos de materiales del equipo de parapente......... 112

Gráfica 26
Índice de capacidad para la variable pero del atleta $C_p = 0,5932$. 133

Gráfica 27
Índice de capacidad para la variable tiempo de recorrido $C_p = 0,6626$. 134

Gráfica 28
Histograma de frecuencia modelo supuesto 136

Gráfica 29
Diagrama de correlación entre el tiempo y kilómetros recorrido 138

Gráfica 30
Histograma de frecuencia de una información deportiva 139

Apéndice: Tabla de parámetros

Signos	Descripción
USL	Límite de especificación superior
LSL	Límite de especificación inferior
C_P	Índice de capacidad óptimo o a corto plazo
σ	Desviación estándar de la población
s	Desviación estándar de la la muestra
ϕ	Función acumula de la distribución normal estándar
t	Distribución t de Student
χ^2	Distribución chi cuadrado
F	Distribución F de Fiseher

Prólogo

Este texto de investigación es un aporte en el área de la **Estadística Aplicada** a profesionales del deporte; son pocos los documentos en el campo deportivo que involucran el análisis de datos como parte esencial que permitirán un análisis más asertivo del comportamiento de las actividades; esto es realizado con formulaciones no complejas que permita a los lectores con conocimientos básicos del tema, una comprensión sencilla.

La bibliografía en este campo es sumamente amplia, pero es considerado pertinente redactar un libro, específicamente como curso introductorio de Estadística en el área deportiva. Esta delimitación pretende facilitar, el trabajo de los estudiantes de pregrado y postgrado, dedicados a esta área en particular, exponiendo de manera precisa los conceptos esenciales de la estadística descriptiva e inferencial. Estas concepciones son desarrolladas progresivamente a lo largo del texto; los estudiantes del área deportiva encontrarán menos dificultades para su formación, profundizando las técnicas estadística más avanzadas.

INTRODUCCIÓN

En el ámbito deportivo es frecuente que la información sea analizada con herramientas estadísticas, que permitan conocer los requerimientos estratégicos de cómo asumir la competencia de un equipo y no menos importante de un atleta en particular. La experiencia muestra muchos ejemplos, que evidencian la utilidad de la estadística en las habilidades o destrezas de un participante o grupo de participantes en una competencia deportiva. Esto ha permitido que el educador o entrenador obtenga un panorama más clarificado de la situación a que somete el grupo o atleta en una competencia., Este texto pretende encausar a las personas que en su quehacer cotidiano se encuentran en la actividad deportiva, ya sea como educador o entrenador.

En la primera unidad se describe la forma correcta de obtener la información, con el objeto de poseer la mayor veracidad, de manera de alcanzar una decisión incontrovertible del fenómeno o actividad.

El siguiente tema trata de la etapa de depuración de la información y una vez realizada esta acción se organizan los datos en las tablas de frecuencia. A partir de esta Tabla, se construye el histograma de frecuencia que permite de una manera sencilla observar el comportamiento de la información. También se encuentra la elaboración de otros diagramas muy conocidos como lo son: el diagrama de barras y el de torta que es aplicado para casos particulares de información.

Posteriormente el texto presenta los conceptos de medida de localización o posición, variabilidad y de forma. La importancia de obtener estos estadísticos en la información. Medidas de localización muy conocidas como la media aritmética, la mediana, la varianza, desviación y otras medidas. El coeficiente de asimetría y la curtosis, que permite evaluar la forma o el tipo de modelo que toma la información. Seguidamente, se muestra la elaboración de los diagramas de barras y sectores para análisis de variables cualitativas. En los últimos temas de esta unidad, se desarrolla la importancia de la correlación existente en el campo bivariado y el modelo lineal simple entre estas variables, con el objeto de realizar una estimación de la variable que se toma como respuesta.

El texto presenta los criterios de evaluar una medida de probabilidad asumiendo normalidad en la información. Inicialmente se muestra cómo se estandariza una variable aleatoria, puntuación *Z*, y la importancia para la interpretación de un fenómeno.

En la unidad número cuatro se presenta el procedimiento estadístico de algunos deportes, que evidencia la utilidad de la estadística en el monitoreo, el análisis y la posterior toma de decisiones de la forma como un atleta o equipo debe asumir una competencia. En la quinta unidad se observan algunos ejemplos de estadísticas descriptivas de algunos de los deportes más conocidos; en la sexta unidad se encuentran herramientas estadísticas, inferencias como son los intervalos de confianza y prueba de hipótesis.

En la séptima unidad se desarrolla la tabla de análisis de varianza como una adecuada metodología de decisión, cuando es comparando dos o más poblaciones o niveles en la actividad deportiva desarrollada. En la octava unidad se muestra la utilidad del modelo lineal simple, como herramienta inicial de asociación de dos variables.

En la unidad nueve, es presentada la tabla de contingencia 2 x 2 que permite realizar un análisis de la dependencia o independencia o grado se asociación, entre dos variables cualitativas nominales o denominados factores. A partir de estas tablas se construyen modelos de regresión logística, analizando las dependencias de una variable binaria de otras variables que representan otras características consideradas en forma conjunta.

UNIDAD 1

ANÁLISIS PRELIMINAR DE LA INFORMACIÓN

INTRODUCCIÓN

El concepto de **población y muestra** es fundamental en el análisis estadístico. Se denomina **población** al conjunto completo de elementos finitos o infinitos de un fenómeno de interés, evaluando en esta una característica como objeto de estudio. En el momento de seleccionar aleatoriamente, un número de elementos representativos de esta población; la designación conceptualmente de estos elementos es acreditada como tamaño de la **muestra**. En esta última, las características o resultados evaluadas son denominados estadísticos

Toma de la información de una muestra

La información numérica o conceptual que se obtiene de un proceso de una actividad deportiva, tales como el tiempo transcurrido por un nadador, el tiempo de un atleta en la carrera de los cien metros, los metros recorridos por una jabalina en el momento de ser lanzada. Una vez tomada esta información es necesario determinar el comportamiento de los datos, con el objetivo de obtener el comportamiento de la información. Sin lugar a duda, la forma más sencilla de lograrlo es, utilizando la estadística descriptiva, es decir diagramas o gráficas que permitan evaluar este comportamiento, así como, estadísticos sencillos que permitan una clara interpretación del fenómeno evaluado.

Depuración de la información

Se debe realizar un análisis de la información que se selecciona de la actividad deportiva que en ese momento se evalúa con el objetivo de acuerdo con la experiencia obtenida, eliminar o depurar información ajena a las condiciones naturales de la actividad. Por ejemplo, un deportista que, en el momento de realizar la actividad, se encuentra en condiciones no aptas, la información obtenida de este atleta por lo general será un dato atípico al normal desarrollo de la actividad. Es por tanto imperativo tomar la información de este atleta y eliminarla, ya que sesgará el análisis que de ahí se desprenda.

CONSTRUCCIÓN DE UNA TABLA DE FRECUENCIA

La Tabla de frecuencia es una organización de la información seleccionada de la actividad que se pretende evaluar. Su estructura tradicional consta de seis columnas: la primera presenta los intervalos de clase con sus respectivos límites inferior y superior, en la segunda se calcula el punto medio de ese intervalo de clase, la tercera columna muestra la frecuencia absoluta o de clase, una cuarta presenta la frecuencia relativa de clase, en la quinta se observa la frecuencia acumulada y la última la frecuencia relativa acumulada.

Etapas en la construcción de una Tabla de frecuencia

El primer paso es determinar el rango del conjunto de observaciones existentes diferenciando los valores extremos de la información suministrada; la ecuación aplicada para evaluar el rango es la siguiente,

$$\mathrm{R} = \text{valor máximo} - \text{valor mínimo} \qquad (1)$$

Una vez evaluado el rango se determina el número de clases o intervalos de clase utilizando la regla de Sturges (STURGES, 1926), para evaluar el número de clases o intervalos:

$$\mathrm{k} = 1 + 3.32\log(\mathrm{n}) \qquad (2)$$

donde n representa el tamaño de la muestra seleccionada en la población N, obtenida de la expansión de la distribución binomial

$$n = \sum_{j=0}^{k-1}\binom{k-1}{i} = (1+1)^{k-1} = 2^{k-1} \qquad \text{donde } 0 \leq n \leq N$$

Scott presenta una alternativa para determinar el número de clases o intervalo de clase, (SCOTT, 1979).

$$h = 3.5sn^{-1/3}$$

donde s es la desviación estándar muestral, *n* el tamaño de la muestra.

$$h = 2(IQ)n^{-1/3}$$

donde IQ es el rango intercuantil de la muestra

Posteriormente, es determinar la amplitud del intervalo. La amplitud está definida como la razón entre el rango y el número de clase. Es decir, la razón entre lo calculado en el paso uno y dos. La ecuación de esta razón se define como $A = \frac{R}{k}$. Posteriormente se construye la Tabla de frecuencia tal como se muestra en la Tabla 1.

Tabla 1

Esquema de una Tabla de frecuencia.

Intervalo de clase	Punto medio	Frecuencia de clase	Frecuencia relativa de clase	Frecuencia acumulada	Frecuencia relativa acumulada
LIMINF-LIMSUP	$x_i = \frac{limsup - liminf}{2}$	f.	$f_r = f/n$	F.	$F_r = \frac{F}{n}$.
.	.	.	.	.	..

⊙ **Fuente:** elaboración por el Autor

La columna de las marcas de clase de cada intervalo es el valor medio entre los límites inferior y superior de cada intervalo de clase. En la siguiente columna es la frecuencia de clase f se define el número de observaciones existentes en el intervalo de clase. En la cuarta columna es determinada la frecuencia relativa f_r de clase, definida como la razón entre la frecuencia de clase y el número total de observaciones. La penúltima columna es la frecuencia acumulada de clase *F*, evaluada en la acumulación de las frecuencias de clase, en el descenso de los intervalos de clase. La última columna. F_r definida como la frecuencia relativa acumulada de clase, estimada de la razón, entre la frecuencia acumulada y el número total de observaciones.

Ejercicio 1.1

La información presentada en la Tabla 2, muestra el tiempo desarrollado por un nadador en competencia. ¿Qué porcentaje de las veces el tiempo está comprendido entre 18,5 minutos y 26.3?

Tabla 2

Tiempo desarrollado por un nadador en competencia

22	21	18
19	32	33
20	16	35
15	16	17

⊙ **Fuente:** elaboración por el autor

El rango e intervalo de clase de las ecuaciones (1) y (2), es respectivamente $R = 35 - 15 = 20$ y $k = 1 + 3.32\log(12) = 4.58 \approx 5$ donde el tamaño de la muestra es $n = 15$. La amplitud es obtenida mediante $A = \frac{R}{k} = \frac{20}{5} = 4$.

Tabla de frecuencia en R

Es de aclarar, la salida del programa R, genera resultados que difieren del procedimiento expuesto en el ítem 1.2.1. Esto es debido al número de intervalos que genera este software estadístico.

nombre<-c(dato[1],..,dato[n])	**#### lectura de los datos**
nuevonombre<-table(nombre)	### crea una nueva variable “nuevonombre”
nuevonombre	### salida de la nueva variable
cumsum(nuevonombre)	##### crea una Tabla de la variable “nuevo-nombre”, se obtiene la frecuencia absoluta.
cumsum(nuevonombre)/margin.table(nuevonombre)	### genera la frecuencia relativa porcentual acumulada.

Comandos utilizados en el ejercicio 1

```
tiempo<-c(22,21,18,19,32,33,20,16,35,15,16,17)
tabletiempo<-table(tiempo)
tabletiempo
cumsum(tabletiempo)
cumsum(tabletiempo)/margin.table(tabletiempo)
```

En este caso, se presenta el procedimiento anteriormente mencionado, distinto a lo generado en el programa R, se observa que el primer intervalo se obtiene de la suma del valor mínimo más la amplitud, es decir $15.0 + 4.0 = 19.0$, Tabla 3. En el caso del segundo intervalo se toma el anterior cálculo (límite superior del primer intervalo) como límite inferior del segundo intervalo y posteriormente se suma la amplitud, para obtener el límite superior del segundo intervalo de clase.

Tabla 3

Tabla de frecuencia del tiempo desarrollado por un nadador en competencia.

Intervalo de clase	Punto medio x_i	Frecuencia de clase f	Frecuencia relativa de clase f_r	Frecuencia acumulada F	Frecuencia relativa acumulada F_r
15,0-19,0	17	6	0,0833	6	0,5833
19,0-23,0	21	3	0,4167	9	0,7500
23,0-27,0	25	0	0,2500	9	0,7500
27,0-31,0	29	0	0,0000	9	0,7500
31,0-35,0	33	3	0,2500	12	1,0000

⊙ **Fuente:** elaboración por el Autor

HISTOGRAMA DE FRECUENCIA

Wand, M.P. indica la importancia de este diagrama; entre estas se encuentra visualizar el comportamiento de la información, ajustándolo a un modelo estadístico conocido; entre los modelos más usuales se encuentran, (WAND, 1995).

a. El histograma tipo normal o de campana que se presenta en la información de fenómenos que requieren especificaciones de cumplimiento, por ejemplo, tiempos de embalaje de un producto, su peso o alguna característica de calidad.
b. El histograma de frecuencia de tipo exponencial se presenta cuando la información proviene de tiempos de servicio de un proceso, en donde se establece ya sea un tiempo mínimo o un tiempo máximo de cumplimiento.
c. El histograma de frecuencia tipo variable, es indicativo de que la información posee mucha variabilidad, lo que implica inestabilidad en la actividad que se esté realizando.

Existen diferentes comportamientos del histograma, es el caso bimodal, en donde se presenta la mezcla de dos poblaciones. Otro de los histogramas es el multimodal, en donde se presentan más de dos poblaciones en la información seleccionada. Para elaborar el histograma de frecuencia se selecciona de la tabla de frecuencia los límites del intervalo de clase, en el diagrama se presenta en el eje de las abscisas. Posteriormente se selecciona la frecuencia absoluta o de clase (se puede tomar la frecuencia relativa en reemplazo de la frecuencia absoluta) y se presenta en el diagrama en el eje de las ordenadas.

Ejercicio 1.2

La siguiente información, Tabla 4, presenta los pesos obtenidos en libras de unos boxeadores de la rama juvenil, de la categoría pesada de la WBA. Se desea establecer si esta variable posee un comportamiento tipo normal o de campana.

Tabla 4
Información de los pesos en libras de boxeadores de la categoría Pesada

199,83	199,89	198,92	200,04	199,90	198,73
200,85	199,50	200,40	200,71	199,09	200,13
199,27	197,11	200,60	199,88	201,78	198,15
198,47	200,69	199,61	198,73	200,36	197,93
202,81	199,59	199,70	201,16	199,03	200,51
200,37	198,78	200,28	199,96	201,57	200,81

⊙ **Fuente:** elaboración del autor

El rango es $R = 201.78 - 197.1104.67$. El número de clases o intervalos de clase $i = 1 + 3.32\log(36) = 6.16 \approx 7$, donde $n = 36$. La amplitud del intervalo es $A = \frac{R}{k} = \frac{4.67}{7} = 0.66$.

El siguiente paso, es la construcción de la tabla de frecuencia, Tabla 5, como se muestra a continuación,

Tabla 5

Información de los pesos en libras de boxeadores de la categoría Pesada

Intervalo de clase	Punto medio x_i	Frecuencia de clase f	Frecuencia relativa de clase f_r	Frecuencia acumulada F	Frecuencia relativa acumulada F_r
197,11-197,77	197,44	1	0,0278	1	0,0278
197,77-198,44	198,11	2	0,0556	3	0,0833
198,44-199,11	198,77	6	0,1667	9	0,2500
199,11-199,77	199,44	4	0,1111	13	0,3611
199,77-200,44	200,11	12	0,3333	25	0,6944
200,44-201,11	200,77	7	0,1944	32	0,8889
201,11-201,78	201,44	4	0,1112	36	1,0000

⊙ **Fuente elaboración del autor**

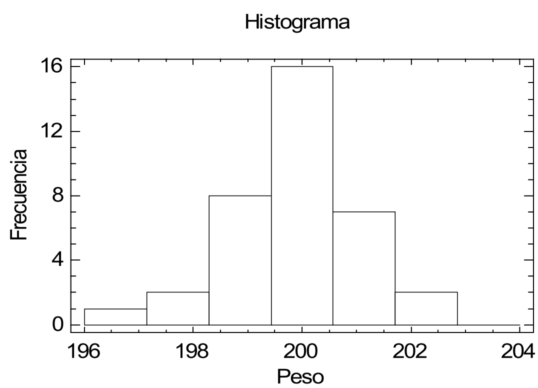
La Gráfica 1, muestra la información del peso de los deportistas presenta un histograma tipo normal o de campana de Gauss.

Histograma de frecuencia en R

nombre<-c(dato[1],..,dato[n])	#### lectura de los datos
hist(nombre)	### crea el histograma de frecuencia

Comandos utilizados en el ejercicio 2

```
pesos boxeadores <-c(199.83,199.89,198.92,200.04,199.90,198.73,200.85,199.50,200.40,200.71,199.09,200.13,199.27,
197.11,200.60,199.88,201.78,198.15,198.47,200.69,199.61,198.73,200.36,197.93,202.81,199.59,199.70,201.16,199.03,
200.51,200.37,198.78,200.28,199.96,201.57,200.81)
hist(pesosboxeadores)
```



Gráfica 1

Histograma de los pesos en libras de boxeadores de la categoría Pesada

⊙ **Fuente:** salida programa R

Ejercicio 1.3

Los tiempos de un corredor de 40 mts planos, presentadas en la Tabla 6. El propósito es elaborar una Tabla de frecuencia o distribución empírica; obtenga la medida de probabilidad de que el tiempo de recorrido se encuentre entre 5.67 y 6.23.

Tabla 6

Tiempos de un corredor de 40 mts. planos

7,29	7,03	4,55
7,35	6,91	5,05
7,35	7,27	4,83
7,38	6,00	6,05

⊙ **Fuente:** elaboración por el Autor

En la información anterior, el rango obtenido es $R = 7.38 - 4.55 = 2.83$ el número de clase obtenido es $k = 1 + 3.32 \log 12 = 4.58 = 5$ y la amplitud es $A = R/k = 2.83/5 = 0.56$, posteriormente se elabora Tabla de frecuencia generada, la Tabla 7

Tabla 7

Frecuencia de los tiempos de un corredor de 40 mts planos

Intervalo de clase		Punto medio x_i	Frecuencia de clase f	Frecuencia relativa de clase f_r	Frecuencia acumulada F	Frecuencia relativa acumulada F_r
4,55	5,11	4,83	3	0,25	3	0,25
5,11	5,67	5,39	0	0,00	3	0,25
5,67	6,23	5,95	2	0,16	5	0,41
6,23	6,79	6,51	0	0,00	5	0,41
6,79	7,38	7,85	7	0,58	12	1,00

⊙ **Fuente:** elaboración por el autor

La medida de probabilidad del intervalo de tiempo recorrido entre 5,67- 6,23 es de 0,16, en terminos de porcentaje, el 16 %.

Ejercicio 1.4

Se suministra la información de los tiempos, en minutos, obtenidos de las diferentes actividades de un nadador que es preparado para los Juegos Nacionales, Tabla 8.

Tabla 8

Información de los tiempos, en minutos, obtenidos de las diferentes actividades de un nadador

5,32	4,95
5,41	5,10
5,89	5,04
5,12	4,93
4,90	4,95
5,42	5,90

⊙ **Fuente:** elaboración por el autor

A. Elabore una tabla de frecuencia del tiempo tomado por el nadador

B. Determine el comportamiento del nadador basado en la información suministrada por el histograma.

C. ¿Cuál es la posibilidad de clasificar de este nadador, si el comité exige que, para poder clasificar a un evento en particular, es necesario obtener un tiempo por debajo a 5 minutos?

La Tabla 9, frecuencia del tiempo del nadador, muestra que el mayor tiempo de las actividades de este nadir se encuentra entre 5,28-5,52 minutos con el 25 %.

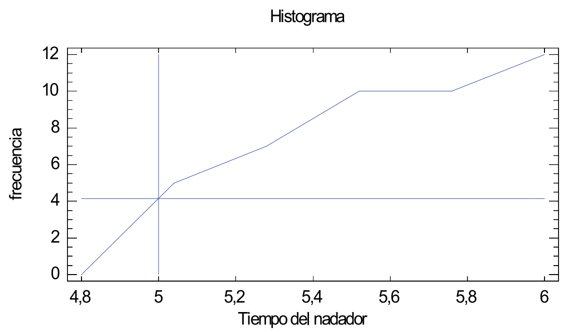
Tabla 9

Frecuencia de los tiempos del nadador

Intervalo de clase		Punto medio x_i	Frecuencia de clase f	Frecuencia relativa de clase f_r	Frecuencia acumulada F	Frecuencia relativa acumulada F_r
4,8	5,04	4,92	5	0,4167	5	0,4167
5,04	5,28	5,16	2	0,1667	7	0,5833
5,28	5,52	5,4	3	0,2500	10	0,8333
5,52	5,76	5,64	0	0,0000	10	0,8333
5,76	6,0	5,88	2	0,1667	12	1,0000

⊙ **Fuente:** elaboración por el autor

El 83,3 % de las veces, el nadador posee un tiempo por debajo de 5,52, es decir el tiempo de este atleta permanece generalmente por debajo de este valor. El polígono de frecuencia muestra que el 4,15 % de las veces el nadador obtiene valores por debajo de 5 minutos.



Gráfica 2

Polígono de frecuencia

⊙ **Fuente:** salida programa R

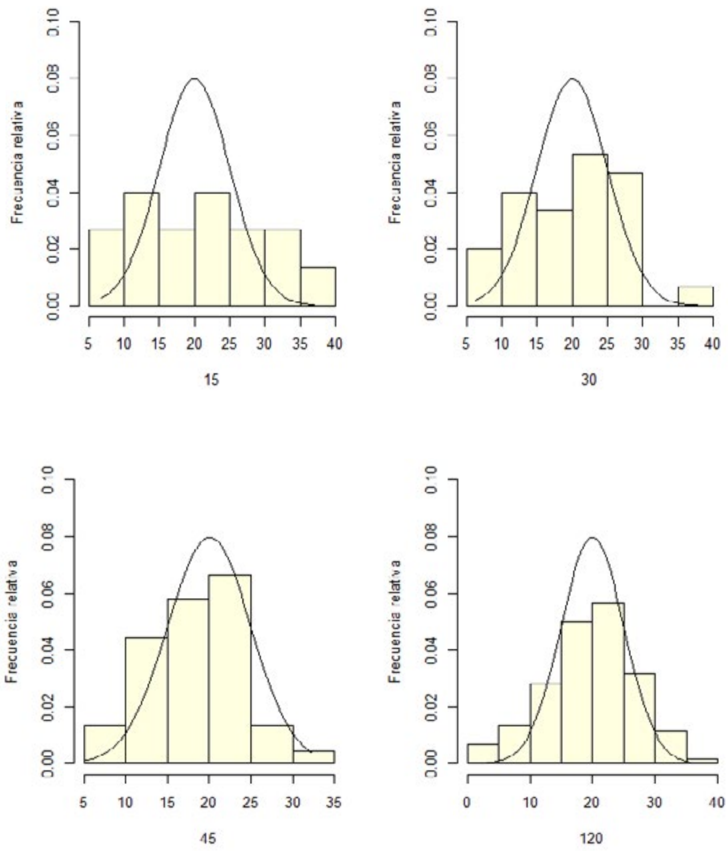
Por otra parte, se observa en la Grafica 3, el comportamiento del histograma de frecuencia, para diferentes tamaños de muestra, es evidente, para tamaños de muestra superiores, el comportamiento es en forma de campana de Gauss, distribución normal.

Curva operativa u ojiva de frecuencia en R

nombre<-c(dato[1],..,dato[n])	#### lectura de los datos
y<-plot.ecdf(nombre,main="distribución acumulada",-col.01line= "blue", verticals = TRUE,ylab = "frecuencia relativa acumulada", pch=20)	### crea el polígono de frecuencia

Histograma con diferente tamaño en R

```
set.seed(18)
N=c(15,30,45,120) #Vector que guarda los tamaños de muestra a comparar
par(mfrow=c(2,2))
for(i in 1:length(N)){
n=N[i]
x=round(rnorm(n,20,7),digits=2)
hist(x, probability=TRUE,main="Histograma de diferentes tamaños con Sturges", xlab=n, ylab="Frecuencia relati-
va",ylim=c(0,0.1), col="lightyellow")
curve(dnorm(x,20,5),from =min(x),to=max(x),add=T)
}
```



Gráfica 3
Histograma de frecuencia con diferente tamaño Sturges

⊙ **Fuente:** salida programa R

LAS GRÁFICAS DE BARRAS Y DE TORTA

Diagramas de barras

El diagrama de barras es aplicado cuando se manejan variables de tipo cualitativo, como también cuantitativos. En el plano cartesiano se presenta en el eje horizontal la característica o modalidad seleccionada y en el eje de las ordenadas o vertical se presenta la frecuencia relativa de cada una de las características o modalidad.

Una vez asignado las modalidades se edifican barras verticales de amplitud o base constantes, separadas una de otra, en donde la altura de la barra es proporcional a la frecuencia relativa o el conteo de cada una de las características o modalidades evaluadas en el estudio. El esquema del diagrama de barras requiere la organización de los datos en una tabla en donde se presente la modalidad y la frecuencia de ocurrencia de cada modalidad, como se muestra en la Tabla 10.

Tabla 10

Frecuencia de ocurrencia de las modalidades

Tipos de Modalidad	Frecuencia
Modalidad A	f_A
Modalidad B	f_B
$\vdots$	
Modalidad Z	f_Z

⊙ **Fuente:** elaboración por el autor

Ejercicio 1.5

Realice una diagrama de barras de la información suministrada de la Tabla 11, donde es presentado el número de juegos clasificados como excelente, bueno, regular y malo de un equipo en una temporada.

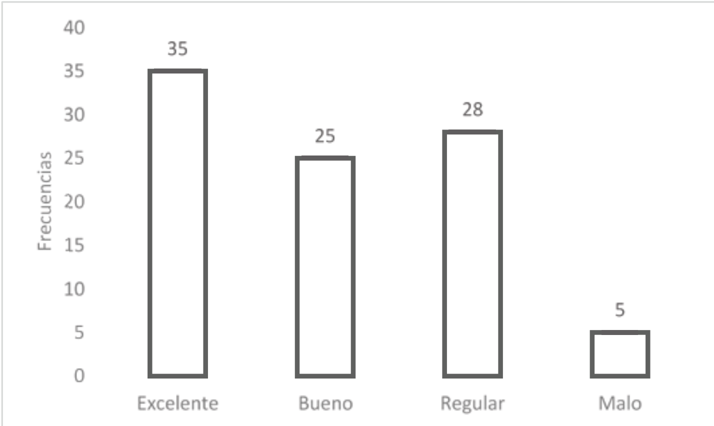
Tabla 11

Número de juegos clasificados de un equipo en una temporada

Tipo de juegos	Número de juegos
Excelente	35
Bueno	25
Regular	28
Malo	5

⊙ **Fuente:** elaboración por el autor

La Gráfica 4 muestra el comportamiento del número de juegos clasificados como excelente, bueno, regular y malo en la temporada del equipo de interés.



Gráfica 4

Diagrama de barras del tipo de juego del equipo

⊙ **Fuente:** elaboración por parte del autor, salida de R

Ejercicio 1.6

Un estudio de un equipo de béisbol es evaluado el número de errores durante una temporada de tres posiciones en el campo de juego. Los resultados Se presentan en la Tabla 12.

Tabla 12

Número de errores durante una temporada de tres posiciones en el campo de juego

Agente	Errores obtenidos en la temporada	Frecuencia relativa
Jardinero	4	4/13=0,307
Lanzador	5	5/13=0,384
Primera base	4	4/13=0,307
Total	13	

⊙ **Fuente:** elaboración por el Autor

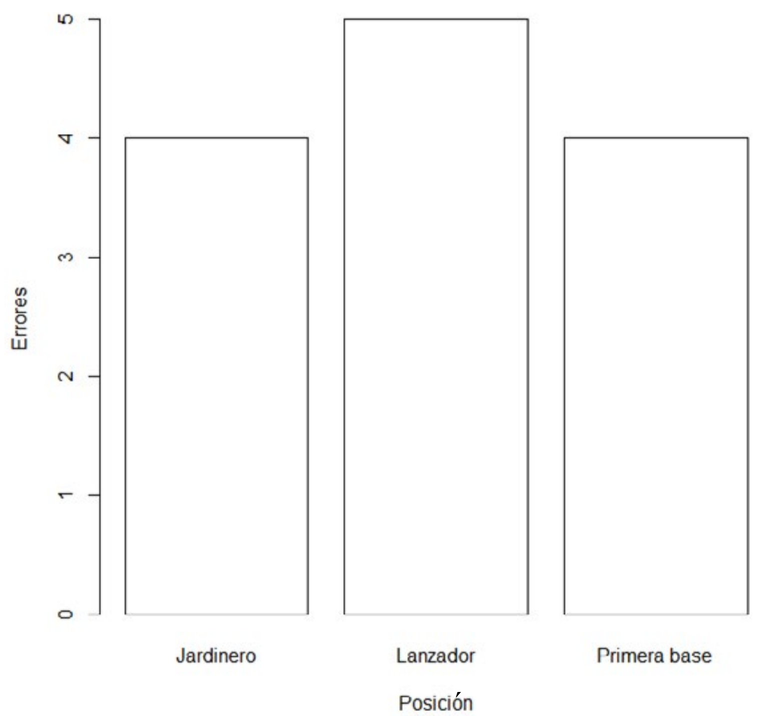
Diagrama de barras en R

nombre<- data.frame(nombreX=c("frame1","frame2","-frame3"), nombreY=c(conteoframe1,conteoframe2,conteo-frame3))	**#### lectura de los datos**
head(nombre)	### crea títulos de cabecera
barplot(nombre$nombreY,xlab ="nombreX",ylab="nombreY", names.arg = c("frame1","frame2","frame3"), col = c("color1", "color2", "color3"))	Elabora el diagrama de barras

Comando programa R para elaborar diagrama de barras

```
beisbol <- data.frame(Posicion=c("Jardinero","Lanzador","Primera base"), Errores=c(4,5,4))
head(beisbol) # así es como deben lucir los datos
barplot(beisbol$Errores,xlab ="Posicion",ylab="Errores", names.arg = c("Jardinero","Lanzador","Primera base"), col
= c("white", "white", "white"))
```

El diagrama de barras Gráfico 5, visualiza que la mayor frecuencia de errores en el juego de beisbol, lo presenta el lanzador con un valor de cinco.



Gráfica 5

Diagrama de barras de los jugadores de béisbol

⊙ **Fuente:** elaboración por el Autor.

Polígonos de frecuencia

Para la elaboración del polígono de frecuencias se presentan los puntos medios x_i en el eje de las abscisas de cada uno de los intervalos de clase y se unen mediante segmentos lineales, en el eje de las ordenadas se muestra las frecuencias relativas. Este diagrama también se puede elaborar, trazando líneas tomando el límite o borde superiores de cada uno de los intervalos de clase, Gráfico 6.

Ejercicio 1.7

La Tabla 13 presenta, en forma organizada, los tiempos obtenidos en un monoplaza en un velódromo. Elabore un polígono de frecuencia para describir el comportamiento de los tiempos.

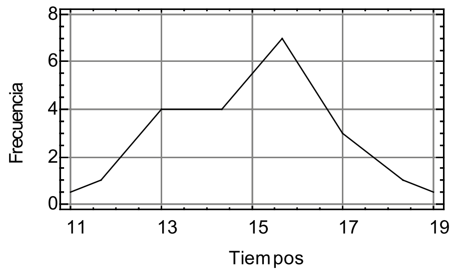
Tabla 13

Información de los tiempos de un velódromo

Intervalo de clase	Punto medio x_i	Frecuencia de clase f	Frecuencia relativa de clase f_r	Frecuencia acumulada F	Frecuencia relativa acumulada F_r
11,00-12,33	11,66	1	0,05	1	0,05
12,33-13,66	13,00	4	0,20	5	0,25
13,66-15,00	14,33	4	0,20	9	0,45
15,00-16,33	15,66	7	0,35	16	0,80
16,33-17,66	17,00	3	0,15	19	0,95
17,66-19,00	18,33	1	0,05	20	1,00

⊙**Fuente:** elaboración por el autor

Eligiendo los resultados de la columna frecuencia relativa acumulada, con respecto al intervalo de tiempo, el polígono de frecuencia obtenido, es presentado em la Gráfica 6.



Gráfica 6

Polígono de frecuencia de los tiempos de espera

⊙ **Fuente elaboración por el autor**

Ejercicio 1.8

En la Tabla 14 se presenta la distancia recorrida por un futbolista en 20 partidos de la Liga Colombiana de Fútbol.

Tabla 14

Información de la distancia recorrida por el jugador de fútbol en la Liga Colombiana

7,8	7,5	9,4	10,5
7,1	8,3	9,3	5,7
8,4	10,2	8,1	9,6
10,2	11,2	11,2	8,0
9,8	10,6	11,0	8,4

⊙ **Fuente:** elaboración por el autor

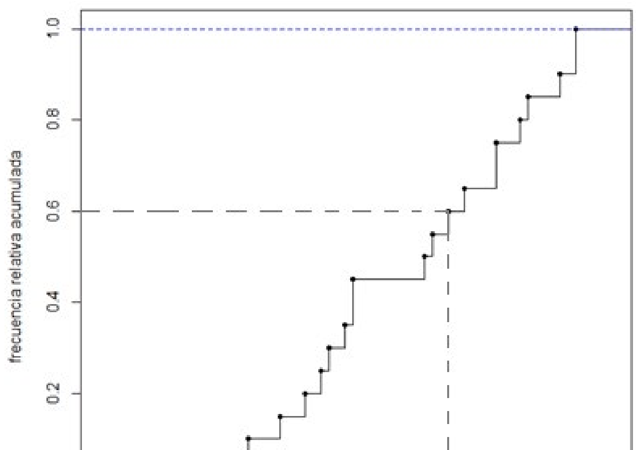
Histograma, polígono y curva operativa de frecuencia en R

nombre<-c(dato[1],..,dato[n])	#### lectura de los datos
y<-hist(nombre, breaks=c(0, 5, 10, 15, 20), col = "pink", border = 1, main = "distribución de nombre", xlab = "distancia" , ylab = "frecuencia")	##Elabora el histograma con opción de color
lines(c(min(y$breaks),y$mids,max(y$breaks)),c(0,y$counts,0),type="l")	
y<-plot.ecdf(nombre,main="distribución acumulada",col.01line= "blue", verticals = TRUE,ylab = "frecuencia relativa acumulada", pch=20)	### crea el polígono de frecuencia

Comandos en R Ejercicio 9

```
distancia <- c(7.8,7.1,8.4,10.2,9.8,9.4,9.3,8.1,11.2,11.0,7.5,8.3,10.2,11.2,10.6,10.5,5.7,9.6,8.0,8.4)
y<-hist(distancia, breaks=c(0, 5, 10, 15, 20), col = "pink", border = 1, main = "distribución de las distancia del jugador", xlab = "distancia" , ylab = "frecuencia")
lines(c(min(y$breaks),y$mids,max(y$breaks)),c(0,y$counts,0),type="l")
y<-plot.ecdf(distancia,main="distribución acumulada",col.01line= "blue", verticals = TRUE,ylab = "frecuencia relativa acumulada", pch=20)
```

El Gráfico 7, presenta el porcentaje acumulado de la distancia recorrida por el jugador; este diagrama permite evaluar el percentil obtenido dada una distancia recorrida, por ejemplo, este atleta el 60 % de las veces recorre distancias por encima de 9,5 kilómetros.



Gráfica 7
Polígono de frecuencia de la distancia recorrida por el jugador de fútbol en la Liga Colombiana, salida en R

Diagrama de torta o sector

Esta grafica también es conocida como diagrama de pastel, al igual que el diagrama de barras se aplica a variables cualitativas y cuantitativas. Su función es obtener porciones representativas de cada una de las modalidades o características evaluadas. Inicialmente el procedimiento para construir un diagrama de torta es obtener una conversión de cada frecuencia obtenidas en la modalidad o característica que es seleccionada. Esta modalidad es representada en un círculo de tal manera que el ángulo de cada sector o modalidad es proporcional a la frecuencia absoluta correspondiente. Esta conversión es realizada aplicando la siguiente formulación,

$$\theta = \frac{360^o}{n} \times f \qquad (3)$$

En donde n es el tamaño de la muestra y f es la frecuencia absoluta o de clase de la modalidad o características observada en la Tabla de frecuencia. Una vez determinado el ángulo de la modalidad o categoría, el siguiente paso es obtener cada uno de los trazos lineales del diagrama de torta o pastel, mediante el uso de un transportador como herramienta para mejorar la precisión.

Ejercicio 1.9

En la Tabla 15 se muestra la información del número de no conformidades o fallas presente en un juego de basquetbol.

Tabla 15

Información del número de no conformidades o fallas presente en un juego de basquetbol

Modalidad	Número de fallas	Ángulo
Triples	5	45
Encesta tiro Libre	17	153
Reversos	2	18
Mates	12	108
Rebote	4	36
Total	40	

Los cálculos de cada uno de los sectores de la torta, basada en la ecuación (3) del número de no conformidades o fallas presente en un juego de basquetbol, son los siguientes,

$\theta_1 = \frac{360^\circ}{40} \times 5 = 45$, $\theta_2 = \frac{360^\circ}{40} \times 17 = 153$, $\theta_3 = \frac{360^\circ}{40} \times 2 = 18$, $\theta_4 = \frac{360^\circ}{40} \times 12 = 108$ y

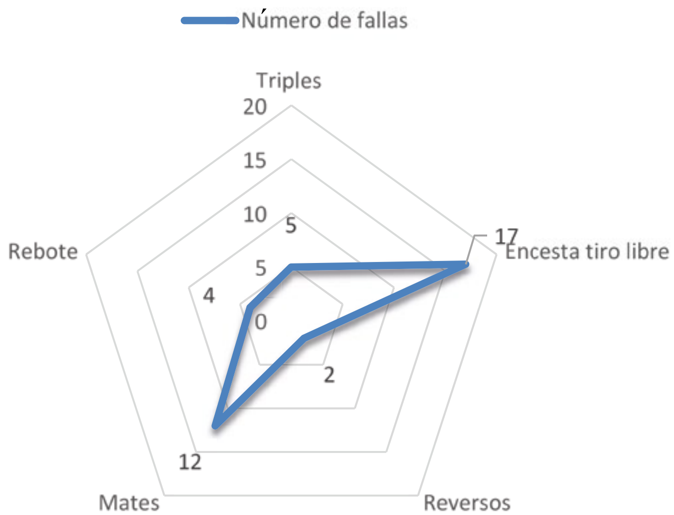
$\theta_5 = \frac{360^\circ}{40} \times 4 = 36$

La Gráfica 8 muestra la partición realizada basada en los cálculos de los ángulos, en este caso representado por los sectores.

Gráfica 8
Diagrama de torta de las fallas presentadas en un juego de basquetbol

⊙ **Fuente:** elaboración por el Autor

El diagrama radial en la Gráfica 9 presenta el conteo de las fallas en el juego de basquetbol, en donde se evidencia un número apreciable de errores en las encesta, en particular el conteo de la encesta de tiro libre enumera el mayor número de errores con 17.



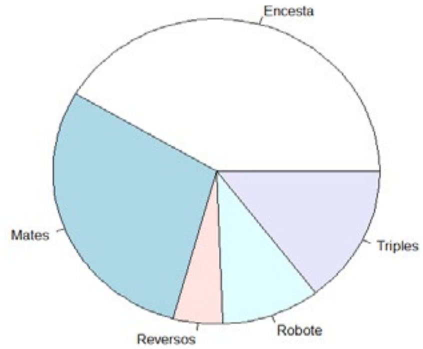
Gráfica 9
Diagrama radial del número de fallas presentadas en un juego de basquetbol

⊙ **Fuente:** elaboración por el Autor.

Histograma de Frecuencia en R

nombre<-c("frame", "frame", "frame",…, "frame")	#### nombre de cada uno las características o eventos ocurridos en el orden de aparición
Pie(table(nombre))	### crea el diagrama de torta

Comando en R diagrama de torta
modalidad<-c("Triples", "Triples", "Triples", "Triples", "Triples", "Triples", "Encesta", "Encesta", "Encesta","Encesta", "Encesta", "Encesta", "Encesta","Encesta", "Encesta", "Encesta", "Encesta", "Encesta", "Encesta", "Encesta", "Encesta", "Encesta", "Encesta", "Reversos", "Reversos", "Mates", "Mates", "Mates", "Mates", "Mates", "Mates", "Mates", "Mates", "Mates", "Mates", "Mates", "Mates", "Mates", "Robote", "Robote", "Robote", "Robote") pie(table(modalidad))

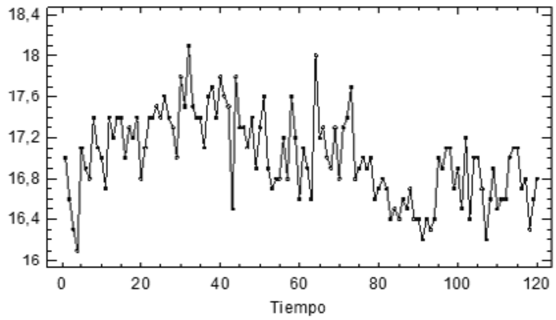


Gráfica 10
Diagrama de torta de las fallas salida en R

⊙ **Fuente:** elaboración por el Autor.

Diagramas de seguimiento

Es el seguimiento en el tiempo, de una característica o variable aleatoria de un fenómeno estocástico, con el objetivo de encontrar alguna variación significativa de la variable evaluada. Por ejemplo, los tiempos realizados por un corredor de fórmula 1, de la temporada del Campeonato Mundial organizada por la Federación Internacional del Automóvil (FIA), en un circuito en particular, se presenta en la Grafica 11:



Gráfica 11
Diagrama de seguimiento o serie de tiempo

⊙ **Fuente:** elaboración por el Autor.

Diagrama de seguimiento en R

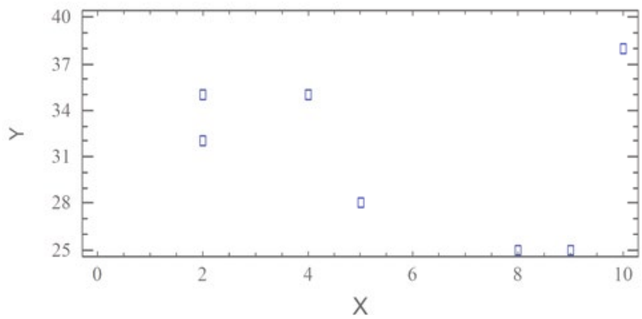
nombre<-c(dato[1],..,dato[n])	
plot(nombre)	
lines(nombre)	

ESTADÍSTICA BIVARIADA

Algunas veces se necesita comparar dos o más variables con el objetivo de encontrar relación entre ellas. Es una manera de encontrar una mejor respuesta a situaciones que son causadas por la relación existente entre ellas.

Gráfica de correlación

Este diagrama se aplica al realizar un análisis de la relación existente, entre dos o más variables de una información seleccionada; el objetivo es encontrar las causas de esta asociación. Existen varios métodos para determinar si dos o más variables se encuentran asociadas. El más simple de estos métodos, es la presentación gráfica y el cálculo del coeficiente de correlación r que mide la intensidad de esta asociación. El diagrama de correlación muestra la asociación entre la variable X que se considera como variable independiente y la variable Y se considera la variable dependiente, Gráfica 12.



Gráfica 12
Correlación muestra la asociación entre la variable X y la variable Y

Fuente: elaboración por el Autor

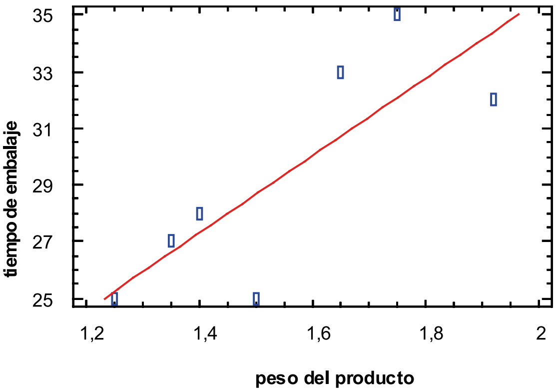
Ejercicio 1.10

La Tabla 16 muestra los datos del peso y el tiempo registrado en un encuentro deportivo. mediante un diagrama de correlación determine si existe alguna relación entre las variables el peso y el tiempo.

Tabla 16
Peso y el tiempo registrado en un encuentro deportivo

Peso	Tiempo
25.0	1.25
35.0	1.75
28.0	1.40
27.0	1.35
25.0	1.25
33.0	1.65
32.0	1.60

El diagrama de correlación muestra una alta asociación entre las variables peso del producto y el tiempo de embalaje, como se presenta en la Gráfica 13, Grafica 14.



Gráfica 13
Diagrama de barras del tipo de juego del equipo

⊙ **Fuente elaboración realizada por el Autor**

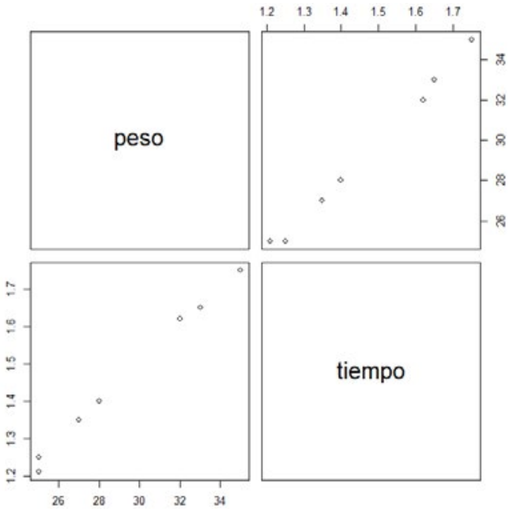
Diagrama de correlación en R

nombre<-c(dato[1],..,dato[n])	**#### lectura de los datos**
hist(nombre)	### crea el histograma de frecuencia

Comando en R del diagrama de correlación

```
peso<-c(25.0, 35.0, 28.0, 27.0, 25.0, 33.0, 32.0)
tiempo<-c(1.25, 1.75,            1.40,        1.35,        1.21,        1.65,        1.62)
pairs (peso ~ tiempo) #permite elaborar un plots de correlación

library (PerformanceAnalytics)
natacion <- data.frame(peso, tiempo)
chart.Correlation(natacion)
cor(peso, tiempo)
```



Gráfica 14
Diagrama de correlación . Salida en R.

RESUMEN

En el momento de seleccionar una muestra de una población es necesario organizarlos mediante tablas de frecuencia o distribución empírica, permiten de esta manera extraer en forma sencilla la información del fenómeno.

La tabla de frecuencia o distribución empírica genera gráficas útiles para un estudio estadístico, tal es el caso del histograma de frecuencias, el polígono de frecuencia y el diagrama de ojiva o curva operativa. Con estos diagramas verificamos la manera en que la información se comporta con respecto a una distribución de referencia.

Existen varios gráficos importantes aplicados en la estadística descriptiva es el caso del diagrama de torta, sectoriza en porcentaje la incidencia de la variable en el fenómeno de estudio. Los diagramas de barras son similares a los gráficos de sectores, donde la altura de cada barra proporciona la frecuencia o porcentaje de incidencia de la variable. En términos generales los diagramas descriptivos son herramientas sencillas, ofrecen amplias posibilidades para analizar la información y algunas veces pueden ser útiles para analizar fenómenos mucho más complejos.

Unidad 2

Medidas de localización, variabilidad y forma

INTRODUCCIÓN

Las medidas de localización, variabilidad y de forma son características evaluadas a partir de la información suministrada en la muestra. Estas características permiten conocer en una forma sencilla el comportamiento de la información y de esta manera se convierte en una herramienta no menos importante en la toma de decisiones. En el primer ítem de esta unidad se plantean los conceptos básicos de las medidas descriptivas. En segundo lugar, se desarrollan las medidas de posición o de localización; entre estas medidas, las más destacadas son la media, la mediana, la moda, la media armónica, la media geométrica. Se presenta consecutivamente la importancia del cálculo de la medida de variabilidad o dispersión en el análisis de la información, específicamente el cálculo de la varianza y desviación estándar en el conjunto de datos, como medidas de dispersión más acentuadas. En el siguiente ítem se encuentra las medidas de forma que permite conocer el comportamiento del fenómeno mediante la evaluación de dos características importantes, el coeficiente de asimetría y el de curtosis. Finalmente se desarrolla el cálculo del coeficiente de correlación cuando se evalúa la relación existe entre dos variables.

MEDIDAS DE LOCALIZACIÓN O DE POSICIÓN

Entre las medidas de localización más aplicada en estadística encontramos la media aritmética o promedio.

La media

La media aritmética representada por $\bar{x}$, se obtiene de la razón de la suma de los valores de la variable y el número total de observaciones, su formulación es la siguiente. El caso de observaciones no organizadas en tabla de frecuencia o distribución empírica (datos no agrupados),

$$\bar{\bar{x}} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{4}$$

La organización de los datos en una tabla de frecuencia, la media es evaluada de la siguiente forma,

$$\bar{\bar{x}} = \sum_{i=1}^{k} \frac{f_i}{n} x_i = \sum_{i=1}^{k} f_r x_i \tag{5}$$

donde x_i es el punto medio, f_i es la frecuencia de clase o absoluta, k es el número de intervalos de clase y f_r es la frecuencia relativa de clase.

Ejercicio 2.1

Determine la media aritmética o promedio de la información del tiempo en segundos que es ejecutado por un nadador, como se observa en la Tabla 17.

Tabla 17

Tiempo en segundos de un nadador

Tiempo		
22	21	18
19	32	33
20	16	35
15	16	17

⊙ **Fuente:** elaboración por el Autor

Tomando la formulación de la media aritmética tenemos que,

$$\bar{x} = \sum_{i=1}^{12} \frac{(22 + 21 + \cdots + 16 + 17)}{12} = \frac{264}{12} = 22$$

La media del nadador es de 22 segundos.

Ejercicio 2.2

En la Tabla 18 se presenta la frecuencia del contenido de colesterol de un equipo de fútbol. Determine la media aritmética del contenido de colesterol.

Tabla 18
La tabla de frecuencia o distribución empírica del contenido de colesterol de un equipo de futbol

Intervalo	Punto medio x_i	Frecuencia f	Frecuencia relativo f_r	Frecuencia acumulada F	Frecuencia relativa acumulada F_r
197.11-197.77	197.44	1	0.0278	1	0.0278
197.77-198.44	198.11	2	0.0556	3	0.0833
198.44-199.11	198.77	6	0.1667	9	0.2500
199.11-199.77	199.44	4	0.1111	13	0.3611
199.77-200.44	200.11	12	0.3333	25	0.6944
200.44-201.11	200.77	7	0.1944	32	0.8889
201.11-201.78	201.44	4	0.1112	36	1.0000

⊙ **Fuente:** elaboración por el autor

En este caso la media aritmética, se calcula de la siguiente manera

$$\bar{x} = \sum_{i=1}^{k} (0.0278) \times 197.44 + (0.0556) \times 198.11 + \cdots + (0.1112) \times 201.44 = 199.9229$$

En este caso el promedio del contenido de colesterol es de 199.9229

La media armónica

Está definida como la razón del número de observaciones y la suma de los recíprocos de cada uno de las observaciones o puntos medios por su frecuencia, se representa mediante la letra H. Entre las características de la media armónica se encuentran: 1. Es sensible a los valores extremos. 2. Intervienen todos los valores de la información. 3. Es un valor único. Para los datos no organizados la formulación aplicada es la siguiente,

$$\bar{X}_A = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$$

donde n es el tamaño de la muestra, x_i es el valor de cada una de las variables.

Para los datos organizados en una Tabla de frecuencia la formulación aplicada es la siguiente,

$$\bar{X}_A = \frac{n}{\frac{1}{x_1}f_1+\frac{1}{x_2}f_2+\cdots+\frac{1}{x_k}f_k} = \frac{n}{\sum_{i=1}^{k}\frac{1}{x_i}} \quad (6)$$

donde n es el tamaño de la muestra, x_i es el valor de cada uno de los puntos medios y k es el número de intervalos de clase o la cantidad de renglones de la Tabla de frecuencia.

Ejercicio 2.3

A continuación, se presentan los valores de glicemia obtenidos de los aspirantes a formar parte del equipo ciclístico de una universidad, Tabla 19. El investigador desea evaluar como medida de localización representativa de la muestra la media armónica.

Tabla 19

Contenido glicémico de los aspirantes a formar parte del equipo ciclístico universitario

199,83	199,89	198,92	200,04	199,90	198,73
200,85	199,50	200,40	200,71	199,09	200,13
199,27	197,11	200,60	199,88	201,78	198,15
198,47	200,69	199,61	198,73	200,36	197,93
202,81	199,59	199,70	201,16	199,03	200,51
200,37	198,78	200,28	199,96	201,57	200,81

⊙ **Fuente:** elaboración por el Autor

El cálculo de la media armónica en la ecuación 6 es la siguiente,

$$\bar{X}_A = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}} = \frac{36}{\frac{1}{199.83}+\frac{1}{199.89}+\cdots+\frac{1}{200.81}} = 199.85$$

Histograma de Frecuencia en R

nombre<-c(dato[1],..,dato[n])	**#### lectura de los datos**
hist(nombre)	### crea el histograma de frecuencia

Comando en R, evaluación de la media armónica

```
glicemia<-
c(199.83,199.89,198.92,200.04,199.90,198.73,200.85,199.50,200.40,200.71,199.09,200.13,199.27,197.11,200.60,199
.88,201.78,198.15,198.47,200.69,199.61,198.73,200.36,197.93,202.81,199.59,199.70,201.16,199.03,200.51,200.37,1
98.78,200.28,199.96,201.57,200.81)
armonic<-1/mean(1/glicemia)
armonic
```

Ejercicio 2.4

En la Tabla 20, muestra la frecuencia obtenida para el tiempo de actividades específicas realizada por uno ciclistas en su entrenamiento. Se desea establecer la media armónica obtenida en el tiempo de operación.

Tabla 20
Tiempo en horas de entrenamiento de los ciclistas

Tiempo en horas de entrenamiento	Frecuencia f
6	4
9	2
12	3
15	1
Total	10

⊙ **Fuente:** elaboración por el Autor

La media armónica se calcula aplicando la siguiente formulación,

$$\bar{X}_A = \frac{n}{\frac{1}{x_1}f_1 + \frac{1}{x_2}f_2 + \cdots + \frac{1}{x_k}f_k} = \frac{10}{\frac{1}{6}(4) + \frac{1}{9}(2) + \cdots + \frac{1}{15}(1)} = 8.29$$

La mediana

Es un valor que separa el conjunto de datos, en porciones iguales, es decir los datos inferiores a este valor representarán el 50 % de la información y los superiores a este valor representarán el 50 % restante. Para calcular la mediana en los datos no organizados, se hace necesario el ordenamiento de la información de menor a mayor con el objetivo de encontrar el orden o posición de cada valor de la variable en el conjunto. Las formulaciones aplicadas para los datos no organizados son las siguientes,

$$\bar{X}_M = \begin{cases} X_{\frac{n}{2}+\frac{1}{2}} & \text{(datos impar)} \\ \\ \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} & \text{(datos par)} \end{cases} \tag{7}$$

Ejercicio 2.5

La siguiente información, Tabla 21, es el peso en gramos de proteína diaria, ofrecido a un atleta de alto rendimiento, se desea determinar el valor de la mediana.

Tabla 21
Peso en gramos de la proteína ofrecida a un atleta de alto rendimiento

Peso de la proteína		
123	156	205
125	164	207
131	164	222
155	194	256

⊙ **Fuente:** elaboración por el Autor

Los datos de la anterior tabla se encuentran ordenados en forma descendente, lo cual facilita encontrar el orden o posición del valor de cada variable. Como se trata de un conjunto par de datos, implica, según las formulaciones anteriores, que existen dos valores centrales. Estos valores se ubican en las siguientes posiciones,

$$\bar{X}_M = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} = \frac{X_6 + X_7}{2} = \frac{164 + 164}{2} = 164$$

Cuando los datos se encuentran organizados en una tabla de frecuencia, el cálculo de la mediana se realiza mediante la siguiente formulación.

$$\bar{X}_M = L_{i-1} + \frac{\left(\frac{n}{2} - F_{i-1}\right)}{f_i}\left(L_i - L_{i-1}\right) \tag{8}$$

donde n es el tamaño de la muestra, L_i y L_{i-1} es el valor del límite superior e inferior del intervalo de clase, F_{i-1} es la frecuencia acumulada anterior al intervalo donde se encuentra la mediana, f_i frecuencia de clase del intervalo donde se ubica la mediana.

Cálculo de la mediana en R

nombre<-c(dato[1],..,dato[n])	#### lectura de los datos
median(nombre)	### crea el histograma de frecuencia

Comando en R, evaluación de la mediana

```
peso<-c(123, 156, 205, 125, 164, 207, 131, 164, 222, 155, 194,256)
median(peso)
```

Ejercicio 2.6

En Tabla 22, se presenta la información de conteo de alumnos de fútbol americano, registro obtenido basado en el peso obtenido.

Tabla 22

Conteo del peso de alumnos de un equipo de futbol americano

Intervalo	Punto medio x_i	Frecuencia f	Frecuencia relativo f_r	Frecuencia acumulada F	Frecuencia relativa acumulada F_r
110-120	60	20	0,18	20	0,18
120-130	125	50	0,45	70	0,63
130-140	135	40	0,37	110	1,00
Total		100			

⊙ **Fuente:** elaboración por el Autor

Se tiene que el valor central de la información se ubica en la posición $\frac{n}{2}$ que en este caso es 55, ya que la muestra n es de 110 (las sumas de la frecuencia absolutas o de clase). El intervalo de clase que posee esta posición es el segundo cuyo límite inferior L_{i-1} es 120 y su frecuencia de clase f_i es 50. La amplitud es la diferencia de los límites, en este caso $(L_i - L_{i-1})$ es 10. El valor de la frecuencia acumulada anterior F_{i-1} es en este caso 20. Por tanto, la mediana de los datos organizados en la anterior Tabla de frecuencia es,

$$\bar{X}_M = 120 + \frac{(50-20)}{50}10 = 126$$

La media geométrica

Está definida como la raíz enésima del producto de los n valores de las observaciones,

$$\bar{X}_G = \sqrt[n]{X_1 \times X_2 \times X_3 \times \ldots \times X_n} \qquad (9)$$

Ejercicio 2.7

El tiempo obtenido de un circuito de fórmula uno se presenta en la Tabla 23. Calcule la media geométrica de la información.

Tabla 23

Tiempo de un circuito de fórmula uno

Tiempo del circuito		
1,24	2,01	1,13
1,40	1,70	1,50
0,8	1,56	1,23
1,3	0,92	1,10

⊙ **Fuente:** elaboración por el Autor

La media geométrica de la ecuación (9), indica que el tiempo obtenido en el circuito es el siguiente,

$$\bar{X}_G = \sqrt[12]{1.24 \times 2.01 \times 1.13 \times \ldots \times 1.10} = 1.285$$

En el caso de los datos organizados, la media geométrica está definida como la raíz enésima de los productos de las observaciones elevada por su frecuencia.

$$\bar{X}_G = \sqrt[n]{X_1^{f_1} \times X_2^{f_2} \times X_3^{f_3} \times \ldots \times X_k^{f_k}} \qquad (10)$$

donde k es la frecuencia de clase en cada intervalo y $n = \sum_{i=1}^{k} f_i$ es la suma de las frecuencias de clase.

Histograma de Frecuencia en R

nombre<-c(dato[1],..,dato[n])	**#### lectura de los datos**
geometric<-function(nombre) exp(sum(log(nombre))/ length(nombre))	### crea el histograma de frecuencia
geometric(nombre)	

Comando en R. evaluación de la media geométrica
circuito<-c(1.24, 2.01, 1.13, 1.40, 1.70, 1.50, 0.8, 1.56, 1.23, 1.3, 0.92, 1.10) geometric<-function(circuito) exp(sum(log(circuito))/length(circuito)) geometric(circuito)

Ejercicio 2.8

El número de dianas obtenidas en los últimos 12 partidos de un equipo de fútbol de una entidad universitaria se presenta en la Tabla 24. Obtenga la media geométrica del número de goles obtenidas por este quipo.

Tabla 24

Número de dianas de un equipo de fútbol

Número de dianas	Frecuencia f
5	2
3	5
1	2
4	3
Total	12

⊙ **Fuente:** elaboración por el Autor

La media geométrica, ecuación (10), para el número de dianas se determina mediante la siguiente formulación,

$$\bar{X}_G = \sqrt[12]{5^2 \times 3^5 \times 1^2 \times 4^3} = 1.622$$

MEDIDAS DE VARIABILIDAD O DISPERSIÓN

La varianza.

Está definida como la suma de la diferencia al cuadrado, del valor de la observación con respecto a la media aritmética. El caso de datos no organizados, la varianza se determina mediante la siguiente formulación,

$$S^2 = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n} = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n} \quad (11)$$

donde x_i es el valor de la variable y $\bar{x}$ es la media o promedio aritmético.

La siguiente formulación es aplicada para los datos organizados,

$$S^2 = \frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + (x_3 - \bar{x})^2 f_3 + \cdots + (x_k - \bar{x})^2 f_k}{n} = \sum_{i=1}^{k} \frac{(x_i - \bar{x})^2 f_i}{n} \qquad (12)$$

Entre las características de la varianza tenemos: a) Es un estadístico sensible a las observaciones extremas, b) La varianza no está expresada en las mismas unidades de los datos, por lo que es conveniente calcular la raíz de esta característica; esta raíz se conoce como desviación estándar o típica.

La desviación estándar

Está definida como la raíz de la varianza, es aplicada con mucha frecuencia porque se expresa con las mismas unidades de las observaciones. La formulación aplicada para los datos no organizados es la siguiente

$$s = \sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n-1}} = \sqrt{\sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n-1}} \qquad (13)$$

datos agrupados u organizados,

$$s = \sqrt{\frac{(x_1-\bar{x})^2 f_1+(x_2-\bar{x})^2 f_2+(x_3-\bar{x})^2 f_3+\cdots+(x_k-\bar{x})^2 f_k}{n}} = \sqrt{\sum_{i=1}^{k} \frac{(x_i-\bar{x})^2 f_i}{n}} \qquad (14)$$

Ejercicio 2.9

En la siguiente información se desea conocer la varianza y la desviación estándar del número de inscritos mensualmente en las actividades deportivas de un colegio de secundaria de la ciudad, Tabla 25.

Tabla 25
Número de inscritos en las actividades deportivas

	Número de inscritos	
113	175	114
145	133	176
103	204	143

⊙ **Fuente:** elaboración por el Autor

Inicialmente se realiza el cálculo de la media aritmética del número de marcadores utilizados,

$$\bar{x} = \sum_{i=1}^{9} \frac{(113 + 175 + 114 + \cdots + 204 + 143)}{9} = 145.1$$

Posteriormente utilizando las ecuaciones (13), (14) se obtiene la varianza y la desviación estándar.

$$S^2 = \frac{(113 - 145.1)^2 + (175 - 145.1)^2 + (114 - 145.1)^2 + \cdots + (143 - 145.1)^2}{9 - 1} = \frac{10393.75}{8}$$
$$= 1154{,}86$$

La desviación estándar se define como,

$$s = \sqrt{\frac{(113 - 145.1)^2 + (175 - 145.1)^2 + (114 - 145.1)^2 + \cdots + (143 - 145.1)^2}{8}} = \sqrt{\frac{10393.75}{8}}$$
$$= \sqrt{1154{,}86} = 33.98$$

Cálculo de la varianza y desviación estándar en R

nombre<-c(dato[1],..,dato[n])	#### lectura de los datos
mean(nombre) **var(nombre)** **sd(nombre)**	### calcula los estadísticos de la media, varianza y desviación estándar respectivamente.

Comando en R, evaluación de la media, la varianza y la desviación estándar
inscritos<-c(113, 175, 114, 145, 133, 176, 103, 204, 143) mean(inscritos) var(inscritos) sd(inscritos)

MEDIDAS DE FORMA O APUNTAMIENTO

Coeficiente de asimetría

Estas medidas permiten comprobar si una distribución de frecuencia posee un comportamiento simétrico o asimétrico. Entre las fórmulas más aplicadas tenemos:

a. El índice de simetría de Pearson, evaluada mediante la siguiente formula,

$$\text{Ske} = \frac{\left(\frac{\sum_{i=1}^{n} x_i}{n} - \text{Mo}\right)}{s} \qquad (15)$$

donde $\bar{x}$ es la media o promedio aritmético, $\bar{X}_M$ es la moda y S es la desviación del conjunto de datos.

b. El índice de simetría de Fisher calculada mediante la siguiente ecuación,

$$\mathrm{Ske} = \frac{1}{n}\sum_{i=1}^{k}\frac{\left(x_i - \frac{\sum_{i=1}^{n} x_i}{n}\right)^3 f_i}{s^3} \qquad (16)$$

donde $\bar{x}$ es la media o promedio aritmético, f_i es la frecuencia de clase o absoluta, S^3 es la desviación elevada al cubo del conjunto de datos; toma el valor de cero cuando la distribución es simétrica, positivo cuando su comportamiento es asimétrico a la derecha y negativa cuando existe asimetría a la izquierda.

Ejercicio 2.10

Los datos que a continuación se presentan: el número de fallas de un automóvil en una competencia de Rally basado en los rangos de kilómetros recorridos, La Tabla 26 establece el comportamiento de la información con el objetivo de planear un mejoramiento del servicio de mantenimiento. Determine el índice de simetría aplicando el método de Fisher.

Tabla 26

Fallas de un automóvil por kilómetros recorridos en competencias de Rally

Intervalo	Punto medio x_i	Frecuencia f	Frecuencia relativo f_r	Frecuencia acumulada F	Frecuencia relativa acumulada F_r
220-320	270	4	0.33	4	0.33
320-420	370	5	0.42	9	0.75
420-520	470	3	0.25	12	1.00

⊙ **Fuente:** elaboración por el Autor

Inicialmente se determina la media aritmética de la información suministrada de la Tabla de frecuencia,

$$\bar{x} = \sum_{i=1}^{3}(x_i)(f_r) = (0.33)(270) + (0.42)(370) + (0.25)(470) = 362$$

la varianza y la desviación estándar,

$$S^2 = \sum_{i=1}^{k}\frac{(x_i - \bar{x})^2 f_i}{n} = \frac{(270-362)^2(4) + (370-362)^2(5) + (470-362)^2(3)}{12} = 5764$$

$$s = \sqrt{\sum_{i=1}^{k} \frac{(x_i - \bar{x})^2 f_i}{n}} = \sqrt{\frac{(270-362)^2(4) + (370-362)^2(5) + (470-362)^2(3)}{12}} = \sqrt{5764} = 75.92$$

La ecuación (16) permite evaluar el índice de asimetría de Fisher en la siguiente forma.

$$Ske = \frac{1}{n}\sum_{i=1}^{k} \frac{(x_i - \bar{x})^3 f_i}{s^3} = \frac{1}{12}\frac{(270-362)^3(4) + (370-362)^3(5) + (470-362)^3(3)}{75.92^3} = 0.12$$

Ejercicio 2.11

Determine el coeficiente de asimetría y la curtosis del Ejercicio 19, del número de inscritos en las actividades deportivas.

Coeficiente de asimetría en R

nombre<-c(dato[1],..,dato[n])	**#### lectura de los datos**
skewness(nombre)	### cálcula el valor del coeficiente de asimetría.

Comando en R, evaluación del coeficiente de asimetría Ejercicio 19

```
library(moments)
inscritos <-c(113, 175, 114, 145, 133, 176, 103, 204, 143)
skewness (inscritos) # el valor de la asimetría de los datos de la variable inscritos
```

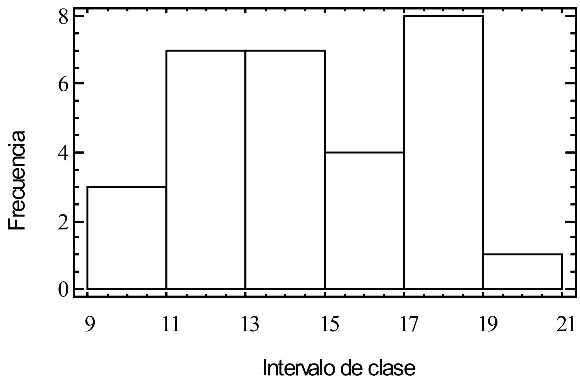
Programa para el cálculo del coeficiente de asimetría y la curtosis en R

nombre<-c(dato[1],..,dato[n])	#### lectura de los datos
m<-mean(nombre) md<-median(nombre) n<-length(nombre) s<-sd(nombre) skew<-sum(nombre-m)^3/n/s^3) kurt<-sum(nombre-m)^4/n/s^4)-3 skew kurt	### calcula el coeficiente de asimetría y la curtosis.

```
inscritos <-c(113, 175, 114, 145, 133, 176, 103, 204, 143)
m <-mean(inscritos)
md<-median(inscritos)
n <-length(inscritos)
s <-sd(inscritos)
skew <-sum((inscritos -m)^3/n/s^3)
kurt <-sum((inscritos -m)^4/n/s^4) - 3
skew
kurt
```

Ejercicio 22

El siguiente histograma representa el comportamiento del porcentaje de grasa de algunos atletas, ver Gráfica 15. Se desea determinar el valor del coeficiente de asimetría de la información suministrada en el diagrama.



Gráfica 15

Diagrama de barras del porcentaje de grasa de un grupo de atletas

⊙ **Fuente:** elaboración por el Autor

La información del histograma de frecuencia elaborado anteriormente es resumida en la Tabla 27,

Tabla 27

Porcentaje de grasa de un grupo de atletas organizada en una Tabla de frecuencia

Intervalo	Punto medio x_i	Frecuencia f	Frecuencia relativo f_r	Frecuencia acumulada F	Frecuencia relativa acumulada F_r
09-11	10	3	0.10	3	0.10
11-13	12	7	0.02	10	0.33
13-15	14	7	0.02	17	0.56
15-17	16	4	0.13	21	0.70
17-19	18	8	0.26	29	0.96
19-21	20	1	0.03	30	1.00
Total		30			

⊙ **Fuente:** elaboración por el Autor

El resultado de la media aritmética de la información es,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{3}(x_i)(f_i) = \frac{1}{30}\{(10)(3) + (12)(7) + (14)(8) + \cdots + (20)(1)\} = 14.66$$

La desviación estándar es,

$$s = \sqrt{\sum_{i=1}^{k}\frac{(x_i - \bar{x})^2 f_i}{n}} = \sqrt{\frac{(10 - 14.66)^2(3) + (12 - 14.66)^2(7) + \cdots + (20 - 14.66)^2(1)}{30}} = \sqrt{7.98}$$
$$= 2.826$$

El índice de asimetría de la ecuación (17) es,

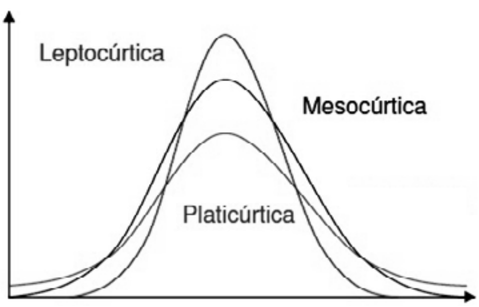
$$Ske = \frac{1}{n}\sum_{i=1}^{k}\frac{(x_i - \bar{x})^3 f_i}{s^3} = \frac{1}{30}\frac{(10 - 14.66)^3(3) + (12 - 14.66)^3(7) + \cdots + (20 - 14.66)^3(1)}{2.826^3} = 0.033$$

La curtosis

La curtosis Permite establecer el nivel de concentración de la información, como también el nivel apuntamiento que verifique el tipo de distribución de la información. Para calcular este coeficiente se utiliza la siguiente formulación,

$$Curt = \frac{\frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^4(f_i)}{\left[\frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2(f_i)\right]^2} - 3 \qquad (17)$$

Cuando toma un valor cercano a cero (0) la distribución es Mesocúrtica, en forma de campana (se asume una distribución Normal). Un valor positivo indica que la distribución es Leptocúrtica y cuando su valor es negativo su distribución es Platicúrtica; estos comportamientos son presentados en la Gráfica 16.



Gráfica 16
Diagramas distribución es Mesocúrtica, Leptocúrtica y Platicúrtica

⊙ **Fuente:** elaboración por el Autor

COEFICIENTE DE CORRELACIÓN.

Este coeficiente mide el grado de correlación de dos variables. El indicador más aplicado para encontrar una relación lineal entre las variables es la estadística de Newman Pearson o coeficiente de correlación. El coeficiente de correlación está definido de la siguiente forma,

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}} \qquad (18)$$

donde n es el número de parejas de datos, y es la variable respuesta o variable dependiente y x es la variable independiente.

Este valor r esta comprendido en el rango de valores de menos uno (-1) a uno (1). Un valor r cercano a uno positivo, implica una alta correlación positiva y un valor cercano a menos uno (-1) una alta correlación negativa. Si el valor del coeficiente de correlación es un valor cercano a cero (0) indica que las variables se consideran independientes.

Ejercicio 22

A continuación, se presenta en la Tabla 28, la muestra de siete datos de peso levantado y el tiempo registrado en un equipo de levantamiento de pesas. Aplicando el coeficiente de correlación r determine si existe una correlación entre el peso y el tiempo.

Tabla 28

Peso levantado y el tiempo establecido

Peso levantado	Tiempo
25.0	1.25
35.0	1.75
28.0	1.40
27.0	1.35
25.0	1.25
33.0	1.65
32.0	1.60

⊙ **Fuente:** elaboración por el Autor

Los cálculos de las sumas de cuadrado, para las diferentes variables de peso y tiempo en el levantamiento de pesas, son presentados en la Tabla 29.

Tabla 29
Cálculo de sumas de cuadrado del peso y tiempo de levantamiento

Peso del producto X	Tiempo de embalaje Y	X^2	Y^2	XY
25.0	1.25	625	21,56	31,25
35.0	1.75	1225	3,06	61,25
28.0	1.40	784	1,96	39,20
27.0	1.35	729	1,82	36,45
25.0	1.25	625	1,56	31,25
33.0	1.65	1089	2,72	54,45
32.0	1.60	1024	2,56	51,20
$\sum x_i = 205$	$\sum y_i = 10.82$	$\sum x_i^2 = 6101$	$\sum y_i^2 = 17.06$	$\sum x_i y_i = 321.54$

⊙ **Fuente:** elaboración por el Autor

Reemplazando en la ecuación (19) del coeficiente de correlación de Pearson tenemos,

$$r = \frac{7(321.54) - (205) \times (10.82)}{\sqrt{[7(6101) - (205)^2][7(17.06) - (10.82)^2]}} = 0.809$$

Existe en este caso, una correlación positiva de 0.809, entre el peso del producto y el tiempo de embalaje.

Determinación de la correlación de las variables en R

nombre1<-c(dato1[1],..,dato1[n]) **nombre2<-c(dato2[1],..,dato2[n])**	**#### lectura de los datos**
cor(nombre1,nombre2)	### calcula la correlación de las variables

Comando en R evaluación del coeficiente de correlación de Pearson

```
peso<-c(25.0, 35.0, 28.0, 27.0, 25.0, 33.0, 32.0)
tiempo<-c(1.25, 1.75,        1.40,      1.35,      1.25,      1.65,      1.60)
cor(peso, tiempo)
```

RESUMEN

Al seleccionar una muestra de n observaciones para evaluar el comportamiento de una población en particular, es importante evaluar ciertos estadísticos (características evaluadas en la muestra seleccionada) para lograr una mejor comprensión del fenómeno en estudio. Estos estadísticos calculados en la muestra son las medidas de localización o centramiento, medidas de variabilidad y las medidas de forma o apuntamiento.

Entre las medidas de localización las más importantes son la media aritmética, la media armónica, la media geométrica, la mediana y la moda; las tres primeras se caracterizan en que su cálculo involucra todas las observaciones, mientras que las restantes no. La medidas de variación más destacados como la varianza, la deviación estándar, el rango. Estas medidas permiten determinar la homogeneidad de la información, es decir el grado de dispersión con respecto a una medida central.

Unidad 3

Las medidas de probabilidad

INTRODUCCIÓN

La medida de probabilidad es un mecanismo por el cual una variable aleatoria o evento de interés, tiene la posibilidad de ocurrencia, con el propósito de tomar las decisiones más asertivas en la evaluación de un fenómeno estocástico en particular. Canavos G. (1993).

Propiedades de las medidas de probabilidad

La teoría de la probabilidad es deducida a partir de axiomas de la teoría de conjunto:

a. sea A' el complemento del evento A, entonces
   $P(A') = 1 - P(A)$

   puesto que $A' \cup A = S$, siendo $A' \cap A = \emptyset$ entonces,

   $P(A' \cup A) = P(S)$ por tanto $P(A') + P(A) = P(S) = 1$, el evento S es el espacio muestral o todos los elementos de la población de interés.

b. la medida de probabilidad de un evento imposible es,
   $P(\emptyset) = 0$

En un espacio muestral S , donde se presenta un evento E, se define como la función de probabilidad $P(E)$ sobre este espacio muestral es definido como:

a. $P(E) > 0$

b. $P(S) = !$

c. $P(E_1 \cup E_2 \cup \ldots \cap E_k) = P(E_1) + \cdots + P(E_k)$
   donde $E_i \cap E_j = \Phi$

d. $P(\tilde{E} \cup E) = P(\tilde{E}) + P(E) = 1$
donde $\tilde{E}$ es el complemento del evento E

PROBABILIDAD CLÁSICA

La probabilidad clásica está definida como la posibilidad de ocurrencia de un evento de interés entre varias condiciones que se produzcan, es decir la frecuencia con la cual obtiene un resultado de un experimento sobre el cual se conocen todos los resultados posibles del espacio muestral. La medida de probabilidad clásica está definida de la siguiente forma,

$$P(E) = \frac{n_i}{\sum_{i=1}^{n} n_i} \qquad (19)$$

donde n_i es el conteo de los elementos del evento E, $\sum_{i=1}^{n} n_i$ son todos los resultados obtenidos en el espacio muestral.

Ejercicio 3.1

A los juegos olímpicos de Brasil 2016 fueron 147 atletas –la delegación más numerosa de la historia del país– que se llevaron 14 diplomas olímpicos y 8 medallas. Por primera vez, además, Colombia consiguió tres preseas doradas y eso le alcanzó para obtener su mejor puesto en un medallero olímpico (23). ¿Qué porcentaje de la delegación obtienen diplomas olímpicos?

La ecuación (19) indica que la medida de probabilidad de recibir un diploma olímpico es definida como,

$$P(i) = \frac{14}{147} = 0.095$$

donde i es el evento de interés, es decir los atletas que recibieron diploma olímpico, $n_i = 14$ y $\sum_{i=1}^{n} n_i = 147$.

El porcentaje de atletas que recibieron diploma olímpico se determina mediante 0.095×100 % es decir 9.5 %.

Regla de la adición

La medida de probabilidad para aplicar la regla de la adición se define mediante la siguiente ecuación,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (20)$$

donde $\mathrm{P(A)}$, $\mathrm{P(B)}$ son las probabilidades marginales definidas como $P(A) = \frac{n_i}{\sum_{i=1}^{n} n_i}$ y $\mathrm{P(B)} = \frac{\mathrm{n_j}}{\sum_{\mathrm{j=1}}^{\mathrm{n}} \mathrm{n_j}}$ y $P(A \cap B)$ es la probabilidad conjunta definida mediante la ecuación $\mathrm{P(A \cap B)} = \frac{\mathrm{n_{ij}}}{\sum\sum_{\mathrm{ij=1}} \mathrm{n_{ij}}}$

TÉCNICAS DE CONTEO

Permite determinar el número de resultados de un espacio muestral. Entre las más aplicadas tenemos:

Regla de la multiplicación de conteo

Una operación se realiza de n_1 formas para cada una de estas formas y de estas formas existen n_2 maneras de realizar la operación, replicando de esta manera un número k de veces, entonces el número de maneras totales es, $n_1 \times n_2 \times \ldots \times n_k$

Ejercicio 3.2

¿Cuántas maneras posibles?, un atleta de salto en garrocha puede estar preparado en su logística personal; cuenta con 3 pantalonetas, 4 camisetas, 5 zapatillas, 7 astas, 8 manillas y dos relojes.

Por tanto, el número de maneras posibles es $n_1 \cdot n_2 \cdot n_3 \ldots n_k = 3 \cdot 4 \cdot 5 \cdot 7 \cdot 8 \cdot 2 = 6720$ maneras.

Combinación

Es el número de resultados obtenidos en una muestra n cuando se toma x elementos en ella, se calcula mediante la siguiente ecuación,

$$_{n}C_{x} = \frac{n!}{(n-x)!x!} \tag{21}$$

Ejercicio 3.3

Obtener el número de maneras posibles, se pueden obtener 3 elementos de una muestra de 10 unidades. Reemplazando en la fórmula $_{10}C_{3} = \frac{10!}{(10-3)!3!} = 120$

permutación

Es la determinación del número ordenado de una operación. Se calcula mediante $n!$

Ejercicio 3.4

Determine el número de maneras posible de ordenar cinco personas en una fila.

lo establecido en la fórmula anterior es

$5! = 120$.

Permutación con muestra

Es el número de resultados ordenados obtenidos en una muestra n cuando se toma x elementos de ella, se calcula mediante,

$$_{n}P_{x} = \frac{n!}{(n-x)!} \qquad (22)$$

Ejercicio 3.5

Evalúe el número de ordenamientos de tres (3) futbolistas, tomados de una muestra de 10 jugadores.

Aplicando la ecuación (21)

$$_{10}p_{3} = \frac{10!}{(10-3)!} = 720$$

PROBABILIDADES CONJUNTA, MARGINAL Y CONDICIONAL

El concepto de probabilidad conjunta, marginal y condicional es definido en la siguiente tabla de contingencia:

Tabla 30

Contingencia en función de los años de experiencia deportiva y su edad

	B_1	B_2	Total
A_1	$n_{11}i$	n_{12}	$n_{1.}$
A_2	$n_{21}i$	$n_{22}i$	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	n

⊙ **Fuente:** elaboración por el Autor.

La probabilidad marginal está definida como la evaluación de la ocurrencia de un evento en particular, por ejemplo, en la Tabla 30, evaluar la medida de probabilidad del evento A_1 implica la suma de los elementos n_{11} y n_{12}, por tanto, la medida de probabilidad marginal, tomando el concepto clásico de probabilidad, está definida mediante la siguiente razón:

$$P(A_1) = \frac{n_{11} + n_{12}}{n} = \frac{n_{1.}}{n} \qquad (23)$$

Entre tanto, seleccionar la ocurrencia de los eventos A_1 y B_1 en forma simultánea, el número de elementos que cumplen con estos dos eventos o atributos es n_{11}, esta medida de probabilidad denominada conjunta está definida como,

$$P(A_1 \cap B_1) = \frac{n_{11}}{n} \tag{24}$$

Ejercicio 3.6

El estudio se llevó a cabo con 130 deportistas de género femenino de hockey sobre hierba, con edades comprendidas entre 13 y 37, ver Tabla 31. Determine las siguientes medidas de probabilidad: ¿Porcentaje de los deportistas menores de 18 años? ¿Porcentaje de los atletas que presentan niveles de experiencia deportiva hasta los 3 años? ¿Porcentaje de los atletas que presentan niveles de experiencia deportiva hasta los 3 años y con edades menores a 18 años? ¿Porcentaje de los atletas que presentan niveles de experiencia deportiva hasta los 3 años o con edades menores a 18 años?

Tabla 31

Contingencia en función de los años de experiencia deportiva y su edad

		Edad < 18	Edad > 18	Total
	Hasta 3 años	26	19	45
Experiencia deportiva	Más de 3 años	44	41	85
Total		70	60	130

⊙ **Fuente:** elaboración por el Autor

En la ecuación (19) indica que la medida de probabilidad marginales,

$$P(\text{Edad} < 18) = \frac{70}{130} = 0.538$$

$$P(\text{Hasta 3 años}) = \frac{45}{130} = 0.346$$

Los participantes en el estudio un 53,8 % de deportistas son menores de 18 años y un 46,1 % son mayores o iguales a 18 años. Los atletas presentan distintos niveles de experiencia deportiva, siendo que un 34,6 % posee hasta 3 años y un 65,4 % más de 3 años.

Seleccionando la ecuación (24) se evalúa la probabilidad conjunta

$$P(\text{Hasta 3 años} \cap \text{Edad} < 18) = \frac{26}{130} = 0.20$$

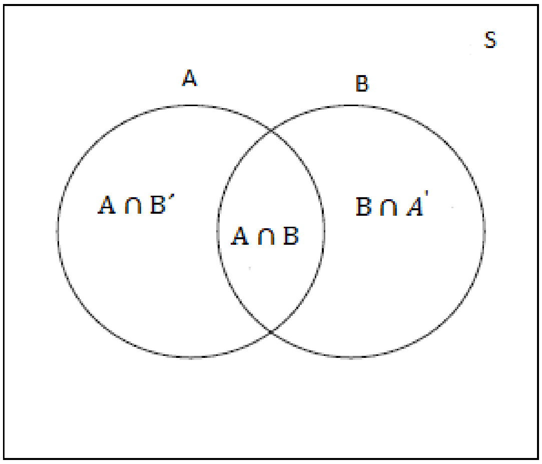
donde $n_{ij} = 26$ y $\sum \sum_{ij=1} n_{ij} = 130$

Es decir, el 20 % de los participantes en el estudio son menores de 18 años y con experiencia menor a 3 años.

$$P(\text{Hasta 3 años} \cup \text{Edad} < 18) = \frac{45}{130} + \frac{70}{130} - \frac{26}{130} = \frac{89}{130} = 0.68$$

El 68 % de los participantes en el estudio son menores de 18 años o con experiencia menor a 3 años.

Regla de la adición



Gráfica 17
Diagrama de Venn de la medida de probabilidad de la unión e intercepción de dos eventos.

La existencia del supuesto de dependencia e independencia de los eventos, defina una segunda regla denominada Regla de la Multiplicación, establecida en las siguientes ecuaciones,

$$B = (A \cap B) \cup (A' \cap B),$$

entonces

$$A \cup B = (A \cap B) \cup (A \cap B') \cup (A' \cap B),$$

por tanto

$$P(A) = P(A \cap B) + P(A \cap B'), \qquad P(A \cap B') = P(A) - P(A \cap B),$$

$$P(B) = P(A \cap B) + P(A' \cap B'), \qquad P(A' \cap B) = P(B) - P(A \cap B),$$

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

esta última ecuación es conocida como la regla de la adición.

Medida de probabilidad condicionada

Algunas veces esta medida de probabilidad de un evento específico está condicionada a la ocurrencia inicial de evento en particular; medida denominada condicionada,

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B/A) \times P(A)}{P(B)} \qquad (25)$$

La probabilidad condicionada debe cumplir con los siguientes axiomas,

a. El cociente de números no negativos arroja un valor no negativo, es decir,

$$P(A/B) \geq 0$$

b. La probabilidad condicionada de un evento seguro es la unidad

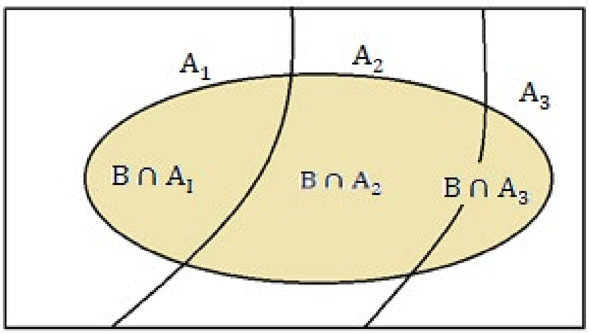
$$P(S|B) = P(S \cap B)/P(B) = P(B)/P(B) = 1$$

c. Dado dos eventos disjuntos A_1 y A_2, es decir $A_1 \cap A_2 = \emptyset$

$$\begin{aligned} P(A_1 \cup A_2|B) &= P(A_1 \cap B) + P(A_2 \cap B)/P(B) = P(A_1 \cap B)/P(B) + P(A_2 \cap B)/P(B) \\ &= P(A_1/B) + P(A_2/B) \end{aligned}$$

Medida de probabilidad total

Por otra parte, la medida de probabilidad marginal de un evento dado probabilidades condicionales se determina mediante la Regla General de la Multiplicación. Esta ecuación se pude comprender analizando la Gráfica 11.



Gráfica 18
Obtención de la Regla General de la Multiplicación del Evento B (color)

⊙ **Fuente:** elaboración por el Autor

Obsérvese que en el Gráfico 18, el evento A es la suma de las intersecciones de la siguiente forma,

$$B = B \cap A_1 + B \cap A_2 + B \cap A_3,$$

entonces,

$P(B) = P(B \cap A_I) + P(B \cap A_2) + P(B \cap A_3)$,

como $P(A \cap B) = P(A/B) \times P(B)$

entonces,

$P(B) = P(B/A_1) \times P(A_1) + P(B/A_2) \times P(A_2) + P(B/A_3) \times P(A_3)$, es decir,

$$P(B) = \sum_{i=1}^{k} P(B/A_i) \times P(A_i), \tag{26}$$

Teorema de Bayes

En el caso de la existencia de un conjunto de eventos A_i, como también un evento cualesquiera B del espacio, algunas veces es necesario evaluar la medida de probabilidad de uno de los eventos A_i condicionada a la ocurrencia de un evento inicial B

$$P(A_I/B) = \frac{P(B/A) \times P(A_I)}{P(B/A_1) \times P(A_1) + P(B/A_2) \times P(A_2) + P(B/A_3) \times P(A_3)} = \frac{P(B/A) \times P(A_I)}{\sum_{i=0}^{k} P(B/A_I) \times P(A_I)} \tag{27}$$

establecida mediante una regla denominada regla de Bayes, Bayes, (1763).

Ejercicio 3.8

Un estudio indica que tres zonas del país aportan la totalidad de la delegación de un país: la zona Caribe, Central y Pacífica con 30, 45, 25 % respectivamente. Esta información se complementa con la probabilidad de obtener medalla olímpica proveniente de las zonas Caribe, Central y Pacífica es de 10, 15 y 5 %. ¿Cuál es la probabilidad de que al seleccionar un atleta que obtenga medalla olímpica, este sea de la zona Caribe?

Solución: El problema presenta la siguiente información: $P(ZCar) = 0.30$, $P(ZCent) = 0.45$ y $P(ZPac) = 0.25$; además $P(Medalla/ZCar) = 0.10$, $P(Medalla/ZCent) = 0.15$ y $P(Medalla/ZPac) = 0.05$. Aplicando la Regla de Bayes de la ecuación (24),

$$P(ZCar/Medalla) = \frac{P(Medalla/ZCar)P(ZCar)}{P(Medalla)} = \frac{010 \times 0.30}{P(Medalla)}$$

La $P(Medalla)$ es obtenida de la ecuación (25) de la Regla general de la Multiplicación de la siguiente forma,

$$\begin{aligned} P(Medalla) &= P(Medalla/ZCar)P(ZCar) + P(Medalla/ZCent)P(ZCent) + P(Medalla/ZPac)P(ZPac) \\ &= 0.10 \times 0.30 + 0.15 \times 0.45 + 0.05 \times 0.25 = 0.11 \end{aligned}$$

Entonces,

$$P(ZCar/Medalla) = \frac{010 \times 0.30}{0.11} = 0.272$$

La probabilidad de seleccionar un atleta que obtenga medalla olímpica sea de la zona Caribe es de 27;2 %.

Medidas de probabilidad para eventos independientes

$$\mathrm{P(A \cap B) = P(A) \times P(B/A) = P(B) \times P(A/B) = P(B \cap A) =}\ \mathrm{P}(B) \times \mathrm{P(B)} \tag{28}$$

Al considerar que los eventos son independientes, la Regla de la Multiplicación para eventos independientes se define en la siguiente forma,

$$\mathrm{P(A \cap B) = P(A) \times P(B) = P(B \cap A)} \tag{29}$$

Ejercicio 3.7

Las estadísticas de los Juegos Centroamericanos y del Caribe indican que el 93 % de las delegaciones presentes, no presentan problemas logísticos en la permanencia en la ciudad sede. ¿Cuál es la probabilidad de que, al seleccionar una muestra de tres de las delegaciones, dos de ellas posean problemas logísticos en la sede de los juegos?

Solución: A, B y C evento delegaciones sin inconvenientes en la logística, por lo que A^0, B^0 y C^0 delegaciones con problemas logísticos. La ecuación (25) indica que la medida de probabilidad es,

$$\sum \mathrm{P(A \cap B \cap C) = P(A) \times P(B^0) \times P(C^0) + P(A^0) \times P(B) \times P(C^0) + P(A^0) \times P(B^0) \times P(C)}$$

$$= 0.93 \times 0.07 \times 0.07 + 0.07 \times 0.93 \times 0.07 + 0.07 \times 0.07 \times 0.93 = 0.014$$

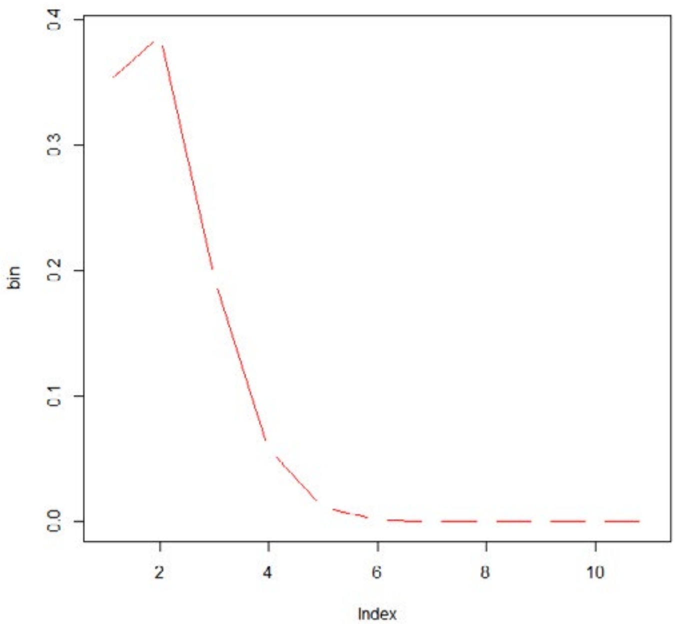
La probabilidad de que dos de tres delegaciones posean problemas de logística en sus instancias es de 1,4%.

Unidad 4

Distribuciones: discretas - continuas de probabilidad

DISTRIBUCIÓNES DISCRETAS: BINOMIAL- POISSON-GEOMÉTRICA

Distribución Binomial



Gráfica 19
Distribución Binomial con parámetro $p = 0.1, n = 10$

⦿ **Fuente:** elaboración por el Autor

Es una distribución de probabilidad utilizada de una variable aleatoria discreta resultante de un experimento de Bernoulli. En donde cada evento tiene solo dos resultados posibles o categorías: la ocurrencia del evento de interés denominado éxito o su no ocurrencia llamado fracaso. Este proceso de Bernoulli posee dos supuestos: a) la probabilidad del resultado de interés permanece fija con el tiempo; b) los eventos son estadísticamente independientes, es decir, el resultado de un evento no afecta al de cualquier evento de ocurrencia posterior en este experimento. La distribución de probabilidad está definida mediante la siguiente ecuación,

$$B(x,n,p) = \frac{n!}{(n-x)!x!} p^x \times (1-p)^{n-x} \qquad (30)$$

donde p es el porcentaje del evento considerado éxito, n es el tamaño de la muestra seleccionada. Es de recordar que la expresión $\frac{n!}{(n-x)!x!}$, es la técnica de conteo denominada combinación.

Gráfico de la distribución Binomial en R

```
bin<-dbinom(0:10,10,0.75)
plot(bin, type="co", col="blue")
```

Esta distribución tiene la función acumulada de la medida de probabilidad definida con la siguiente expresión,

$$P(X \leq x) = F(x) = \sum_0^x \frac{n!}{(n-x)!x!} p^x \times (1-p)^{n-x} \qquad (31)$$

Tabla 32

Distribución Binomial y Poisson

	Fórmula aplicada	Primer momento	Segundo momento	Observaciones
BINOMIAL	$b(x,n,p) = \frac{n!}{(n-x)!x!} p^x \times (1-p)^{n-x}$	np	$np(1-p)$	Dos categorías, p éxito, $(1-p)$ fracaso, eventos independientes con p constante
POISSON	$p(x,\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ	λ promedio del evento, eventos independientes

⊙ **Fuente:** elaboración por el Autor

Medida de probabilidad mediante la distribución Binomial en R

pbinom(c(x), size=n, prob=0.5, lower. tail=TRUE)	**#### medida de probabilidad de la distribución binomial acumulada**
pbinom(c(12), size=25, prob=0.2, lower.tail=-FALSE)	**### Probabilidad de un valor superior**

Ejercicio 4.1

La posibilidad de que un ciclista profesional triunfe en una carrera del calendario sigue una distribución Binomial con un porcentaje de éxito del 15 %. El calendario indica que se realizaran 10 carreras ciclísticas.

¿Cuál es la probabilidad de que este ciclista triunfe en dos (2) de estas carreras?

Solución: Usando las fórmulas de la distribución Binomial ecuaciones (30) y (31) donde la variable aleatoria X es el número de carreras ganadas:

$$P(X=2)=\frac{10!}{(10-2)!2!}0.15^2\times(1-0.15)^{10-2}=\frac{10\times9\times8\times7\times6\times5\times4\times3\times2\times1}{(8\times7\times6\times5\times4\times3\times2\times1)\times2\times1}\times0.15^2\times(0.85)^8=0.275$$

Existe una probabilidad de 0,275, es decir 27,5 % de ganar en dos de las carreras.

Ejercicio 4.2

¿Cuál es la probabilidad de que este ciclista triunfe en más de (una) de estas carreras?

Solución: Utilizando la distribución Binomial Acumulada de la ecuación (23) donde la variable aleatoria X es el número de carreras ganadas:

$$\begin{aligned}P(X>1)&=1-P(X\le 1)=1-[P(X=0)+P(X=1)]\\&=1-\left[\frac{10!}{(10-0)!\,0!}\times0.15^0\times(1-0.15)^{10-10}=\frac{10!}{(10-1)!\,1!}\times0.15^1\times(0.85)^{10-1}\right]\\&=1-0.544=0.456\end{aligned}$$

La probabilidad de que este ciclista triunfe en más de (1) de estas carreras es de 45,6 %.

Ejercicios de medidas de probabilidad mediante la distribución Binomial en R

Ejercicio 4.6

Comando en R, evaluación de la medidas de probabilidad puntual de la distribución de Binomial

```
### en un intervalo
pbinom(c(4, 25), 60, 0.02)
```

Distribución Poisson

Un fenómeno Poisson se identifica por las siguientes condiciones:

a. La ocurrencia del evento se da a una velocidad constante en el tiempo o en el espacio.

b. La frecuencia de ocurrencia es constante y se representa por λ en un intervalo de tiempo o espacio.

c. La medida de probabilidad es proporcional a la longitud del intervalo establecido.

La función de probabilidad está dada por:

$$P(x,\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!} \tag{32}$$

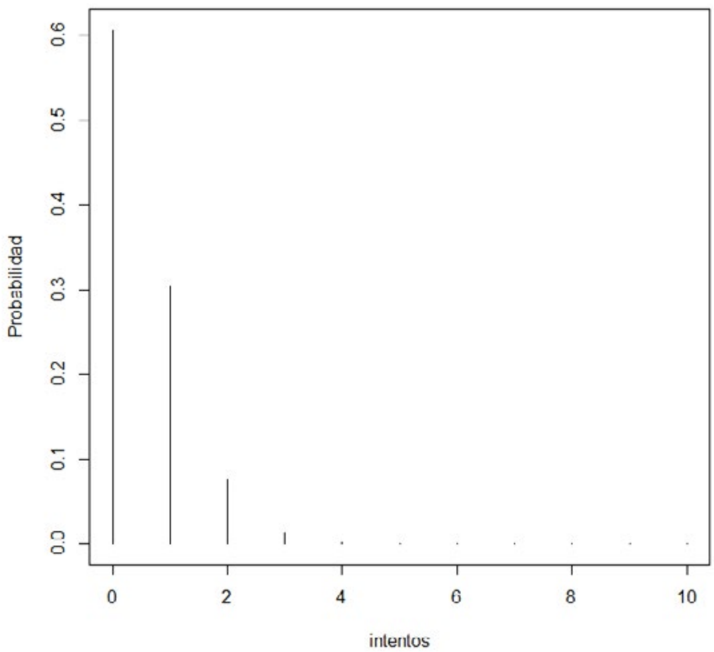
La función acumulada de la distribución de Poisson se presenta de la siguiente forma,

$$\mathrm{P}(\mathrm{X} \leq \mathrm{x}) = \mathrm{F}(\mathrm{x}) = \sum_{0}^{x} \frac{e^{-\lambda}\lambda^{x}}{x!} \tag{33}$$

donde λ es el parámetro definido como el promedio de ocurrencia del evento de interés.

Gráfico de la distribución Poisson en R

```
poi<-dpois(0:10,lambda=0.5) #funcion en masa
plot(0:10, dpois(0:10,0.5), type='h', xlab="intentos",ylab="Probabilidad" )
```



Gráfica 20

Distribución Poisson con parámetro $\lambda = 0.5$

⊙ **Fuente:** elaboración por el Autor

Medida de probabilidad acumulada mediante la distribución Poisson en R

ppois(c(x), lambda=valor del promedio, lower. tail=TRUE)	#### medida de probabilidad de la distribución binomial acumulada
ppois(c(x), lambda=valor del promedio, lower. tail=TRUE)	### Probabilidad de un valor superior
dpois(variable, lambda=promedio)	### valor puntual x

Ejercicio 4.4

El número de fallas que presenta un vehículo de carreras sigue una distribución de Poisson con una media de 0,02 por circuito. ¿Cuál es la probabilidad de que no presente ningún error en el circuito?

Solución: La distribución Poisson según la ecuación (24) donde la variable aleatoria x es el número de fallas que presenta el vehículo y el promedio de falla $\lambda = 0.02$:

$$P(X = 0) = \frac{e^{-0.02} 0.02^0}{0!} = 0.980$$

La probabilidad de que no se presente falla en este vehículo en un circuito es de 0.98, es decir un porcentaje del 98 %.

Comando en R, evaluación de la medidas de probabilidad puntual de la distribución de Poisson
dpois(0, lambda=0.02)

Ejercicio 4.3

El número de goles de un equipo profesional sigue una distribución de Poisson con una media de 2 por encuentro. ¿Cuál es la probabilidad de que este equipo anote exactamente tres goles?

Solución: La distribución Poisson según la ecuación (31) donde la variable aleatoria X es el número de goles anotados por el equipo y el promedio por anotación $\lambda = 2$:

$$P(X = 3) = \frac{e^{-2} 2^3}{3!} = 0.1804$$

La probabilidad de que este equipo anote exactamente tres goles es de 0,1804 es decir un porcentaje de 18,04 %.

Comando en R, evaluación de la medidas de probabilidad puntual de la distribución de Poisson
dpois(3, lambda=2)

¿Cuál es la probabilidad de que este equipo anote más de tres goles en dos partidos?

Solución: La variable aleatoria X es el número de goles anotados por el equipo y el promedio por anotación se modifica $\lambda = 4$, ya que establece para dos partidos, utilizando la ecuación (31)

$$P(X > 3) = 1 - P(X \leq 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 3)]$$
$$= 1 - \left[\frac{e^{-4}4^0}{0!} + \frac{e^{-4}4^1}{1!} + \frac{e^{-4}4^2}{2!} + \frac{e^{-4}4^3}{3!}\right] = 1 - 0.433 = 0.567$$

La probabilidad de que este equipo anote más tres goles en dos partidos es de 0,567 es decir un porcentaje de 56,7 %.

Comando en R, evaluación de la medidas de probabilidad puntual de la distribución de Poisson
ppois(c(3), lambda=4, lower.tail=FALSE)

Distribución Geométrica

La distribución Geométrica debe cumplir con los siguientes supuestos:

a. La probabilidad del resultado de interés p permanece fija con el tiempo

b. Los eventos estadísticamente son independientes.

c. La variable X es la ubicación o lugar en donde se presenta la ocurrencia del evento éxito.

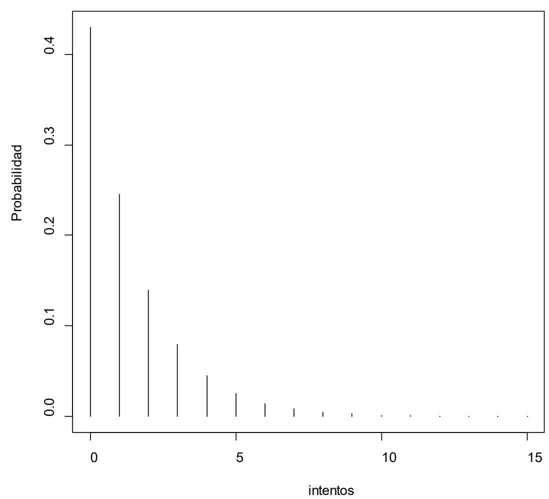
La función de probabilidad está dada por:

$$G(x,p) = p(1-p)^{x-1} \tag{34}$$

donde p es el parámetro definido como la proporción de ocurrencia del evento de interés.

Gráfico de la distribución Geométrica en R

```
plot(0:15, dgeom(0:15,0.43), type='h', xlab="intentos",ylab="Probabilidad" )
```



Gráfica 21
Distribución Geométrica con parámetro $p = 0.43$

Fuente: elaboración por el Autor

Medida de probabilidad acumulada mediante la distribución Poisson en R

dgeom(x-1, p)	## x ubicación del evento de intres
	## p porcentaje del evento de interés

Ejercicio 4.5

El porcentaje de efectividad de un jugador de futbol americano es del 43 %, ¿cuál es la medida de probabilidad de que en la décima segunda oportunidad sea efectivo?

Solución: La distribución Geométrica ecuación (32) donde la variable aleatoria X es la ubicación o momento, en este caso de anotar, es decir $x = 12$, la medida de probabilidad de ocurrir este evento es,

$$P(X = 12) = 0.43 \times (1 - 0.43)^{12-1} = 0.000887$$

La medida de probabilidad de anotar, por parte de este jugador, es de 0.000887, es decir, un porcentaje de 0,0887 %.

Comando en R, evaluación de la medidas de probabilidad puntual de la distribución Geométrica
dgeom(**11, 0.43**)

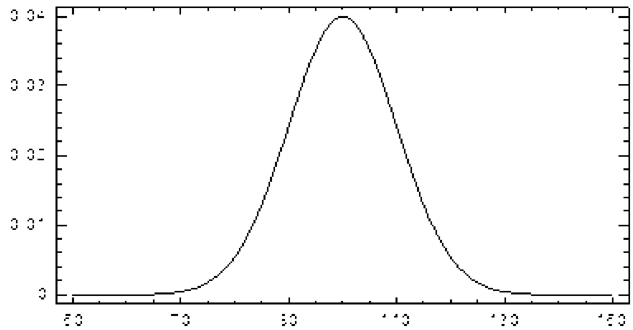
DISTRIBUCIÓN NORMAL

El modelo de probabilidad más frecuente utilizado en estadística es la distribución normal, ver Gráfica 22. Modelo que puede emplearse en la forma general y estandarizada. La importancia de la distribución normal se debe principalmente a que existen variables asociadas a fenómenos naturales que siguen el modelo de la distribución normal; entre otros por ejemplo los caracteres morfológicos o físicos de los individuos, los errores generados por la utilización de un instrumento de medición. La variable X tiene una distribución normal, si cumple las siguientes condiciones: a) X es una variable aleatoria continua, b) si existen los parámetros de la media μ y la desviación estándar σ, c) la función de probabilidad viene dada por la siguiente expresión:

$$X \sim N(x, \mu, \sigma^2) = \int_{\propto}^{t} \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} dt \qquad (35)$$

El modelo normal estándar es trabajado con una función de densidad que comprende integrales, que no pueden ser reducidas a funciones elementales; se dice que es una función normal estándar si su media es cero y su varianza es la unidad, es decir $E(X) = 0$ y $v(x) = 1$. La función de distribución de la variable normal estándar Z es,

$$Z \sim N(0,1) = \int_{-\propto}^{z_i} \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}[z]^2} dz$$



Gráfica 22

Modelo de la distribución Normal

⊙ **Fuente:** elaboración por el Autor

La ecuación (33) representa una distribución de probabilidades, es decir, un modelo que permite explicar el comportamiento de numerosas variables aleatorias físicas, previamente estandarizas. Este modelo tiene dos parámetros: uno de posición o de localización, que marca el lugar central de la curva correspondiendo al valor esperado o media y otro es la dispersión. Esta curva normal posee extremos infinitamente largos, la gran mayoría de las observaciones se encontrarán concentradas alrededor del valor esperado, a una distancia no mayor que tres desviaciones estándar. La función acumulada correspondiente a la distribución estándar dada la medida de probabilidad de la variable normal estándar, asuma valores menores o iguales a Z.

$$\mathbf{F(z) = P[Z \leq z] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}[Z]^2}} \qquad (36)$$

donde la nueva variable Z es de la siguiente forma $z = \frac{x-\mu}{\sigma}$.

Entre las características que posee una distribución normal, se tiene:

a. Las distribuciones normales se caracterizan por dos parámetros: la media μ y la desviación estándar σ.

b. El de la máxima valor de la variable aleatoria X se encuentra en la media como el punto más alto de una curva.

c. La media de una distribución normal pertenece al conjunto de los números reales, es decir puede tomar cualquier valor positivo, negativo o cero.

d. La planicie de la curva normal depende del valor de la desviación estándar, es decir entre mayor sea este, el ancho de la curva es mayor.

Los valores de los intervalos de los percentiles comúnmente utilizados indican que el 68,3 % de los valores de una variable aleatoria normal se encuentran más o menos a una desviación estándar de la media, el 95,4 % se encuentran más o menos dos desviaciones de la media y el 99,7 % de los valores a más o menos tres desviaciones estándar de la media.

CÁLCULO Y UTILIDAD DE LA DISTRIBUCIÓN NORMAL ESTÁNDAR Z

El valor Z es calculado por medio de, la estandarización de la variable X mediante las formulaciones presentada en la Tabla 33:

Tabla 33

Estandarización de una variable con distribución normal y parámetros conocidos

Fórmula aplicada	Primer momento	Segundo momento	Observaciones
$X \sim N(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}$	μ	σ^2	Es la distribución más utilizada en fenómenos físicos.
$Z \sim N(0,1) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}[z]^2}$	0	1	En este caso se estandariza la variable x en la variable z, de la siguiente forma: $z = \frac{x-\mu}{\sigma}$

⊙ **Fuente:** elaboración por el Autor

La distribución normal es el modelo probabilístico frecuentemente más aplicado a variables aleatorias identificadas generalmente con medidas físicas. Esta información al ajustarse este modelo, son utilizables con el propósito de prevenir y planificar el experimento de estudio, dado su comportamiento similar a largo plazo. De este modo se puede saber con antelación, con una medida de probabilidad definida, los posibles valores de la variable aleatoria monitoreada en el estudio.

DETERMINACIÓN DE LA MEDIDA DE PROBABILIDAD MEDIANTE LA DISTRIBUCIÓN NORMAL ESTÁNDAR

Ejercicio 4.6

¿Qué porcentaje de atletas de varios equipos de básquetbol tienen un porcentaje de grasa mayor de 19, sabiendo que la media es de 16 con una desviación estándar de 3,6

Solución: Usando la fórmula de la normal estandarizada:

Identificamos la variable aleatoria X como el porcentaje de grasa, asumiendo una distribución normal de esta variable; los parámetros son conocidos: la media $\mu = 16$ y la desviación estándar $\sigma = 3.6$.

$$P[X > 19] = P\left[Z > \frac{19-16}{3.6}\right] = P[Z > 0.833] = 1 - P[Z \leq 0.833]$$

En la Tabla del **Anexo A**, se encuentra la medida de probabilidad para un valor de $Z = 0.833$, la medida de probabilidad de atletas con grasa corporal arriba de 19 es.

$$P[X > 19] = 1 - 0.7978 = 0.2022$$

El porcentaje de atletas con grasa corporal por encima de 19 es del 20,22 %.

Medida de probabilidad mediante la distribución normal estándar en R

pnorm(c(x), mean=valor de promedio, sd=valor de la desviación , lower.tail=TRUE)	**#### medida de probabilidad de la distribución normal**
pnorm(c(x), mean=valor de promedio, sd=valor de la desviación , lower.tail=FALSE)	### Probabilidad de un valor superior

Comando en R, evaluación de la medida de probabilidad distribución normal.

```
pnorm(c(19), mean=16, sd=3.6 , lower.tail=TRUE)
pnorm(c(19), mean=16, sd=3.6 , lower.tail=FALSE)
```

Ejercicio 4.7

La información presenta en el Ejercicio 3, asuma una distribución normal de los pesos obtenidos en libras de unos boxeadores de la rama juvenil, de la categoría pesada de la WBA. Los parámetros estimados de la media y la desviación estándar son respectivamente $\mu = 199.865$ y $\sigma = 1.134$. Evalúe la medida de probabilidad de que el peso de uno de los boxeadores este por encima de 201 libras.

Solución: Usando la fórmula de la normal estandarizada:

$$P[X > 205] = P\left[Z > \frac{201 - 199.865}{1.134}\right] = P[Z > 1.00] = 1 - P[Z \leq 1.00]$$

Se determina el valor de la medida de probabilidad utilizando la Tabla Normal Estándar Acumulada establecida en el Anexo A

$$P[X > 205] = 1 - 0.8413 = 0.158$$

Esto implica que el 15,8 % de los deportistas poseen un peso por encima de 205 libras en la categoría de pesos Pesado de la WBA.

Comando en R, evaluación de la medida de probabilidad distribución normal.

```
pnorm(c(205), mean=199.865, sd=1.134 , lower.tail=FALSE)
```

Ejercicio 4.8

El número de atletas de una determinada región del país suman 125 individuos. El plan médico indica que el contenido de glicemia sigue una distribución normal con media de 85 y una desviación estándar de 10,5. Los parámetros establecidos por el departamento médico indican que atletas con valores de glicemia por encima de 96 deben someterse a una dieta especial. ¿Cuántos de estos atletas deben someterse a la dieta?

Solución: Usando la fórmula de la normal estandarizada:

$$P[X > 96] = P\left[Z > \frac{96-85}{10.5}\right] = P[Z > 1.04] = 1 - P[Z \leq 1.04]$$

Se determina el valor de la medida de probabilidad utilizando la Tabla Normal Estándar Acumulada establecida en el Anexo A

$$P[X > 96] = 1 - 0.8504 = 0.149$$

Esto implica que el 14,9 % de los atletas deben someterse a la dieta. Por tanto, el número de atletas que deben someterse a una dieta especial es:

$$125 \times 0.149 = 18.6 \approx 19$$

Diecinueve (19) atletas deben someterse a un plan nutricional.

Comando en R, evaluación de la medida de probabilidad distribución normal.
pnorm(c(96), mean=85, sd=10.5 , lower.tail=TRUE)

Tabla 34
Distribuciones continuas de probabilidad

Nombre de la distribución	Formula aplicada	Medida de localización	Medida de variabilidad	Observaciones
Normal	$N \sim (x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}$ $-\infty \leq x \leq \infty$ $-\infty \leq \mu \leq \infty$ $\sigma \geq 0$ $F(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} d(x)$	$E(x) = \mu$	$V(x) = \sigma$	Es la distribución más utilizada sobre todo en fenómenos físicos.
Normal estándar	$N \sim (z, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[Z]^2}$ $-\infty \leq z \leq \infty$ $\mu = 0$ $\sigma = 1$ $F(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[z]^2} d(z)$	$E(z) = \mu = 0$	$V(z) = \sigma^2 = 1$	Es la distribución más utilizada sobre todo en fenómenos físicos. En este caso se estandariza la variable x en la variable z, de la siguiente forma: $z = \frac{x-\mu}{\sigma}$

⊙ **Fuente:** elaboración por el Autor

Tabla 35

Distribuciones continuas de probabilidad (continuación)

Nombre de la distribución	Modelo	Medida de localización	Medida de variabilidad	Observaciones
Gamma	$G(x,\theta,\alpha)=\frac{1}{\tau(\alpha)\theta^{\alpha}}x^{\alpha-1}e^{-\frac{x}{\theta}}$ $x>0$ $\theta,\alpha>0$ $F(x,\theta,\alpha)=p(x\leq x)=1-\begin{bmatrix}1+\frac{x}{\theta}+\frac{1}{2!}(\frac{x}{\theta})^{2}\ldots\\+\frac{1}{(\alpha-1)!}(\frac{x}{\theta})^{\alpha-1}\end{bmatrix}e^{-\frac{x}{\theta}}$	$E(x)=\alpha\theta$	$V(x)=\alpha\theta^{2}$	Aplicada en Teorías de cola, tiempo de garantía de un producto
Exponencial	$E(x,\theta)=\frac{1}{\theta}e^{-\frac{x}{\theta}}$ $x>0$ $\theta>0$ $F(x,\theta)=P(X\leq x)=1-e^{-\frac{x}{\theta}}$	$E(x)=\theta$	$V(x)=\theta^{2}$	Tiempo de un servicio, tiempo de vida de un objeto

⊙ **Fuente:** elaboración por el Autor

La disminución expinencial es aplicada con frecuencia en fenómenos como tiempo de servicio, en el caso de las actividades fisícas, los números obtenidos por un nador en un intervalo de preparación. El modelo exponencial esta definido. Ver tabla 35

$$E(X,\theta)=\frac{1}{\theta}\quad EXP\quad[-X/\theta] \tag{37}$$

Su función acumulado

$$F(X)=1-EXP\quad[-X/\theta] \tag{38}$$

Donde θ es el tiempo promedio

Ejercicio 4.9

Los tiempos de recorrido de un atleta de medio fondo sigue una distribución exponencial con $\theta = 20$ segundos cual es la probabilidad de que ese atleta posea un tiempo entre 18 y 19 segundos?

X: Tiempo de recorrido

$$P\,(18 < x < 19) = F\,(20) - F\,(18) = \left\{1 - EXP\left[\frac{19}{20}\right]\right\} - \left\{1 - EXP\left[\frac{18}{20}\right]\right\}$$

$= 0{,}6132 - 0{,}5934 = 0{,}019$

El 1,9 % de las veces recorridas entre 18 y 19 segundos

UNIDAD 5

ESTADÍSTICA DE ALGUNOS DEPORTES

ESTADÍSTICA EN EL FÚTBOL

El procedimiento para evaluar los puntajes de cualquier equipo que represente un país y obtenga buenos resultados en el fútbol internacional alcanzará ciertos puntos que permitirán ascender en la clasificación mundial FIFA. Los puntos acumulados por cada uno de los equipos afiliados a la FIFA en un cuatrienio se obtienen sumando los siguientes criterios:

a. El número de puntos obtenidos en un encuentro internacional.

b. La media o promedio de puntos obtenidos en los últimos partidos, se estima durante doce meses (12).

c. La media de puntos obtenidos en partidos anteriores a los últimos doce (12) meses (depreciación anual).

Los puntos obtenidos en el ítem a) se estima por partido dependiendo de los siguientes factores,

a. Resultado del encuentro Victoria-Empate $(\mathrm{M} = \mathrm{encuentro})$

b. Importancia internacional del encuentro $(\mathrm{I} = \mathrm{importancia})$

c. Ubicación del equipo rival en la clasificación de la FIFA y la Confederación que está representando. $(\mathrm{T} = \mathrm{TablayC} = \mathrm{confederación})$

Tabla 36

Factor de fuerza de la confederación en las tres últimas copas mundiales y los valores equivalentes

Confederación	Tras la CM 2006	Hasta e incluida la CM 2006
UEFA	1,00	1,00
CONMEBOL	0.98	0.99
CONCACAF	0.85	0.88
AFC	0.85	0.85
CAF	0.85	0.85
OFC	0.85	0.85

⊙ **Fuente:** elaboración por el Autor

Estos factores se resumen en una fórmula para determinar el número total de puntos obtenidos por el Equipo $(P = \text{puntos})$.

$$P = M \times I \times T \times C \tag{39}$$

Lo siguiente son los criterios para asignar los puntos Tabla 37, para la aplicación de la ecuación (39)

Tabla 37

Criterio de puntajes FIFA

Criterio	Puntajes	Observaciones
M: Puntos por victoria	Los equipos ganan 3 puntos por victoria, 1 punto por empate y 0 puntos por derrota.	En una tanda de tiros penales, el ganador obtiene 2 puntos y el perdedor 1 punto.
I: Importancia del partido	Amistoso (incluidos los torneos menores): I = 1.0	
	Eliminatoria mundialista o en el ámbito de la confederación: I = 2.5	
	Competición final de confederación o Copa FIFA Confederaciones: I = 3.0	
	Competición final de la Copa Mundial de la FIFA: I = 4.0	
T: Fuerza de los contendientes	La fuerza de los contendientes se basa en la siguiente fórmula: $200 - \text{el puesto en la clasificación de los contendientes}$	Como excepción de esta fórmula, se asigna siempre al equipo a la cabeza de la clasificación el valor 200 y a los equipos clasificados en el puesto 150, y subsiguientes se les asigna un valor mínimo de 50. El puesto en la Tabla se obtiene de la última Clasificación Mundial FIFA/Coca-Cola publicada.
C: Fuerza de la confederación	Al calcular partidos entre equipos de distintas confederaciones, se emplea el valor medio de las confederaciones a las que pertenecen los equipos que compiten. La fuerza de una confederación se calcula de acuerdo con el número de victorias que ha obtenido en las últimas tres ediciones de la Copa Mundial de la FIFA. Los valores son los siguientes:	CONMEBOL 1.00
		UEFA 0.99
		CONCACAF/AFC/CAF/OFC 0.85

⊙ **Fuente:** elaboración por el Autor

Ejercicio 5.1

Se desea obtener el puntaje P de dos equipos: A y B; el primero es de la **CONCACAF, el segundo equipo** de la **AFC**, basado en los criterios de medición resumidas en las Tablas 31, 32. ¿Cuáles son los puntajes finales P de cada uno de los equipos después del resultado?

Tabla 38
Ejemplo de los criterios de puntajes FIFA

	Equipo A	**Equipo B**
Fecha del partido	16 de junio de 2006	
Importancia del partido	Competición final Copa Mundial FIFA	
Resultado	0	0
Posición en la Tabla	11	59
Fuerza regional	0.88	0.85
M (puntos por el resultado)	1	1
I (importancia del partido)	4	
T (fuerza del adversario)	1.41	1.89
C (fuerza de la confederación)	0.865	
$P = M \times I \times T \times C$	487.9	653.9

⊙ **Fuente:** elaboración por el Autor

Las puntuaciones de los Equipos A y B se pueden apreciar en la Tabla 33, donde el Equipo A presenta un puntaje final de 487,9 menor al Equipo B, con un puntaje final de 653,9, aplicando la ecuación 26. Es importarte resaltar que el valor de la ponderación de la fuerza de confederación final 0,865 es un promedio simple entre 0,88 y 0,85

ESTADÍSTICA EN EL BÉISBOL

Las estadísticas del béisbol son relativamente sencillas de evaluar, es el caso:

a. El promedio de bateo en donde se divide el número de Hits sobre el número de turnos oficiales al bate, lo que resulta la siguiente ecuación:

$$\text{Promedio de bateo} = \frac{\text{Número de Hits realizado por el bateador}}{\text{Número de turnos al bate}} \times 100 \qquad (34)$$

Ejercicio 5.2

Es el caso de un bateador con 121 turnos al bate en la temporada, de los cuales el número de Hits realizado es de 45.

Solución: La ecuación (27) indica que el promedio de bateo es el siguiente:

$$\text{Promedio de bateo} = \frac{45 \text{ Hits realizado por el bateador}}{121 \text{ turnos al bate}} \times 100 = 371$$

El número promedio de bateo para este bateador en particular es de 371.

b. La Efectividad en el lanzamiento, en donde las carreras limpias lanzadas se multiplican por nueve y el resultado se divide en los innings lanzados.

$$\text{Efectividad en el lanzamiento} = \frac{\text{Número de carreras lanzadas} \times 9}{\text{Número de innigs laanzados}} \quad (41)$$

Ejercicio 5.3

Un lanzador realiza 16 carreras en 49,2 innings. ¿Cuál su efectividad de lanzamiento?

Solución: La ecuación (39) indica que la efectividad del lanzamiento está definida de la siguiente forma:

$$\text{Efectividad en el lanzamiento} = \frac{16 \times 9}{49.2} = 2.92$$

La efectividad de lanzamiento es de 2.92

ESTADÍSTICA EN LA FÓRMULA UNO

En el automovilismo, específicamente la fórmula uno, es importante conocer el ranking liderado de los mejores equipos como se muestra en la Tabla 34, en donde se muestra el liderazgo de los últimos años de las dos escuderías: Mercedes y Ferrari como las más **fiables, pero también las más rápidas** de manera constante. En el otro extremo encuentran los equipos como Sauber y McLaren. Este ranking solo toma los **kilómetros recorridos** en toda la temporada, es claro que los equipos que lo lideran, es por pocas paradas en la pista, como también circuitos sin abandono.

Tabla 39

Kilómetros completados en cada equipo

Equipo	Kilómetros
Mercedes	5087.652
Ferrari	4440.87
Williams	3710.035
Force India	3649.52
Haas	3323.67
Sauber	3305.05
Red Bull	3174.71
Renault	2774.38
Toro Rosso	2718.52
McLaren	1978.35

⊙ **Fuente:** elaboración por el Autor

Las escuderías intentan mejorar notablemente en los test la media recorrida en el pilotaje, ya que esta información es una de las variables que puede medir la inestabilidad en la conducción por parte del piloto.

Tabla 40

Ranking de kilómetros promedio por marca de motor

Motor	Kilómetros	Media/equipo
Mercedes	12447.47	4149.156
Ferrari	7764.54	3882.270
Renault	8667.61	2889.203
Ferrari 2016	3305.05	3305.050
Honda	1978.375	1978.375

⊙ **Fuente:** elaboración por el Autor

Tabla 41

Posición de los pilotos de carrera en la temporada 2017

Posición	Piloto	Equipo	Tiempo	Neumático
1º	Kimi Räikkönen	Ferrari	1:18.634	Superblando
2º	Sebastian Vettel	Ferrari	1:19.024	Ultrablando
3º	Valtteri Bottas	Mercedes	1:19.310	Superblando
4º	Lewis Hamilton	Mercedes	1:19.352	Ultrablando
5º	Felipe Massa	Williams	1:19.420	Ultrablando
6º	Max Verstappen	Red Bull	1:19.438	Superblando
7º	Carlos Sainz	Toro Rosso	1:19.837	Ultrablando
8º	Nico Hülkenberg	Renault	1:19.885	Ultrablando
9º	Daniel Ricciardo	Red Bull	1:19.900	Ultrablando
10º	Sergio Pérez	Force India	1:20.116	Ultrablando

⊙ **Fuente:** elaboración por el Autor

UNIDAD 6

ESTIMACIÓN PUNTUAL E INTERVALO DE CONFIANZA. METODOLOGÍA DEL CONTRASTE DE HIPÓTESIS

ESTIMACIÓN DE LOS PARÁMETROS DE UNA POBLACIÓN

El conocimiento de una población está sujeto a la comprensión de la distribución de probabilidad de las variables aleatorias de la misma, es decir, la dificultad se reduce a la estimación de los parámetros poblacionales

una muestra aleatoria de la forma $x_1, x_2 \dots x_n$ es seleccionada de una población, es definida una función de una medida de la muestra $\hat{\theta} = \hat{\theta}(x_1, x_2 \dots x_n)$ de forma que f es una estimación del parámetro poblacional θ, obtenido mediante la función $a = a(x_1, x_2 \dots x_n)$, variante de muestra en muestra. Esto indica $\hat{\theta}$ cómo una variable aleatoria, con una distribución de probabilidad conocida, modelo paramétrico, denominado estadístico.

Los estadísticos como (buen) estimador, deben cumplir con ciertas propiedades:

a. El insesgamiento o centramiento, en el caso de la media, el valor esperado de la muestra coincide con el parámetro de la población, es decir, $E(\hat{\theta}) = \theta$

b. Eficiencia: la existencia de dos estimadores θ_1 y θ_2, se dice que θ_1 es más eficiente que θ_2 cuando $\sigma_{\theta_1} > \sigma_{\theta_2}$

c. Consistencia: al aumentar el tamaño de la muestra se aproxima asintóticamente al valor del parámetro de la población, $\lim_{n \to \infty} (\hat{\theta}) = \theta$

DISTRIBUCIÓN DE MUESTREO

Las distribuciones de muestreo constituyen en el área de la estadística los modelos muestrales que permite entre otras razones:

a. La estimación de la viabilidad de la población basada en la muestra seleccionada en el experimento.

b. Estimación de la media de la población, ajustado al Teorema del Límite Central, en donde existe una medida de probabilidad de encontrarse en un intervalo establecido.

Entre los modelos de muestreo más aplicados se observa en la Tabla 37, donde las estimaciones de la media se utilizan la distribución de muestreo Normal y la t de Students; en el caso de la estimación de la varianza o razón de varianza se realiza con las distribuciones de muestreo Chi-cuadrado y F de Fisher,

Tabla 42

Estandarización de los modelos de distribución de muestreo

Nombre	Formulación simplificada	Aplicación
Normal	$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$	Promedio con σ^2 conocida
T de Student	$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$	Promedio con σ^2 desconocida
Chi – cuadrado	$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$	Varianza σ^2
F de Fisher	$F = \frac{s_1^2}{s_2^2}$	Razón de varianzas σ_1^2 / σ_2^2 (comparación)

⊙ **Fuente:** elaboración por el Autor

ESTIMACIÓN PUNTUAL

En la Estimación Puntual, el parámetro de la población se infiere mediante el valor de un estadístico, tomado de la muestra; es el caso por ejemplo de la media muestral, bajo el supuesto de normalidad, este estadístico es una estimación del parámetro de la media μ.

ejercicio 6.1

Sea $\bar{x}$ el promedio obtenido de una muestra seleccionada aleatoriamente de un proceso. Deduzca si es un estimador insesgado de μ.

El promedio muestral está definido como,

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Aplicando Esperanza

$$E[\bar{x}] = E\left[\sum_{i=1}^{n} \frac{x_i}{n}\right] = \frac{1}{n} E\left[\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\left[\sum_{i=1}^{n} E(x_i)\right]$$

por definición $E(x_i) = \mu$, entonces,

$$E[\bar{x}] = \frac{1}{n}\left[\sum_{i=1}^{n} \mu\right] = \frac{1}{n} n\mu = \mu$$

Por tanto, que $\bar{x}$ es un estimador insesgado, es decir el estadístico toma el valor del parámetro.

ESTIMACIÓN POR INTERVALO DE LA MEDIA

Una estimación puntual proporciona un valor no preciso del parámetro poblacional, esto implica el desconocimiento de la medida de probabilidad de su cercanía o no del verdadero valor del parámetro. Entre tanto, el interés es arrojar una estimación, precisa de la incertidumbre de la evaluación, mediante la estimación por intervalos de confianza; es calculado este intervalo, con cierta probabilidad de ocurrencia de que el parámetro poblacional desconocido esté contenido en este. El intervalo de confianza consiste en calcular dos estimadores: un límites inferior $L1$ y uno superior $L2$, el intervalo de confianza es resumido en la pareja de valores $I = [L1, L2]$ denominada estimador por intervalo, los límites del intervalo serán función de los valores de la siguiente forma,

$$L_1 = \boldsymbol{\theta}_1(X_1, X_2, \ldots, X_n)\ ;\ L_2 = \boldsymbol{\theta}_2(X_1, X_2, \ldots, X_n).$$

Esto implica, un valor definido en el intervalo aleatorio obtenido de la muestra para estimar el parámetro $\boldsymbol{\theta}$ de la población, estimados mediante un intervalo de confianza cuya notación general es:

$$P[L_1 \leq \boldsymbol{\theta} \leq L_2] = 1 - \alpha,$$

el caso de la media con varianza conocida:

$$P\left[-\boldsymbol{z}_{\frac{\alpha}{2}} \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \boldsymbol{z}_{1-\frac{\alpha}{2}}\right] =$$

$$P\left[-\boldsymbol{z}_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{x} - \mu \leq \boldsymbol{z}_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

entonces

$$P\left[\bar{x} - \boldsymbol{z}_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \boldsymbol{z}_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha \qquad (42)$$

donde α es el nivel de significancia o error tipo I asumido por el investigador, n es el tamaño de la muestra y $z_{1-\frac{\alpha}{2}}$ es el valor de la variable normal estándar dado un nivel de significancia α.

El intervalo de confianza, para la media μ conocida la varianza poblacional σ^2 esta dada por la siguiente ecuación:

$$\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

En resumen, el intervalo de confianza de la media, con varianza σ^2 conocida es,

$$I = \left[\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] \tag{43}$$

En la Tabla 43 se presentan algunos valores de $z_{\frac{\alpha}{2}}$ dado los valores asumidos del nivel de significancia α.

Tabla 43

Valores estandarizados de Z basado en el nivel de significancia confiabilidad

Nivel de significancia	Valor de la puntuación Z	Confiabilidad
0.100	1.645	90.0
0.050	1.965	95.0
0.010	2.576	99.0
0.001	3.290	99.9

⦿ **Fuente:** elaboración por el Autor

El intervalo de confianza, para la media μ desconocida la varianza poblacional σ^2 esta dada por:

$$P\left[\bar{x} - t_{t(\alpha/_2;\, n-1)} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{t(1-\alpha/_2;\, n-1)} \frac{s}{\sqrt{n}}\right] = 1 - \alpha$$

entonces.

$$\bar{x} - t_{(\alpha/_2;\, n-1)} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{(1-\alpha/_2; n-1)} \frac{s}{\sqrt{n}} \tag{44}$$

donde $t\alpha/_2$ es el percentil de la distribución t de Students con $v = n - 1$ grados de libertad.

El intervalo

$$I = \left[\bar{x} \pm t_{(\alpha/_2;\, n-1)} \frac{s}{\sqrt{n}}\right] \tag{45}$$

Ejercicio 6.2

Estime el intervalo de confianza para el verdadero promedio μ del tiempo de un nadador, el estadístico de localización obtenido en la muestra de cinco datos seleccionados de las cinco primeras pruebas es de $\bar{x} = 110$ segundos, con varianza conocida $\sigma^2 = 36$. El error tipo I o nivel de significancia aplicada para esta estimación es de 5%.

Solución: El nivel de significancia del 5 % indica que $z_{0.025} = 1.96$, utilizando la ecuación (43) se tiene que:

$$110 - 1.96\frac{6}{\sqrt{5}} \leq \mu \leq 110 + 1.96\frac{6}{\sqrt{5}}$$

Lo que implica que el verdadero promedio de tiempo se encuentra en el intervalo 104.74 a 115.25 con una confiabilidad del 95 %.

Ejercicio 6.3

Estime el intervalo de confianza para el promedio μ, del ejercicio anterior asumiendo que la varianza de la población no es conocida, los datos obtenidos de los tiempos en las cinco pruebas realizadas son: 112, 113, 110, 115 y 108.

Solución: El nivel de significancia del 5 % indica que el valor de $t_{0.025\,(4)} = 2.77$, con cuatro grados de libertad, el promedio y la desviación muestrales es respectivamente 111.6 y 2.70 segundos, utilizando la ecuación (45):

$$111.6 - 2.77\frac{2.70}{\sqrt{5}} \leq \mu \leq 111.6 + 2.77\frac{2.70}{\sqrt{5}}$$

Lo que implica que el verdadero promedio de tiempo de este nadador se encuentra en el intervalo 108.25 a 114.94 segundos, con una confiabilidad del 95 %.

ESTIMACIÓN POR INTERVALO DE LA DIFERENCIA DE MEDIAS

El intervalo de confianza para la diferencia de las medias con varianzas conocidas se define en la siguiente ecuación,

$$[\bar{x}_1 - \bar{x}_2] - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq [\bar{x}_1 - \bar{x}_2] + z_{(1-\alpha/2)}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

El intervalo de confianza de diferencias de medias con varianzas conocidas es,

$$\boldsymbol{I} = \left[[\bar{x}_1 - \bar{x}_2] \pm z_{(\alpha/2)}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right] \qquad (46)$$

En el caso que las varianzas son desconocidas el intervalo de confianza para la diferencia de las medias se define como,

$$[\bar{x}_1 - \bar{x}_2] - t_{(\alpha/_2;\,v)}\, s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq [\bar{x}_1 - \bar{x}_2] - t_{(1-\alpha/_2;\,v)}\, s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \qquad (47)$$

donde la desviación ponderada $s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$, el percentil $t_{\alpha/_2,v}$ de la distribución t de Students con $v = n_1 + n_2$ grados de libertad.

El intervalo de confianza $1-\alpha$, para diferencia de medias con varianzas desconocidas pero iguales es,

$$I = \left[[\bar{x}_1 - \bar{x}_2] \pm t_{(\alpha/_2;n_1+n_2-2)}\, s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right] \qquad (48)$$

El desconocimiento de las varianzas de poblacionales implica modificar la ecuación de la siguiente forma,

$$P\left\{[\bar{x}_1 - \bar{x}_2] - t_{(\alpha/_2;\,v)}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq [\bar{x}_1 - \bar{x}_2] - t_{(1-\alpha/_2;\,v)}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right\} = 1-\alpha$$

Los grados de libertad están definidas como,

$$v = \frac{\left[\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right]^2}{\frac{\left[\frac{s_1^2}{n_1}\right]^2}{n_1+1}+\frac{\left[\frac{s_2^2}{n_2}\right]^2}{n_2+1}} - 2$$

el intervalo

$$I = \left[[\bar{x}_1 - \bar{x}_2] \pm t_{(\alpha/_2;v)}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right] \qquad (49)$$

Ejercicio 6.4

La siguiente información muestra los pesos levantados por dos atletas en forma independiente, realizado en un intervalo de tiempo equidistante para evitar la fatiga muscular. La información se presenta en la Tabla 44. ¿Se pueden considerar similares los pesos levantados por los atletas? Utilice un nivel de significancia de 5 %

Tabla 44

Peso levantado por dos atletas en intervalos de tiempo equidistantes

Atleta A	Atleta B
232	225
231	223
235	234
230	228
236	

⊙ **Fuente:** elaboración por el Autor

Solución: El nivel de significancia del 5 % indica que el valor de $t_{0.025\,(8)} = 2.36$, con ocho grados de libertad, el promedio muestral para cada atleta es de 232.5 y 277.5 respectivamente, utilizando la ecuación (48):

$$[232.5 - 277.5] - 2.36 \times 3.69\sqrt{\frac{1}{5}+\frac{1}{5}} \leq \mu_1 - \mu_2 \leq [232.5 - 277.5] + 2.36 \times 3.69\sqrt{\frac{1}{5}+\frac{1}{5}}$$

La desviación ponderada es calculada de la siguiente forma,

$$S_p = \sqrt{\frac{(5-1)\times 2.58^2 + (5-1)\times 4.79^2}{5+5-2}} = 3.69$$

La diferencia de medias arroja los siguientes valores en el intervalo, -50.50 y -39.49. Los valores resultantes son negativos, lo que implica que el primer atleta levanta menos peso que el segundo atleta, con una confiabilidad del 95 %.

ESTIMACIÓN PARA PROPORCIONES

El caso en donde la población posee un comportamiento ajustado a una distribución Binomial, con parámetro p, como razón entre el número de éxitos del evento de interés, sobre la muestra seleccionada; generalmente el intervalo para esta fracción se aproxima a una distribución normal con media y varianza de la distribución muestral de una proporción: $\mu_p = np$ y $\sigma^2 = p(1 - p)/n$, el intervalo de confianza para la proporción es.

$$P\left[\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

donde $\hat{p} = \frac{x}{n}$, x es el número de éxitos, n es la muestra seleccionada

el intervalo

$$I = \left[\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] \qquad (43)$$

ESTIMACIÓN PARA LAS DIFERENCIAS DE PROPORCIONES

El intervalo de confianza para la comparación entre los parámetros, p_1 y p_2 , de dos distribuciones binomiales. El estimador puntual de esta diferencia es la diferencia de proporciones $\hat{p}_1 - \hat{p}_2$, donde $\hat{p}_1$ es la proporción de éxitos en una muestra de tamaño n_1 de la primera población, y $\hat{p}_2$ es la proporción en una segunda muestra de tamaño n_2

En cuanto a la medición de la varianza muestral proporcional, es expresada de la siguiente forma,

$$\sigma_p^2 = \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}$$

entonces la medida de probabilidad,

$$P\left[(\hat{p}_2 - \hat{p}_1) - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \leq p_1 - p_2 \leq (\hat{p}_2 - \hat{p}_1) + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right] = 1-\alpha$$

donde los valores $\hat{p}_1 = \frac{x_1}{n_1}$, siendo x_1 el número de éxitos de la primera población y $\hat{p}_2 = \frac{x_2}{n_2}$, x_2 el número de éxitos de la segunda población.

el intervalo se define,

$$I = \left[(\hat{p}_2 - \hat{p}_1) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right] \qquad (51)$$

Ejercicio 6.5

Es comparando dos bateadores designados. El procedimiento es realizar el mayor número sencillos en 400 lanzamientos; el bateador A, presenta 32 ponches, el bateador B, solo 28 ponches. Estimar las diferencia entre las proporciones de ponches entre los bateadores usando una confianza igual a 0,95. ¿¡Estadísticamente cuál de los bateadores es más efectivo?

La confiabilidad del 95 % en este intervalo implica un valor de $z_{\frac{\alpha}{2}} = 1.96$

los valores de las muestras $n_1 = 400$ y $n_2 = 400$, por lo tanto, $\hat{p}_1 = \frac{368}{400} = 0.92$ y $\hat{p}_2 = \frac{372}{400} = 0.93$

entonces

$$(0.92 - 0.93) - z_{0.025}\sqrt{\frac{0.92(0.08)}{400}+\frac{0.93(0.07)}{400}} \leq p_1 - p_2$$
$$\leq (0.92 - 0.93) + z_{0.025}\sqrt{\frac{0.92(0.08)}{400}+\frac{0.93(0.07)}{400}}$$

evaluando

$$(0.01) - 1.96(0.01862) \leq p_1 - p_2 \leq (0.01) + 1.96(0.01862)$$

por tanto,

$$-0.0264 \leq p_1 - p_2 \leq 0.0465$$

el intervalo de confianza no es considerablemente distinto.

Ejercicio 6.6

Las encuestas Time-Yankelovich, periódicamente publicada en la revista Time, informan que acerca de los resultados de consultas telefónicas a unas 1000 personas en el mes de enero, el 60 % de los que respondieron dijo que le preocupaba el mantenimiento de los escenarios deportivos. En una encuesta semejante en junio de ese mismo año, solo el 50 % dijo que le preocupaba. Estimar la diferencia verdadera en esas proporciones, en un intervalo de confianza del 95 %.

Los resultados en su orden en amabas encuestas,

$\hat{p}_1 = \frac{600}{1000} = 0.60,$ y $\hat{p}_2 = \frac{500}{1000} = 0.50$

entonces

$$I = \left[(0.60 - 0.50) \pm 1.96\sqrt{\frac{0.60(0.40)}{1000} + \frac{0.50(0.50)}{1000}}\right] = (0.1) - 1.96(0.022) \leq p_1 - p_2$$
$$\leq (0.1) + 1.96(0.022) = 0.0566 \leq p_1 - p_2 \leq 0.1434$$

Intervalo de confianza para la varianza en R

ic.var <- function(nombre de la variable, conf. level=valor de la confiabilidad)	**#### nombre de la variable, confiabilidad**
alfa <- 1 - conf.level	## nivel de significancia
n <- length(x)	### tamaño de muestra
l1 <- (n - 1) * var(x) / qchisq(1 - alfa / 2,n - 1)	##Límite inferior del intervalo
l2 <- (n - 1) * var(x) / qchisq(alfa / 2,n - 1)	##Límite superior del intervalo
ic <- c(l1,l2)	####intervalo
return(ic)	
ic.var(x)	## intervalo de la varianza

Unidad 7

Metodología del contraste de hipótesis

La metodología científica es caracterizada dentro de su sistemática metodología, por la construcción de hipótesis acerca del comportamiento o funcionamiento de los fenómenos de interés, pretendiendo simplificar la decisión de rechazar o no la hipótesis planteada, mediante la selección de una muestra experimental representativa de la población evaluada. A través de este contraste de hipótesis, la metodología estadística presenta procedimientos adecuados para determinar el rechazo o no de la hipótesis planteada en la investigación con un margen de error predeterminado por el investigador.

PROCEDIMIENTO DE UN CONTRASTE DE HIPÓTESIS

El contraste de hipótesis es una herramienta aplicada para determinar si una afirmación o conjetura es corroborada por la evidencia muestral, acerca del valor de un parámetro o parámetros, como también es posible contrastar el comportamiento de una información bajo el supuesto de un modelo conocido.

El proceso metodológico del contraste de hipótesis induce a la presentación de dos tipos de hipótesis:

La hipótesis nula H_0 es la afirmación realizada acerca del parámetro. Por ejemplo, he de afirmar que el promedio de la cantidad de colesterol de un atleta en particular es de 145 mg, es decir $H_0: \mu = 145$. Esta afirmación en el proceso investigativo debe contrastarse con la evidencia muestral, por lo que se plantea la hipótesis alterna, definida como la negación de la hipótesis nula; planteando tres formas o maneras de negar la hipótesis nula un contraste bilateral y dos unilaterales.

a. El caso de los contrastes bilaterales, en el ejemplo expuesto, la hipótesis nula $H_1: \mu \neq 145$, indica un promedio diferente a 145, puede poseer un valor menor o mayor a 145

b. Entre tanto, el contraste unilateral a la izquierda, el valor del promedio es menor a 145, $H_1: \mu < 145$,

c. Mientras el contraste unilateral a la derecha, el promedio es mayor a el valor de145, $H_1: \mu > 145$,

El siguiente es el procedimiento general de un contraste de hipótesis:

1. Plantear la hipótesis nula $H_0: \theta = \theta_0$
2. 2. Negación de la hipótesis nula, mediante la hipótesis alterna de tres formas: en forma bilateral $H_1 \theta \neq \theta_0$, unilateral a derecha $H_1 \theta > \theta_0$, unilateral a la izquierda $H_1 < \theta_0$.
3. Asignación del nivel de significancia o error Tipo I (α).
4. Definir el estadístico de prueba con su área de rechazo; está definida mediante los siguientes estadísticos de prueba: de manera bilateral $d \leq d_{\alpha/2}$ o $d \geq d_{1-\alpha/2}$ unilateral a la izquierda $d < d_\alpha$, unilateral a la derecha $d > d_{1-\alpha}$. Los estadísticos de prueba d se presentan en la Tabla 45.
5. Determinación de los estadísticos de prueba d.
6. Decisión: rechazar o no rechazar la hipótesis nula

Tipos de errores en un contraste de hipótesis

Existen dos tipos de errores en una prueba de hipótesis, el error tipo I o nivel de significancia α alfa, y el error tipo II β beta. El error Tipo I ocurre cuando es rechazada la hipótesis nula siendo esta verdadera y el error Tipo II, ocurre cuando no es rechazada la hipótesis nula, siendo esta falsa, el resumen se presenta en la Tabla 45.

Tabla 45

Errores en los contrastes de hipótesis

Decisión	**H_0 verdadera**	**H_0 falsa**
No se rechaza H_0	Decisión correcta	Error tipo II β
Se rechaza H_0	Error tipo I α	Decisión correcta

⊙ **Fuente:** elaboración por el Autor

En este caso el rechazo de la hipótesis nula H_0, cuando es verdadera, esto implica un error denominado error tipo I; entre tanto, al no rechazar la hipótesis nula H_0, cuando es falsa, indica un error de tipo II.

Tabla 46

Estadísticos de prueba en contraste de hipótesis

Estadístico de prueba d	Parámetros evaluados
$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$	La media con varianza conocida
$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$	La media con varianza desconocida
$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$	La varianza muestral
$F = \frac{s_1^2}{s_2^2}$	Razón de varianza
$z = \frac{\hat{p} - p}{\frac{p(1-p)}{n}}$	La proporción poblacional
$Z = \frac{[\bar{x}_1 - \bar{x}_2] - [\mu_1 - \mu_2]}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	Diferencia de medias con varianza conocida para muestras independientes
$t = \frac{[\bar{x}_1 - \bar{x}_2] - [\mu_1 - \mu_2]}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	Diferencia de medias con varianzas desconocidas para muestras independientes
$z = \frac{(\hat{p}_2 - \hat{p}_1) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$	Diferencia de proporciones para muestras independientes

⦿ **Fuente:** elaboración por el Autor

Decisión mediante el P valor

La decisión de rechazar o no la hipótesis nula se puede resolver mediante el valor P, definida de la siguiente forma: cuando el valor $P = P[D \geq d] < \alpha$ se debe rechazar la hipótesis nula, en pruebas bilaterales el valor $P = 2P[D \geq d]$.

Contraste de hipótesis de la media con varianza conocida σ^2

El siguiente es el procedimiento general de un contraste de hipótesis:

1. Plantear la hipótesis nula $H_0: \mu = \mu_0$
2. La negación de la hipótesis nula, mediante la hipótesis alterna en forma bilateral $H_1 \mu \neq \mu_0$,
3. Asignación del nivel de significancia α.
4. El estadístico de prueba está definida mediante,

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \tag{52}$$

con su área de rechazo,

$z \leq z_{\frac{\alpha}{2}}$ o $z \geq z_{1-\frac{\alpha}{2}}$

Contraste de hipótesis de la media con varianza desconocida σ^2

El siguiente es el procedimiento general de un contraste de hipótesis:

1. Plantear la hipótesis nula $H_0: \mu = \mu_0$
2. La negación de la hipótesis nula, mediante la hipótesis alterna en forma bilateral $H_1 \mu \neq \mu_0$,
3. Asignación del nivel de significancia α.
4. El estadístico de prueba está definida mediante,

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} \tag{53}$$

con su área de rechazo,

$t \leq t_{\frac{\alpha}{2}}$ o $t \geq t_{1-\frac{\alpha}{2}}$

con ν grados de libertad

Contraste de hipótesis de la varianza σ^2

El siguiente es el procedimiento general de un contraste de hipótesis:

1. Plantear la hipótesis nula $H_0: \sigma^2 = \sigma_0^2$
2. La negación de la hipótesis nula, mediante la hipótesis alterna en forma bilateral $H_1: \sigma^2 \neq \sigma_0^2$,
3. Asignación del nivel de significancia α.
4. 4. El estadístico de prueba está definida mediante,

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \tag{54}$$

con su área de rechazo,

$\chi^2 < \chi^2_{\frac{\alpha}{2},v}$ o $\chi^2 > \chi^2_{1-\frac{\alpha}{2},v}$

con v grados de libertad

Contraste de hipótesis para la proporción p

El siguiente es el procedimiento general de un contraste de hipótesis:

1. Plantear la hipótesis nula $H_0: p = p_0$
2. La negación de la hipótesis nula, mediante la hipótesis alterna en forma bilateral $H_1: p \neq p_0$,
3. Asignación del nivel de significancia α.
4. El estadístico de prueba está definida mediante,

$$z = \frac{\hat{p} - p}{\frac{p(1-p)}{n}} \tag{55}$$

con su área de rechazo,

$z \leq z_{\frac{\alpha}{2}}$ o $z \geq z_{1-\frac{\alpha}{2}}$

Contraste de hipótesis para la diferencias de medias $\mu_1 - \mu_2$ con varianzas conocidas σ_1^2 y σ_2^2

El siguiente es el procedimiento general de un contraste de hipótesis:

1. Plantear la hipótesis nula $H_0: \mu_1 = \mu_2$
2. La negación de la hipótesis nula, mediante la hipótesis alterna en forma bilateral $H_1: \mu_1 \neq \mu_2$,
3. Asignación del nivel de significancia α.
4. El estadístico de prueba está definida mediante,

$$z = \frac{[\bar{x}_1 - \bar{x}_2] - [\mu_1 - \mu_2]}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (56)$$

con su área de rechazo,

$z \leq z_{\frac{\alpha}{2}}$ o $z \geq z_{1-\frac{\alpha}{2}}$

Contraste de hipótesis para la diferencias de medias $\mu_1 - \mu_2$ con varianzas desconocidas σ_1^2 y σ_2^2

El siguiente es el procedimiento general de un contraste de hipótesis:

1. Plantear la hipótesis nula $H_0: \mu_1 = \mu_2$
2. La negación de la hipótesis nula, mediante la hipótesis alterna en forma bilateral $H_1: \mu_1 \neq \mu_2$,
3. Asignación del nivel de significancia α.
4. El estadístico de prueba está definida mediante,

$$t = \frac{[\bar{x}_1 - \bar{x}_2] - [\mu_1 - \mu_2]}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (57)$$

con su área de rechazo,

$t \leq t_{\frac{\alpha}{2},v}$ o $t \geq t_{1-\frac{\alpha}{2},v}$

con v grados de libertad

Contraste de hipótesis para la diferencias de proporciones $p_1 - p_2$

El siguiente es el procedimiento general de un contraste de hipótesis:

1. Plantear la hipótesis nula $H_0: p_1 = p_2$
2. La negación de la hipótesis nula, mediante la hipótesis alterna en forma bilateral $H_1: p_1 \neq \boldsymbol{p_2}$,
3. Asignación del nivel de significancia α.
4. El estadístico de prueba está definida mediante,

$$z = \frac{(\hat{\boldsymbol{p}}_2 - \hat{\boldsymbol{p}}_1) - (p_1 - p_2)}{\sqrt{\frac{\hat{\boldsymbol{p}}_1(\mathbf{1} - \hat{\boldsymbol{p}}_1)}{\boldsymbol{n_1}} + \frac{\hat{\boldsymbol{p}}_2(\mathbf{1} - \hat{\boldsymbol{p}}_2)}{\boldsymbol{n_2}}}} \tag{58}$$

con su área de rechazo,

$z \leq z_{\frac{\alpha}{2}}$ o $z \geq z_{1-\frac{\alpha}{2}}$

Contraste de hipótesis para la razón de varianzas σ_1^2 / σ_2^2

El siguiente es el procedimiento general de un contraste de hipótesis:

1. Plantear la hipótesis nula $H_0: \boldsymbol{\sigma_1^2} = \boldsymbol{\sigma_2^2}$
2. La negación de la hipótesis nula, mediante la hipótesis alterna en forma bilateral $H_1: \boldsymbol{\sigma_1^2} \neq \boldsymbol{\sigma_2^2}$,
3. Asignación del nivel de significancia α.
4. El estadístico de prueba está definida mediante,

$$f = \frac{s_1^2}{s_2^2} \tag{59}$$

con su área de rechazo,

$f < F_{\frac{\alpha}{2}, v_1, v_2}$, o $f > F_{1-\frac{\alpha}{2}, v_1, v_2}$

con v_1, v_2 grados de libertad

Ejercicio 7.1

El nivel de ácido graso polisaturado (en porcentaje) es evaluado para un conjunto de futbolistas; se seleccionó una muestra de seis de estos futbolistas con los siguientes resultados: 16.8, 17.2, 17.4, 16.9, 16.5, 17.1. ¿Se puede afirmar que el promedio de ácido polisaturado en el equipo de futbol es de 17%? Utilice un nivel de significancia de 0.01.

Solución: El parámetro de interés es el promedio de ácido polisaturado, en la muestra se determina $\bar{x} = 16.982$, $s = 0.31885$

1) $H_0: \mu = 17$. 2) la hipótesis alterna en forma bilateral $H_1: \mu \neq 17$. 3) Asignación del nivel de significancia o error Tipo I $\alpha = 0.01$.4) El área de rechazo está definida mediante los siguientes estadísticos de prueba: de manera bilateral $t \leq -4.032$ o $t \geq 4.032$.5) Determinación de los estadísticos de prueba $t = \frac{16.983 - 17}{0.31885 / \sqrt{6}} = -0128$ 6) Decisión: Basado en el estadístico de prueba, la muestra evidencia que no se debe rechazar la hipótesis nula.

Utilizando el valor P como la prueba es bilateral es multiplicada por dos (2). $P = 2P[t \geq 4.032] = 2 \times 0.005 = 0.01$ el valor es cercano al nivel de significancia establecido, esta metodología impide la toma de una decisión.

Prueba de hipótesis e intervalo de confianza de la media en R

t.test(nombre de la variable, mu=valor del promedio, conf.level = valor de la confiabilidad)	**#### prueba de hipótesis e intervalo de confianza**

UNIDAD 9

EL ANÁLISIS DE VARIANZA ANOVA

DISEÑO ALEATORIZADO UNIFACTORIAL CON EFECTO FIJO

Es el diseño más sencillo, compara el efecto de $k \geq 2$ niveles de un solo factor sobre una determinada variable problema o respuesta. Los niveles del factor son los tratamientos aplicándose en forma aleatoria a un conjunto virtualmente homogéneo de unidades experimentales. El modelo de diseño del diseño completamente aleatorizado con efecto fijo posee el siguiente modelo estadístico:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad (65)$$

donde $i = 1, 2, \ldots, a$ y $j = 1, 2, \ldots, n$, siendo y_{ij} la ij estima observación, μ es un parámetro común a todos los tratamientos denominado promedio general, τ_i es el efecto del i-ésimo tratamiento y ε_{ij} es el error aleatorio.

En este tipo de modelo el efecto de los tratamientos está definido como la variación del promedio de cada tratamiento con respecto al promedio general, se define como $\sum_{i=1}^{a} \tau_i$, por tanto, las hipótesis a plantear son las siguientes:

$H_0: \mu_1 = \mu_2 = \cdots = \mu_a$

$H_1: \mu_i \neq \mu_j$

También se puede plantear el contraste de hipótesis en términos de los efectos del experimento.

$H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0$

$H_1: \tau_i \neq 0$

Descomposición de las sumas de cuadrados en la Tabla de análisis de varianza

La suma de cuadrados total, definida como la variabilidad de las observaciones en el experimento se puede descomponer o particionar en las siguientes sumas de cuadrados:

$$\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{..}\right)^2=\sum_{i=1}^{a}\left(\bar{y}_{i.}-\bar{y}_{..}\right)^2+\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{i.}\right)^2$$

donde a es el número de niveles, n es el número de réplicas en cada tratamiento, y_{ij} son las respuestas.

Las sumas de cuadrados se pueden simplificar en formulaciones algebraicamente equivalentes, pero más sencillas de utilizar. En su orden las ecuaciones (66) son: la suma de cuadrados totales suma de cuadrados del tratamiento y la suma de cuadrados del error,

$$SST=\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{..}\right)^2=\sum_{i=1}^{a}\sum_{j=1}^{n}y_{ij}^{\,2}-\frac{y_{..}^2}{N}$$

$$SCRT=\sum_{i=1}^{a}y_{i.}^{\,2}-\frac{y_{..}^2}{N}$$

$$SSE=SST-SCRT \qquad (66)$$

Esta descomposición y el contraste de hipótesis es resumida en la Tabla 47, denominada Tabla de análisis de varianza.

Tabla 53

Descomposición de la suma de cuadrados y el contraste de la hipótesis

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios MS	Valor de F
Tratamiento	$a-1$	$SCRT=\sum_{i=1}^{a}y_{i.}^{\,2}-\frac{y_{..}^2}{N}$	$\frac{\sum_{i=1}^{a}y_{i.}^{\,2}-\frac{y_{..}^2}{N}}{a-1}$	$\frac{\frac{\sum_{i=1}^{a}y_{i.}^{\,2}-\frac{y_{..}^2}{N}}{a-1}}{\frac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2}{N-a}}$
Error	$N-a$	$SSE=\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2$	$\frac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2}{N-a}$	
Total	$N-1$	$SST=\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2$		

El cuadrado medio del error o varianza del error σ^2 es la razón de la suma de cuadrados del error y los grados de libertad $MSE=\frac{\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{i.}\right)^2}{N-a}$.

EJERCICIO 41

Una compañía automotriz está investigando cuatro tipos de alerón delanteros con el propósito de obtener un dispositivo más eficiente en la carga aerodinámica, se obtuvieron los siguientes datos de las cargas para cada uno de los alerones evaluados, Tabla 48.

Tabla 54

Carga aerodinámica en porcentaje, de cada uno de los alerones de prueba

	Alerones				
	A	**B**	**C**	**D**	**Total**
	58.2	56.3	50.1	52.9	
	57.2	54.5	54.2	49.9	
	58.4	57.0	55.4	50.0	
	55.8	55.3		51.7	
	54.9				
Total	284.5	223.1	159.7	204.5	871.8
Promedio	56.900	55.775	53.233	51.125	

⊙ **Fuente:** elaboración por el Autor

En este experimento se presentan cuatro niveles $a = 4$ en este experimento, el número total de observaciones $N = 16$.

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N} = 58.2^2 + 57.2^2 + \cdots + 51.7^2 - \frac{871.8^2}{16} = 120.237$$

$$SCRT = \sum_{i=1}^{a}\frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N} = \left[\frac{284.5^2}{5} + \frac{223.1^2}{4} + \frac{159.7^2}{3} + \frac{204.5^2}{4}\right] - \frac{871.8^2}{16} = 85.675$$

$$SSE = SST - SCRT = 120.237 - 85.675 = 34.594$$

Esta información es resumida en la Tabla 49 de análisis de varianza ANOVA, con el propósito de contrastar si existen diferencias entre los cuatro tipos de alerones.

Tabla 49

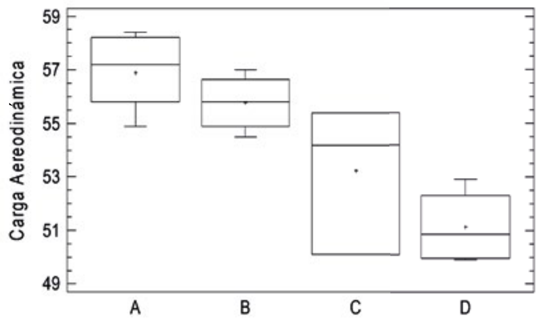
Descomposición de la suma de cuadrados y el contraste de hipótesis

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Valor de F	P-valor
Alerones	3	85.675	28.55	9.913	0.001439
Error	12	34.594	2.88		
Total	15	120.237			

⊙ **Fuente:** elaboración por el Autor

El valor del estadístico de prueba $F = 9.913$, esto indica que se debe rechazar la hipótesis nula de igualdad de los niveles de los tratamientos; esto se verifica con el $P - \text{valor} = 0.001439$ inferior al nivel de significancia establecido del 5 %. En esta Tabla se muestra también que la varianza del error o cuadrado medio del error es $\sigma^2 = 2.88$.

La muestra indica que existen diferencias estadísticas entre los alerones. El siguiente paso es determinar estadísticamente sus diferencias y encontrar el dispositivo más eficiente para el automóvil; inicialmente es recomendable realizar mediante diagramas descriptivos como los diagramas box plots presentados en la Gráfica 23.



Gráfica 23
Diagrama box plots de los diferentes tipos de alerones

⊙ **Fuente elaboración por el Autor**

El diagrama muestra que el alerón A posee la mayor carga aerodinámica, contrario a la carga presentada en el alerón D.

COMPARACIÓN DE PAREJAS DE MEDIAS MEDIANTE EL MÉTODO LSD

Una vez descrito en la Tabla de análisis de varianza ANOVA la diferencia entre los niveles mediante el P-valor; es importante entonces identificar las desigualdades entre los niveles, es decir se contrasta las siguientes hipótesis de igualdad de medias,

$$H_0: \mu_i = \mu_j$$

$$H_1: \mu_i \neq \mu_j$$

Este método consiste en comparar una diferencia entre un par de medias μ_i y μ_j con el estadístico LSD.

$$|\bar{y}_{i.} - \bar{y}_{j.}| > t_{\alpha/2,(N-a)} \times \sqrt{CME\left[\frac{1}{n_i} + \frac{1}{n_j}\right]} \qquad (67)$$

donde el estadístico de esta prueba denominada Mínima Diferencia Significativa se está dada por la siguiente ecuación,

$$\mathrm{LSD} = t_{\alpha/2,(N-a)} \times \sqrt{\mathrm{CME}\left[\frac{1}{n_i} + \frac{1}{n_j}\right]} \qquad (68)$$

Si la diferencia resultase mayor que el estadístico, se concluye que existen diferencias entre los tratamientos.

Ejercicio 4.2

En el problema anterior verifique qué parejas de tratamientos, se consideran estadísticamente diferentes, aplicando la prueba Mínima Diferencia Significativa.

$|\bar{y}_{A.} - \bar{y}_{B.}| = |56.900 - 55.775| = 1.125 < 1.65$

$|\bar{y}_{A.} - \bar{y}_{C.}| = |56.900 - 53.233| = 3.667 > 1.96*$

$|\bar{y}_{A.} - \bar{y}_{D.}| = |56.900 - 51.125| = 5.775 > 1.65*$

$|\bar{y}_{B.} - \bar{y}_{C.}| = |55.775 - 53.233| = 2.542 > 2.14*$

$|\bar{y}_{B.} - \bar{y}_{D.}| = |55.775 - 51.125| = 4.650 > 1.84*$

$|\bar{y}_{C.} - \bar{y}_{D.}| = |53.233 - 51.125| = 2.108 > 1.84*$

El anexo B indica que el valor de $t_{0.025,(12)} = 2.17$, por tanto, de la ecuación (39) los valores LSD son,

$$LSD = 2.17 \times \sqrt{2.88\left[\frac{1}{5} + \frac{1}{4}\right]} = 1.65$$

$$LSD = 2.17 \times \sqrt{2.88\left[\frac{1}{5} + \frac{1}{3}\right]} = 1.96$$

$$LSD = 2.17 \times \sqrt{2.88\left[\frac{1}{4} + \frac{1}{3}\right]} = 2.14$$

$$LSD = 2.17 \times \sqrt{2.88\left[\frac{1}{4} + \frac{1}{4}\right]} = 1.84$$

La prueba LSD muestra una diferencia significativa en las mayorías de las parejas, exceptuando los alerones A y B.

En cuanto a la evaluación de que tanto explicar el modelo.

Es determinado mediante el coeficiente,

$$R^2 = 1 - \frac{SSE}{SSE}$$ 69

denominado coeficiente de determinación, comprendido entre $0 \leq R^2 \leq 1$

Ejercicio 41

Es realizado una competencia experimental de parapentes, en donde se prueba cuatro tipos de materiales, para ello, es recabado los datos de vuelo de los diferentes tipos de materiales. El experimento midió el tiempo de vuelo desde una altura determinada, a su vez cada tiempo se iba anotando para un total de 40 réplicas como muestra la Tabla 56. ¿Existen diferencias en el tiempo de vuelo de los tipos de materiales utilizados en el parapente?

Gráfica 24

El parapente, materiales utilizados para aumentar el tiempo de vuelo

⊙ **Fuente:** elaboración por el Autor

Tabla 56

Tiempo recabado de cada uno de los paracaídas, utilizado en el experimento

	Tipos de materiales			
	A	**B**	**C**	**D**
	1.36	1.41	1.64	1.75
	1.51	1.28	1.49	1.94
	1.55	1.33	1.56	1.86
	1.48	1.37	1.63	1.80
	.	.	.	.
	.	.	.	.
Total	14.64	13.51	15.87	18.29
Promedio	1.464	1.351	1.587	1.829

⊙ **Fuente:** elaboración por el Autor

En este experimento se presentan cuatro niveles $a = 4$ en este experimento, el número total de observaciones $N = 16$.

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N} = 1.43^2 + 1.51^2 + \cdots + 1.84^2 - \frac{62.31^2}{40} = 4.5271$$

$$SCRT = \sum_{i=1}^{a} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N} = \left[\frac{14.64^2}{10} + \frac{13.51^2}{10} + \frac{15.87^2}{10} + \frac{18.29^2}{10}\right] - \frac{62.31^2}{40} = 1.25967$$

$$SSE = SST - SCRT = 4.5271 - 1.25967 = 3.26743$$

Esta información es resumida en la Tabla 57 , el análisis de varianza ANOVA, con el propósito de contrastar si existen diferencias entre los cuatro tipos de materiales.

Tabla 57

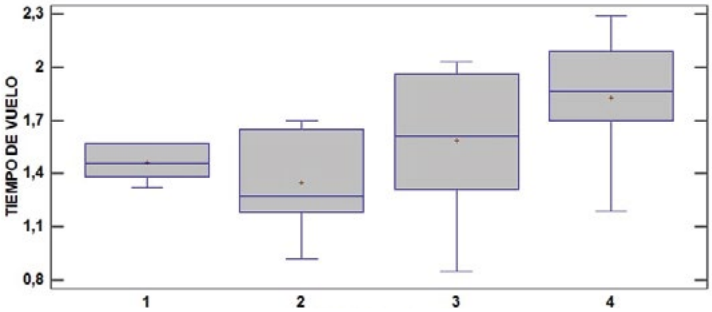
Descomposición de la suma de cuadrados y el contraste de hipótesis

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Valor de F	P-valor
Tipo de material	3	1.2596	0.419889	4.63	0.0077
Error	36	3.26743	0.0907619		
Total	39	4.5271			

⊙ **Fuente:** elaboración por el Autor

El valor del estadístico de prueba $F = 4.63$, esto indica rechazo de la hipótesis nula de igualdad de los tipos de materiales; verificada con el P-valor 0.0077 inferior al nivel de significancia establecido del 5 %..

La muestra indica que existen diferencias estadísticas entre los tipos de materiales, por tanto, el siguiente paso es determinar estadísticamente sus diferencias, encontrar el material más eficiente; el diagrama de caja muestra claramente el tipo de material D, en donde el tiempo de vuelo es el más elevado.



Gráfica 25

Diagrama de box plots para los cuatro tipos de materiales del equipo de parapente

⊙ **Fuente elaboración por el Autor**

En la tabla de análisis de varianza, se muestra un resumen de las varianzas, se evidencia que el cuadrado medio entre grupos o de tratamiento (0.419889) es mucho mayor que el de intragrupal o del error (0.09076), por lo cual se rechaza la hipótesis de igualdad de medias del tiempo de vuelo como variable respuesta. Entre tanto el coeficiente de determinación.

$$R^2 = 1 - \frac{3.26743}{4.5271} = 0.2181$$

Se procede a determinar el R^2 ajustado para indicar que en los datos para probar el tiempo de vuelo de los drones vigías, se ajusta el 21,81 % de la variabilidad en los tiempos de vuelo.

Tabla 52

Prueba LSD de los tipos de materiales para equipo de parapentes

Tipos de materiales			
Nivel	**Medias**	**Grupo**	
B	1.351	X	
A	1.464	X	
C	1.587	XX	
D	1.829		X
.	.	.	.

⊙ **Fuente:** elaboración por el Autor

La Tabla 56, muestra que el tipo de material D, presenta el mayor tiempo de vuelo, es considerado entonces como el mejor material para el parapente.

UNIDAD 8

MODELO DE REGRESIÓN LINEAL SIMPLE

MODELO LINEAL SIMPLE

Es un modelo matemático que relaciona dos variables: la variable y conocida como variable respuesta o variable dependiente y la variable x denominada variable independiente o factor de incidencia. El modelo lineal de regresión posee la siguiente formulación:

$$y_{ij} = \beta_0 + \beta_1 x + \epsilon_{ij} \tag{60}$$

donde β_0 y β_1 son parámetros conocido como intercepto y pendiente respectivamente; $\in$ es el error aleatorio cuyo comportamiento se asume como normal estándar.

Estos parámetros son estimados mediante el Método de los Mínimos Cuadrados utilizando las siguientes formulaciones,

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{61}$$

La determinación de la idoneidad del modelo de regresión con respecto al fenómeno en estudio, es necesario evaluar: a) el coeficiente de determinación R^2, b) la Tabla de análisis de varianza y el coeficiente de correlación:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} \tag{62}$$

En el caso de que el valor de $r = 1$ hay una asociación lineal perfecta con pendiente negativa. El valor de $r = 1$ hay también una asociación lineal perfecta con pendiente negativa y un valor de r está cercano a cero indica que no existe una asociación lineal entre las variables.

El coeficiente de determinación R^2 está definido mediante la siguiente expresión matemática,

$$R^2 = \frac{SSM}{SST} \tag{63}$$

siendo $SSM = \hat{\beta}_1 SS_{xy}$ como la suma de cuadrados de regresión y $SST = SS_{YY}$ es la suma de cuadrados totales. Las sumas de cuadrados para el modelo se establecen de la siguiente forma:

$$SS_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}$$

$$SS_{yy} = SST = \sum_{i=1}^{n} y^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$$

$$SS_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n} \tag{64}$$

donde r es la estimación del coeficiente de correlación, es la suma de cuadrados de SS_{xy} las variables evaluadas, SS_{xx} es la suma de cuadrados de la variable X y SS_{yy} es la suma de cuadrados de la variable Y. La Tabla de análisis de varianza resultante es la que aparece en el siguiente recuadro, siendo p el número de variables que posee el modelo de regresión lineal.

Tabla 47

La suma de cuadrados del modelo de regresión simple

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Valor F
Regresión	1	$SSM = \beta_1 SS_{xy}$	$\frac{SSM}{1}$	$f = \frac{\frac{SSM}{1}}{\frac{SSE}{n-2}}$
Error	$n-2$	$SSE = SST - SSM$	$\frac{SSE}{n-2}$	
Total	$n-1$	$SST = \sum_{i=1}^{n} y^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$		

⊙ **Fuente:** elaboración por el Autor

El contraste de hipótesis plantea $H_0: \beta_1 = 0$, esto implica que rechazar esta hipótesis evidencia un modelo que se ajusta a las condiciones del experimento. Este rechazo ocurre cuando el valor f es mayor que $F_{\alpha,(1,n-2)}$ se considera idóneo el modelo.

EJERCICIO 43

El objetivo de una Entidad Deportiva es obtener un modelo de regresión lineal que involucre las siguientes variables: Número de veces en donde existe una disminución de la velocidad, variable respuesta, y el porcentaje de concentración de oxígeno en el entorno. Para estudiar la relación entre las dos variables se tomaron las observaciones que se encuentran en la Tabla 54, a) obtenga el modelo lineal simple, b) Elabore la Tabla e análisis de varianza.

Tabla 48

Variables relacionadas: conteo de la disminución de la velocidad y porcentaje de oxígeno del entorno

Conteo de la disminución de la velocidad	Concentración de Oxígeno (%) en el entorno
18	60
17	65
15	68
12	72
10	80
6	85
8	87
5	90

⊙ **Fuente:** elaboración por el Autor

La variable dependiente y el conteo en la disminución de la velocidad y la variable independiente x es el porcentaje de concentración de oxígeno en el entorno, aplicando el método de los mínimos cuadrados de la ecuación (61), se tiene que:

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{-369.725}{870.875} = -0.436$$

$$\beta_0 = \bar{y} - \beta_1\bar{x} = 11.375 - (-0.435) \times (75.875) = 44.449$$

Luego la ecuación de línea estimada de la ecuación (60) es:

$$\hat{y} = 44.449 - 0.436x$$

La medida del grado de asociación lineal está dada por el coeficiente de correlación muestral r, el cual se calcula mediante:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{-369.725}{870.875} = 0.98$$

Lo cual indica que hay una buena asociación lineal entre las dos variables; utilizando las ecuaciones (64) para determinar las sumas de cuadrado con el propósito de construir la Tabla de análisis de varianza Tabla 47.

$$SS_{xy} = 6525 - \frac{(607)(91)}{8} = -379.625$$

$$SS_{yy} = SST = 1207 - \frac{(91)^2}{8} = 171.875$$

$$SS_{xx} = 46.927 - \frac{(607)^2}{8} = 870.875$$

Tabla 49

Modelo de regresión simple del conteo en la disminución de la velocidad y la variable independiente el porcentaje de concentración de oxígeno en el entorno

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Valor F
Regresión	1	165.483	165.483	156.11
Error	6	6.38	1.06	
Total	7	171.875		

⊙ **Fuente:** elaboración por el Autor

Lo cual induce a concluir que el modelo de regresión es significativo, ya que el valor f calculado es mayor que el valor obtenido en la Tabla $F_{0.05,(1,6)} = 5.99$.

Ejercicio 47

Una investigación del Ministerio del Deporte evalúa el rendimiento de los atletas considerando dos variables relacionadas; el peso y el tiempo en segundos obtenido. la Tabla 50. ¿Cuál es la capacidad del rendimiento de los atletas?

Tabla 50

Variable peso del Atleta y tiempo de recorrido en segundos

N°	Peso del Atleta(X)	Tiempo en segundo (Y)
1	59.51	14.790
2	58.01	14.640
3	59.00	14.730
4	60.17	14.160
5	61.50	14.970
6	61.02	14.625
7	59.63	15.660
8	60.24	14.025
9	58.95	15.030
10	57.98	14.340
11	60.44	14.700
12	58.91	14.190
13	59.18	15.030
14	59.46	14.640
15	59.51	14.790

⊙ **Fuente:** Elaboración de los autores.

La medida del grado de asociación lineal está dada por el coeficiente de correlación muestral r, el cual se calcula mediante:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}} = \frac{0.34122}{\sqrt{13{,}6746 \times 2.37699}} = 0.063055$$

Lo cual indica que hay una buena asociación lineal entre las dos variables; utilizando las ecuaciones (64) para determinar las sumas de cuadrado con el propósito de construir la Tabla de análisis de varianza Tabla 51.

$$SS_{xy} = 11494993{,}2 - \frac{(893{,}51)(220{,}32)}{15} = 0.34122$$

$$SS_{yy} = SST = 3238{,}43715 - \frac{(220{,}32)^2}{15} = 2{,}37699$$

$$SS_{xx} = 53237{,}6827 - \frac{(893{,}51)^2}{15} = 13.6746933$$

Tabla 51

Modelo de regresión simple del conteo en la disminución de la velocidad y la variable independiente el porcentaje de concentración de oxígeno en el entorno

Fuente	Suma de Cuadrados	Grados de libertad	Cuadrado Medio	Razón-F	Valor-P
Modelo	0,0093785	1	0,0093785	0,05	0,8356
Residuo	2,34939	12	0,195783		
Total	2,35877	13			

Fuente: elaboración por el Autor

Lo cual induce a concluir que el modelo de regresión no es significativo, ya que el valor f calculado es menor que el valor crítico obtenido en la tabla $F_{0.05,(1,12)} = 5.99$. Esta decisión es verificada por el valor P de 0.8356.

Tabla 52

Coeficientes del modelo de regresión simple

	Mínimos Cuadrados	Estándar	Estadístico	
Parámetro	Estimado	Error	T	Valor-P
Intercepto	13,1212	7,12988	1,84031	0,0906
Pendiente	0,0261917	0,11967	0,218867	0,8304

Fuente: elaboración por el Autor

El factor de correlación dio un valor muy bajo (0,063) lo que indica que la relación entre las variables es débil. Por lo tanto, la ecuación de regresión línea, $\text{tiempo de recorrido} = 13.1212 + 0.02619 \times \textbf{peso del atleta}$ no brindará un modelo eficiente para predecir la correspondencia entre estas variables.

UNIDAD 10

TABLAS DE CONTINGENCIA-LOS MODELOS LOGÍSTICOS-CAPACIDAD DE RENDIMIENTO DEPORTIVO

TABLAS DE CONTINGENCIA

La Tabla de contingencia es una tabla varias entradas, la más sencilla por supuesto es la de doble entrada, Tabla 59, donde en cada casilla figurará el número de casos o individuos que poseen un nivel de uno de los factores o características analizadas y otro nivel del otro factor analizado.

Tabla 59
Ejemplo de Tabla de contingencia de doble entrada

		Factor A		
		A1	**A2**	**Total**
Factor B	**B1**	$k_{11}(e_{11})$	$k_{21}(e_{21})$	$k_{1.}(\pi_{1.})$
	B2	$k_{12}(e_{12})$	$k_{22}(e_{22})$	$k_{2.}(\pi_{2.})$
Total		$k_{.1}(\pi_{.1})$	$k_{.2}(\pi_{.2})$	N

⦿ **Fuente:** elaboración por el Autor

Entre las preguntas que se realizan en este tipo de información, se tiene:

¿Esta información sugiere que existe asociación entre el Factor A y B? Para responder a esta pregunta, se realiza una prueba de independencia. En el Factor B se selecciona las categorías B1 y B2 con probabilidades de $\pi_{1.}$ y $\pi_{2.}$ respectivamente. En la variable Factor A **se** seleccionan dos categorías: **A1 y A2,** con probabilidades de $\pi_{.1}$ y $\pi_{.2}$

Formulación de las hipótesis

Las hipótesis nula planteada para las proporciones en la Tabla de contingencia se presenta de la siguiente forma,

$H_0: \pi_{ij} = \pi_{i.}\pi_{.j} \; \forall \; i, j \; Vs \; H_1: \pi_{ij} \neq \pi_{i.}\pi_{.j}$ para al menos un par ij

En la **hipótesis nula** que se plantea la no existencia de asociación entre las variables que resume en la siguiente afirmación: "La presencia del Factor B es independiente al Factor B", contra la **hipótesis alternativa:** de que hay asociación entre las variables. El estadístico de prueba para este contraste de hipótesis es aproximadamente una distribución χ_0^2 con v grados de libertad bajo la hipótesis nula:

$$\chi_0^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}} \tag{70}$$

En este caso la región de rechazo o región crítica es focalizada mediante la siguiente expresión,

$$RC = \{\chi_0^2 : \chi_0^2 > \chi_{1,1-\alpha}^2\} \tag{71}$$

Los valores esperados son calculados en cada una de las celdas de la Tabla de contingencia, mediante la siguiente formulación:

$$e_{11} = \frac{k_{1.} \times k_{.1}}{N} \quad e_{12} = \frac{k_{1.} \times k_{.2}}{N} \quad e_{21} = \frac{k_{2.} \times k_{.1}}{N} \quad e_{22} = \frac{k_{1.} \times k_{.2}}{N} \tag{72}$$

EJERCICIO 44

Una liga de fútbol se centra en dos equipos importantes Rojo y Azul, se realiza un ordenamiento de la información en la Tabla 60. El objetivo inicial es determinar si existe relación entre los resultados obtenidos y el equipo de futbol.

Tabla 60

Contingencia para las variables Equipos x Resultado

		Resultado		
		Ganados	**Perdidos**	Total
Equipos	**Rojo**	273	482	755
	Azul	39	137	176
Total		312	619	931

⊙ **Fuente:** elaboración por el Autor

Los valores esperados son calculados a partir de las ecuaciones (48) presentados en la Tabla 61.

$$e_{11} = \frac{k_{1.x}k_{.1}}{N} = \frac{755 \times 312}{931} = 253.01826 \qquad e_{12} = \frac{k_{1.x}k_{.2}}{N} = \frac{755 \times 619}{931} = 501.98174$$

$$e_{21} = \frac{k_{2.x}k_{.1}}{N} = \frac{176 \times 312}{931} = 58.98174 \qquad e_{22} = \frac{k_{1.x}k_{.2}}{N} = \frac{176 \times 619}{931} = 117.01826$$

Tabla 61

Valores esperados de la Tabla de contingencia para las variables Equipos x Resultado

		Resultado		
		Ganados	**Perdidos**	**Total**
Equipos	**Rojo**	273(253.0182)	482(501.98174)	755
	Azul	39(58.98174)	137(117.01826)	176
Total		312	619	931

⊙ **Fuente:** elaboración por el Autor

Tabla 62

Contingencia para las variables Equipos x Resultado cálculo del estadístico de prueba

n_{ij}	m_{ij}	$(n_{ij} - \hat{m}_{ij})$	$(n_{ij} - \hat{m}_{ij})^2$	$\frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}}$
253,01826	273	19,9817401	399,269936	1,578028147
58,9817401	39	19,9817401	399,269936	6,76938211
501,98174	482	19,9817401	399,269936	0,79538737
117,01826	137	19,9817401	399,269936	3,412031047
			$X^2 \sim \chi^2$	12,55482867
			P valor	0,000395184

⊙ **Fuente elaboración por el Autor**

El cálculo del estadístico X^2 se realiza de la siguiente forma,

$$X^2 \sim \chi_0^2 = \sum_{i=1}^{s}\sum_{j=1}^{s} \frac{(n_{ij} - m_{ij})^2}{m_{ij}} = \frac{(273 - 253.01826)^2}{253.01826} + \dots + \frac{(137 - 117.01826)^2}{114.5} = 12.5548$$

Tabla 63

Valores calculados del Estadístico

	Valor	**gl**	**Sig. asintótica (bilateral)**
Chi-cuadrado de Pearson	12.555a	1	.0003
Corrección por continuidad	11.934	1	.001
Razón de verosimilitudes	13.274	1	.000
Estadístico exacto de Fisher			
Asociación lineal por lineal	12.541	1	.000
N de casos válidos	931		

⊙ **Fuente:** elaboración por el Autor

El valor del estadístico de prueba de chi-cuadrado de Pearson $X^2 = 12.554$ en contraste con $X^2_{1,0.95}$= 3.84 equivalentemente mediante el P-valor $P(X^2_{1,0.95} > 12.554) = 0.0003$. La prueba de **razón de verosimilitud** verifica los resultados anteriores, aplicando la siguiente formulación,

$$G^2 = \sum_{i=1}^{9} -2n_{ij}\ln\frac{m_{ij}}{n_{ij}} \quad (73)$$

Tabla 64

Contingencia para las variables Equipos x Resultado, cálculo razón de verosimilitud.

			$-2n_{ij}\ln\frac{m_{ij}}{n_{ij}}$
Prueba de razón de verosimilitud	n_{ij}	m_{ij}	41,5015339
	253,01826	273	-32,2659682
	58,9817401	39	-39,1573239
	501,98174	482	43,1963563
G^2			13,2745981
P Valor			0,00026911

⊙ **Fuente:** elaboración por el Autor

donde $G^2 > X^2_{1,0.95} = 3.84$, se rechaza la hipótesis nula de asociación entre las variables.

Tabla 65

Medidas simétricas

		Valor	Sig. Aproximada
Nominal por nominal	Phi	0.116	.000
	V de Cramer	0.166	.000
N de casos válidos		931	

⊙ **Fuente:** elaboración por el Autor

De acuerdo con la información arrojada por la Tabla 63, la chi cuadrado de Pearson es igual a 12.554, con un P-valor de 0.0003.

El test de V de Cramer presenta resultados similares: **0.166** (no es muy cercano a 0, indicando presencia de asociación) y un valor de P-valor igual a **0.000.** Como el valor observado de χ^2, pertenece a la región crítica, existen evidencias muestrales para rechazar H_0. Se concluye entonces, con un 95 % de confianza, que el factor **Equipos** no es independiente del Factor

RIESGOS RELATIVOS Y RAZÓN DE VENTAJA (ODDS RATIO)

Tabla 66

Frecuencias en una Tabla de contingencia 2×2

Factor 1	Factor 2 (X)		
	1 (x_1)	2 (x_2)	**Total**
1	k_{11}	k_{12}	$k_{1.}$
2	k_{21}	k_{22}	$k_{2.}$
Total	$k_{.1}$	$k_{.2}$	1

Es un **cociente de ventajas** o **razón de productos cruzados** poblacionales y se denota por ODDS RATIO o OR. Es la razón entre la ventaja a favor de la primera columna en lugar de la segunda columna, para los individuos de la primera fila y la misma ventaja para los individuos de la segunda fila. Esta razón es definida para una Tabla de contingencia dos por dos, ver Tabla 55, de la siguiente forma

$$\mathrm{OR} = \frac{k_{11}}{k_{12}} \times \frac{k_{22}}{k_{21}} > 1 \tag{55}$$

donde debe cumplirse la siguiente condición $\frac{k_{11}/k_{1.}}{k_{12}/k_{1.}} > \frac{k_{21}/k_{2.}}{k_{22}/k_{2.}}$

Luego si $\mathrm{OR} > 1$se puede afirmar que la probabilidad de que un individuo seleccionado al azar se clasifique en la primera columna es mayor para la Fila 1 que para la Fila 2.

El error de estimación en la Tabla de contingencia se define mediante la siguiente ecuación,

$$e.e = \sqrt{\frac{1}{n_{1/1}} + \frac{1}{n_{1/2}} + \frac{1}{n_{2/1}} + \frac{1}{n_{2/2}}} \tag{75}$$

La estimación del intervalo de confianza para un nivel de significancia del 5 % del logaritmo de la razón de ventajas:

$$\left[\mathrm{Ln}\left(\frac{n_{2/1}n_{1/2}}{n_{1/1}n_{2/2}}\right) - 1.96 \times e.e\ ;\ \mathrm{Ln}\left(\frac{n_{2/1}n_{1/2}}{n_{1/1}n_{2/2}}\right) + 1.96.\times e.e\right] \tag{76}$$

Por otra parte, el estimador de la razón de ventajas con un nivel de significancia del 5 % es:

$$\left[e^{\text{lim inferior del Log}}\ ;\ e^{\text{lim superior del Log}}\right] \tag{77}$$

La razón de ventaja obtenida en la ecuación (50), se puede evaluar de la siguiente forma

$$\mathrm{OR} = \frac{p_1/1-p_1}{p_2/1-p_2} = \frac{p_1(1-p_2)}{p_2(1-p_1)} \tag{78}$$

donde $p_1 = {}^{k_{11}}/_{k_{1.}}$ es la probabilidad de que un individuo con valor x_1 de la variable X presente la característica de interés; $p_2 = {}^{k_{21}}/_{k_{2.}}$ es el análogo para un individuo con valor x_2 de la variable X.

EJERCICIO 45

Evalúe los riesgos **relativos** y la **razón de ventaja** ODDS de los datos de la Tabla 60 entre los Factores: Resultados y Equipo; en la liga profesional con dos de los Equipo que se encuentran al final de la Tabla de clasificación en los últimos 10 años.

Solución: Se calcula el riesgo relativo del Factor Resultado con el nivel o modalidad Perdido entre el Equipo Azul en relación al Equipo Rojo, aplicando la ecuación (74):

$$\text{ODDS}_{\text{Perdido}}\,\text{Equipo} = \frac{\text{Color Azul}}{\text{Color Rojo}} = \frac{137}{482} = 0.284$$

Es decir, en cada partido de la columna de Resultado Perdido, existe aproximadamente 0,284 de esta modalidad de Equipo Azul con respecto al Equipo Rojo.

Mientras que la Ventaja o razón ODDS de los partidos de la columna de Resultado Ganado es:

$$\text{ODDS}_{\text{Ganado}}\,\text{Equipo} = \frac{\text{Color Azul}}{\text{Color Rojo}} = \frac{39}{273} = 0.1428$$

Aproximadamente 0.1428 de esta modalidad de Equipo Azul con respecto al Equipo Rojo. La razón de ventaja ODDS, que en este caso se evalúa el cambio en el riesgo de **Equipos,** al pasar de las subpoblaciones con Resultado Perdido a Resultado Ganado.

$$\text{OR} = \frac{\text{ODDS}_{\text{Perdido}}\,\text{Equipo}}{\text{ODDS}_{\text{Ganado}}\,\text{Equipo}}$$

La ventaja u ODDS, seleccionando la columna de **Resultado Ganado**, se define de la siguiente manera,

$$\text{OR} = \frac{\frac{137}{482}}{\frac{39}{273}} = \frac{273 \times 137}{39 \times 482} = 1.9896 \approx 2$$

Es decir, se puede afirmar que los partidos realizados por los dos equipos en la liga de futbol, clasificados en la modalidad de **Resultados Ganados,** tienen casi el doble (2) de posibilidad de ser del **Equipo Azul** que del **Equipo Rojo**.

Ejercicio 46

Evalué el intervalo de confianza del 95 % de las ventajas de los partidos de la liga profesional referenciado en el ejercicio anterior.

Solución: se determina el error del riesgo y el intervalo de confianza de la ventaja con las ecuaciones (75), (77) y (78).

$$e.e = \sqrt{\frac{1}{273}+\frac{1}{39}+\frac{1}{482}+\frac{1}{137}} = \sqrt{0.03867799} = 0.1966672$$

$$\left[\mathrm{Log}\left(\frac{273\times 137}{39\times 482}\right) - 1.96\times 0.196672\,;\ \mathrm{Log}\left(\frac{273\times 137}{39\times 482}\right) + 1.96\times 0.196672\right] = [-0.0867 - 0.385477\,;\ -0.0867 + 0.385477]$$
$$= [-0.4721; 0.2998]$$

Estimador por intervalo de confianza del 95% de la razón de ventajas:

$$[e^{-0.4721};\ e^{0.29877}] = [0.6223;\ 1.3482]$$

Con una confianza del 95 % se puede decir que el verdadero valor del parámetro se encuentra entre los valores 0,6223 y 1,3452. Como el intervalo de confianza para OR contiene la unidad (1), entonces es razonable pensar que el Factor Equipos afecta por igual los partidos con modalidad Resultado Ganado y los partidos con modalidad Perdido; confirmando de esta manera los resultados de la **independencia** entre estas dos variables, que se obtuvo en la prueba chi-cuadrado.

MODELO LOGÍSTICO

El modelo logístico más sencillo, es obtenido en una Tabla de contingencia dos por dos, es decir se ajusta a un modelo de tipo binario tal que la variable Y toma dos variables $Y = 1$ si el individuo presenta la característica de interés e $Y = 0$ en caso contrario. Por lo que se considera la variable respuesta Y sigue una distribución binomial con parámetros n y p con media np. El modelo logístico se realiza mediante la trasformación logit de la medida denominada ventaja o razón de ODDS, obtenida mediante el cociente: $\mathrm{ODDS} = \frac{p}{1-p}$ mediante el logaritmo neperiano de este parámetro,

$$\log\frac{p}{1-p} = \mathrm{logit}(p) = \beta_0 + \beta_1 x_1 + \beta_1 x_2 + \cdots + \beta_n x_n \qquad (79)$$

Modelo logístico binario simple

En el contexto de que solo se dispone de una variable predictora, es decir un solo factor de riesgo el modelo logístico toma la forma,

$$\log\frac{p}{1-p} = \mathrm{logit}(p) = \beta_0 + \beta_1 x \qquad (60)$$

donde p representa la probabilidad de éxito y X es la variable predictora, por lo que la expresión de la ecuación (80) es modificada en su forma equivalente,

$$\frac{p}{1-p} = e^{\beta_0+\beta_1 x} \qquad (81)$$

En la ecuación (81) se despeja el parámetro p por lo que el modelo logístico se puede reescribir de la siguiente forma,

$$E(Y) = p = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

Esta expresión contiene términos no negativos como también el numerador es siempre menor o igual que el denominador, por lo tanto, el cociente variara entre cero (0) y uno (1), rango permitido en las medidas de probabilidad. Las estimaciones máxima verosimilitud de los coeficientes referenciando la Tabla 66 son los siguientes

$$\hat{\beta}_0 = \log \frac{k_{21}}{k_{2.}-k_{21}}$$

$$\hat{\beta}_1 = \log \frac{\frac{k_{11}}{k_{1.}-k_{11}}}{\frac{k_{21}}{k_{2.}-k_{21}}} \quad (82)$$

En el caso específico de evaluar el efecto de dos individuos con valores x_1 y x_2 de la variable X, considerando el modelo expresado en la ecuación (79) se puede expresar el modelo para el primer individuo como

$$\text{logit}(p_1) = \log \frac{p_1}{1-p_1} = \beta_0 + \beta_1 x_1 \quad (83)$$

y, para el segundo individuo

$$\text{logit}(p_2) = \log \frac{p_2}{1-p_2} = \beta_0 + \beta_1 x_2$$

donde p_1 es la probabilidad de que un individuo con valor x_1 de la variable X presente la característica de interés; p_2 es el análogo para un individuo con valor x_2 de la variable X. Al diferenciar los logit de estos dos individuos se tiene entonces,

$$\log \frac{p_1}{1-p_1} - \log \frac{p_2}{1-p_2} = [\beta_0 + \beta_1 x_1] - [\beta_0 + \beta_1 x_2] = \beta_1 [x_1 - x_2]$$

reescribiendo la expresión anterior, por propiedades logarítmicas

$$\log \frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}} = \log \frac{p_1(1-p_2)}{p_2(1-p_1)} = \log(OR) = \beta_1 [x_1 - x_2] \quad (84)$$

considerando un caso particular en donde la diferencia $x_1 - x_2$ de la variable X es la unidad, la expresión de la ecuación (84) se puede presentar como

$$OR = e^{\beta_1} \quad (85)$$

Una de las alternativas para la construcción de un intervalo de confianza para un nivel de significancia del 5 %, está dada por la ecuación (59)

$$\left[e^{\beta_1 + 1.96 \times ee(\beta_1)} \; ; \; e^{\beta_1 - 1.96 \times ee(\beta_1)}\right] \quad (85)$$

Ejercicio 47

Una investigación del Ministerio del Deporte evalúa el bajo rendimiento de los atletas considerando dos niveles de colesterol 240 y 210 mgr como se presenta en la Tabla 56. ¿Cuál es el modelo logístico realizado en el estudio?

Tabla 67

Nivel de colesterol y rendimiento del atleta

		Rendimiento		
		Bajo	**Alto**	**Total**
Nivel del Colesterol	240 mgr.	31	439	470
	210 mgr.	8	246	254
Total		39	685	724

⊙ **Fuente:** elaboración por el Autor

La estimación de la razón de ODDS o ventaja de presentar bajo rendimiento en cada uno de los cohortes; se presenta de la siguiente forma:

La primera fila es decir colesterol de 240 mgr

$$\frac{\hat{p}_1}{1-\hat{p}_1} = \frac{31/470}{439/470} = \frac{0.0660}{0.9340} = 0.0707$$

y, para la segunda fila, colesterol de 210 mgr.

$$\frac{\hat{p}_2}{1-\hat{p}_2} = \frac{8/254}{246/254} = \frac{0.0315}{0.9685} = 0.0325$$

por tanto, una estimación de la razón de ventajas de presentar bajo rendimiento entre los atletas con 240 mgr de colesterol a los que tienen 210 es

$$OR = \frac{\dfrac{31/470}{439/470}}{\dfrac{8/254}{246/254}} = \frac{0.0707}{0.0325} = 2.17$$

La razón de ventaja también se puede determinar utilizando la ecuación (78)

$$OR = \frac{p_1(1-p_2)}{p_2(1-p_1)} = \frac{\left(31/470\right)\left(246/254\right)}{\left(8/254\right)\left(439/470\right)} = \frac{(31)\times(246)}{(8)\times(439)} = 2.17$$

Las estimaciones máxima verosimilitud de los coeficientes basado en la ecuación (82) son los siguientes,

$$\hat{\beta}_0 = \log\frac{8}{254-8} = -3.426 \qquad \hat{\beta}_1 = \log\frac{\frac{31}{470-31}}{\frac{8}{254-8}} = \log(2.17) = 0.775$$

Utilizando la ecuación (83), el modelo logit se define de la siguiente forma,

$$\text{logit}(\hat{p}) = \log\frac{\hat{p}}{1-\hat{p}} = -3.426 + 0.775x_1$$

La razón OR también se puede evaluar mediante la ecuación

$$OR = e^{0.775} = 2.17$$

El valor 2.17 es una estimación del número de veces al que está a más riego el atleta de bajo rendimiento, durante el periodo de estudio, un atleta con 240 mgr de colesterol que otro atleta con 210 mgr. Otra alternativa para la construcción de un intervalo de confianza para un nivel de significancia del 5 %, está dada por la ecuación (77)

$$[e^{0.775-1.96\times 0.4044}\ ;\ e^{0.775+1.96\times 0.4044}]$$

donde $e.e(\hat{\beta}_1) = \sqrt{\frac{1}{31}+\frac{1}{8}+\frac{1}{439}+\frac{1}{246}} = \sqrt{0.1636} = 0.4044$

Los valores del intervalo son $[0.982;\ 4.795]$

Por otra parte, utilizando la ecuación (52) para construir el intervalo de confianza del 95 % se tiene

$$\left[\text{Log}\left(\frac{31\times 246}{8\times 439}\right) - 1.96\times 0.4044;\ \text{Log}\left(\frac{31\times 246}{8\times 439}\right) + 1.96\times 0.4044\right] = [-0.4561-0.7926;\ -0.4561+0.7926]$$
$$= [-1.2487; 0.3365]$$

Estimador por intervalo de confianza del 95% de la razón de ventajas;

$$[e^{-1.2487};\ e^{0.3365}] = [0.2868;\ 1.400]$$

Con una confianza del 95 % se puede decir que el verdadero valor del parámetro se encuentra entre los valores 0,2868 y 1.400. Como el intervalo de confianza para OR contiene la unidad (1), entonces es razonable pensar que el nivel de colesterol afecta por igual el rendimiento de los atletas. En los valores predichos bajo la Tabla 56, están definidos en las ecuaciones

$$\text{logit}(\hat{p}_1) = \log\frac{\hat{p}_1}{1-\hat{p}_1} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Es reemplazada en la anterior expresión el valor de la variable de interés $x_1 = 1$, por lo que el valor predicho del logit es

$$\text{logit}(\hat{p}_1) = \log\frac{\hat{p}_1}{1-\hat{p}_1} = \hat{\beta}_0 + \hat{\beta}_1(1)$$

Estimando de esta manera el $\text{logit}(\hat{p}_1)$ y permite evaluar la probabilidad de ocurrencia del evento

$$\hat{p} = \frac{e^{\hat{\beta}_0+\hat{\beta}_1(1)}}{1+e^{\hat{\beta}_0+\hat{\beta}_1(1)}}$$

permite evaluar el valor esperado del número de individuos que poseen el evento de interés $\hat{p} \times k_{.1}$.

Ejercicio 48

Realice la predicción del número de atletas con bajo rendimiento, sujeto a los niveles de colesterol en su sangre del Ejercicio 47. El modelo logit se define para evaluar la característica de bajo rendimiento del atleta $x_1 = 1$,

$$\text{logit}(\hat{p}) = \log\frac{\hat{p}}{1-\hat{p}} = -3.426 + 0.775(1) = -2.651$$

por lo que la probabilidad de tener bajo rendimiento debido a niveles de colesterol de 240 mgr es

$$\hat{p} = \frac{e^{-2.651}}{1 + e^{-2.651}} = 0.0659$$

Esto que el modelo predice que el 6,59 % de los atletas con bajo rendimiento posee un nivel de colesterol de 240 mgr. Por lo tanto, estos atletas, se espera que, $39 \times 0.0659 = 2.57 \approx 3$ solo tres de ellos, posean niveles de colesterol alto.

PROPUESTA DE INDICADOR DE CAPACIDAD POLINOMIAL DEL RENDIMIENTO DE UN ATLETA

Algunas investigaciones sobre el monitoreo de perfiles lineales hacen referencia a una variable independiente, es decir a un perfil lineal simple, Noorossana R.et al. (2011).

En este modelo lineal se presenta la siguiente ecuación, con un perfil lineal cuadrático o en general polinómico, debido al comportamiento de la variable tiempo muchos (curva de aprendizaje) en muchos de los deportes,

$$y_{ij} = A_0 + A_1 X_i + A_2 X_i^2 + \varepsilon_{ij}\ , \quad i = 1,2,\dots,n \quad , \ j = 1,2,\dots,k$$

donde ε_{ij} son variables aleatorias independientes distribuidas normalmente con media cero y varianza σ^2. Las pendiente A_1, A_2 y la intersección A_0 se denominan en general coeficientes de regresión.

Hosseinifard et al. (2012), se concentró en el índice de capacidad de proceso de un perfil lineal simple bajo el supuesto de no normalidad de la variable respuesta. Introdujeron la distribución Burr XII[1] en la variable de respuesta de cada nivel de la variable explicativa, y luego, usaron el método Clements para calcular C_p, C_{pu}, y C_{pl}, para la variable

1 I. W. Burr, (1973). Parameters for a general system of distribution to match a grid of α3 a α4. Commun Stat. 2:1, 1-21.

respuesta en cada nivel de la variable explicativa. En dicho método el C_{pk} de la variable respuesta se calcula en n niveles de la variable explicativa, y el C_{pk} se introduce como el índice de capacidad del proceso en el perfil lineal simple. El C_p, y C_{pk}, es calculada de la siguiente manera:

$$C_p = \frac{1}{6}[\phi^{-1}(1 - P_U) - \phi^{-1}(P_L)]$$

$$C_{pk} = \frac{1}{3}\min[\phi^{-1}(1 - P_U) - \phi^{-1}(P_L)] \quad (87)$$

donde P_U se estima usando la ecuación (70) y (71)

$$P_U = 1 - \prod_{i=1}^{n} P_r(y_{ij} < USL),$$

entonces:

$$P_U = 1 - \prod_{i=1}^{n} \phi(\frac{USL_i - \mu_i}{\sigma}) \quad (88)$$

P_L se estima con la ecuación (72) y (73)

$$P_L = 1 - \prod_{i=1}^{n} P_r(y_{ij} > LSL)$$

por consiguiente

$$P_L = 1 - \prod_{i=1}^{n}\left[1 - \phi\left[\frac{lei_i - \mu}{\sigma}\right]\right] \quad (89)$$

μ y σ en las ecuaciones (88) y (89) son la media y la desviación estándar de la variable de respuesta en diferentes niveles de la variable explicativa.

El índice de capacidad del proceso C_p que se definió en la ecuación (87) es una comparación entre los límites de tolerancia natural y los límites de especificación de un proceso. En un perfil lineal cuadrático por ejemplo $y = A_0 + A_1X + A_2X^2$ es la línea de referencia del proceso, $a_0 + a_1x + a_2x^2$ es la media condicional de y y x, entonces, μ se calcula de la siguiente manera $\mu = a_0 + a_1x + a_2x^2$.

La variable aleatoria normal Y posee una media de $a_0 + a_1x + a_2x^2$ y varianza de σ^2, y a_0, a_1 y a_2 son estimaciones de A_0, A_1 y A_2; se determinan con las formulaciones: $a_0 = (\sum_j^k a_{0j})/k$ $a_1 = (\sum_j^k a_{1j})/k$ y $a_2 = (\sum_j^k a_{2j})/k$ respectivamente. a_{0j}, a_{1j} y a_{2j} son los interceptos y las pendientes estimadas en el j-ésimo perfil de muestra.

La varianza del proceso σ^2 se estima utilizando MSE, calculada como $\mathrm{MSE} = (\sum_j^k \mathrm{MSE_j})/k$, donde, $\mathrm{MSE_j}$ es la varianza estimada en el j. ésimo perfil de muestra. Por lo tanto, el UNTL y LNTL obtenida las siguientes ecuaciones,

$$\mathrm{UNTL_j} = \mu + 3\sigma = a_0 + a_1 x + a_2 x^2 + 3\sigma$$

$$\mathrm{LNTL_j} = \mu + 3\sigma = a_0 + a_1 x + a_2 x^2 - 3\sigma \qquad (90)$$

UNTL y LNTL son líneas paralelas con distancia entre ellos a 6σ, es conocido μ, UNTL y LNTL son funciones de x, como $\mu_y(x) = a_0 + a_1 x + a_2 x^2$, $\mathrm{UNTL_y}(x) = a_0 + a_1 x + a_2 x + 3\sigma$ y $\mathrm{LNTL}(x) = a_0 + a_1 x + a_2 x - 3\sigma$. Suponga que los límites de especificación de la variable y son dos funciones de x, tal como las obtiene las ecuaciones.

$$\mathrm{USL_y}(x) = a_{0u} + a_{1u} x + a_{2u} x^2$$

$$\mathrm{LSL_y}(x) = a_{0u} + a_{1u} x + a_{2u} x^2 \qquad (91)$$

Actualmente, C_P de un perfil lineal simple tiene una forma funcional, como se presenta en la siguiente ecuación,

$$C_{p(x)} = \frac{\mathrm{USL_y}(x) - \mathrm{LSL_y}(x)}{\mathrm{UNTL_y}(x) - \mathrm{LNTL}(x)}$$

$$x \in [x_1, x_2] \qquad (92)$$

Al usar $C_{p(x)}$ como el índice de capacidad del proceso del perfil lineal simple, es posible evaluar la capacidad en cada nivel de x. la capacidad del proceso en cada nivel de la variable explicativa propone información detallada del proceso. Sin embargo, es necesario tener un valor único del índice de capacidad del proceso para un perfil lineal simple en todos los rangos de la variable explicativa para dar un juicio general sobre la capacidad del proceso.

Es recomendable utilizar el área limitada entre USL_y y LSL_y para calcular $\mathrm{USL_y}(x)$ y $\mathrm{LSL_v}(x)$ y, también el área limitada entre $\mathrm{UNTL_y}$ y $\mathrm{LNTL_y}$ para calcular $\mathrm{UNTL_y}(x)$ y $\mathrm{LNTL}(x)$. Por lo que lo propuesto por los autores para determinar un valor único para el C_p de un perfil lineal simple es:

$$C_{p(\mathrm{profile})} = \frac{\int_{x_l}^{x_u} [\mathrm{USL_y}(x) - \mathrm{LSL_y}(x)]dx}{\int_{x_l}^{x_u} [\mathrm{UNTL_y}(x) - \mathrm{LNTL}(x)]dx}$$

$$x \in [x_l, x_u] \qquad 93$$

$\mathrm{UNTL_y(x)}$ y $\mathrm{LNTL(x)}$ son dos líneas paralelas. Se supone como $\mathrm{USL_y(x)}$ y $\mathrm{LSL_y(x)}$ son dos líneas paralelas como $\mathrm{USL_y(x)} = \mathrm{a_{0u}} + \mathrm{a'_1 x} + \mathrm{a'_2}x^2$, y $\mathrm{LSL_y(x)} = \mathrm{a_{0l}} + \mathrm{a'_1 x} + \mathrm{a'_2}x^2$.

donde $\mathrm{a_{0u}}$, $\mathrm{a_{0l}}$, son las intercepciones de $\mathrm{USL_y(x)}$ y $\mathrm{LSL_y(x)}$, $\mathrm{a'_1}$, $\mathrm{a'_2}$ las pendientes de $\mathrm{USL_y(x)}$ y $\mathrm{LSL_y(x)}$. La distancia de estas líneas paralelas se puede considerar como su diferencia.

Por lo que $C_{p(profile)}$ se calcula de la siguiente manera:

$$\mathrm{C_{p(profile)}} = \frac{\mathrm{a_{0u}} - \mathrm{a_{0l}}}{6\sigma}$$

$\mathrm{C_{pk}(x)}$ se determina en la siguiente ecuación:

$$\mathrm{C_{pk}(x)} = \min\left\{\frac{\mathrm{USL_y(x)} - \mu_y(\mathrm{x})}{\mathrm{UNTL_y(x)} - \mu_y(\mathrm{x})}, \frac{\mu_y(\mathrm{x}) - \mathrm{LSL_y(x)}}{\mu_y(\mathrm{x}) - \mathrm{LNTL(x)}}\right\} \tag{94}$$

$x \in [x_l, x_u]$

$\mu_y(x)$, es la función de la línea de referencia $C_{pk}(x)$, genera el valor de $\mathrm{C_{pk}}$ de un proceso simple para cada nivel de x. El $\mathrm{C_{pk}}$ de un perfil lineal simple, se calcula:

$$\mathrm{C_{pk\,(profile)}} = \min\left\{\frac{\int_{x_l}^{x_u}[\mathrm{USL_y(x)} - \mu_y(\mathrm{x})]dx}{\int_{x_l}^{x_u}[\mathrm{UNTL_y(x)} - \mu_y(\mathrm{x})]dx}; \frac{\int_{x_l}^{x_u}[\mu_y(\mathrm{x}) - \mathrm{LSL_y(x)}]dx}{\int_{x_l}^{x_u}[\mu_y(\mathrm{x}) - \mathrm{LNTL(x)}]dx}\right\} \tag{95}$$

El índice de capacidad del proceso ($\mathrm{C_{pk}}$) cuando solo se encuentra disponible los límites de especificación funcional superior o inferior, se puede calcular mediante las siguientes ecuaciones:

$$\mathrm{C_{pu(profile)}} = \frac{\int_{x_l}^{x_u}[\mathrm{USL_y(x)} - \mu_y(\mathrm{x})]dx}{\int_{x_l}^{x_u}[\mathrm{UNTL_y(x)} - \mu_y(\mathrm{x})]dx}$$

$$\mathrm{C_{pl(profile)}} = \frac{\int_{x_l}^{x_u}[\mu_y(\mathrm{x}) - \mathrm{LSL_y(x)}]dx}{\int_{x_l}^{x_u}[\mu_y(\mathrm{x}) - \mathrm{LNTL(x)}]dx}$$

Si $\mathrm{USL_y(x)}$ es más grande que $\mu_y(x)$ en $[x_l, x_m]$ y menor que $\mu_y(x)$ en $[x_m, x_u]$.

Entonces el índice de capacidad mínimo se calcula de la siguiente manera:

$$\mathrm{C_{pk\,(profile)}} = \min\left[\left\{\mathrm{C_{pu\,(profile)}} = \frac{\int_{x_l}^{x_m}[\mathrm{USL_y(x)} - \mu_y(\mathrm{x})]dx - \int_{x_m}^{x_u}[\mu_y(\mathrm{x}) - \mathrm{USL_y(x)}]dx}{\int_{x_l}^{x_u}[\mathrm{UNTL_y(x)} - \mu_y(\mathrm{x})]dx}, \mathrm{C_{pl\,(profile)}} = \frac{\int_{x_l}^{x_u}[\mu_y(\mathrm{x}) - \mathrm{LSL_y(x)}]dx}{\int_{x_l}^{x_u}[\mu_y(\mathrm{x}) - \mathrm{LNTL(x)}]dx}\right\}\right]$$

Ejercicio 47

Una investigación del Ministerio del Deporte evalúa el rendimiento de los atletas considerando dos variables relacionadas; el peso y el tiempo en segundos obtenido. la Tabla 68. ¿Cuál es la capacidad del rendimiento de los atletas?

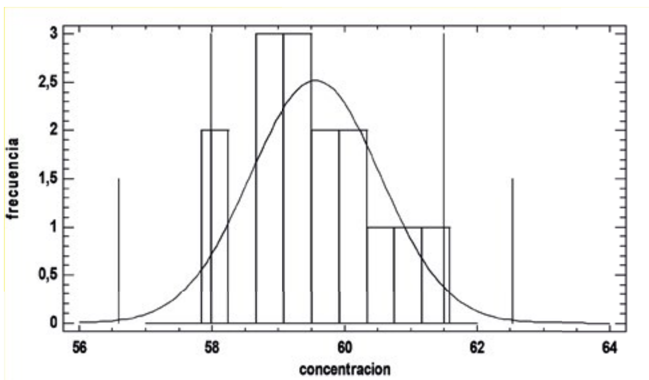
Tabla 68

Variable peso del Atleta y tiempo de recorrido en segundos

N°	Peso del Atleta(Y)	Tiempo en segundo (X)
1	59.51	14.790
2	58.01	14.640
3	59.00	14.730
4	60.17	14.160
5	61.50	14.970
6	61.02	14.625
7	59.63	15.660
8	60.24	14.025
9	58.95	15.030
10	57.98	14.340
11	60.44	14.700
12	58.91	14.190
13	59.18	15.030
14	59.46	14.640
15	59.42	14.640

⊙ **Fuente:** Elaboración de los autores.

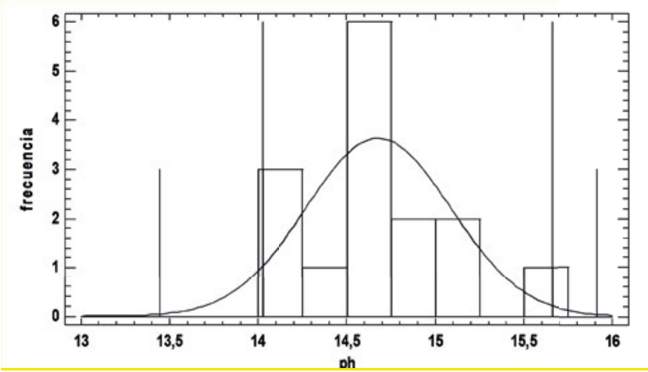
Inicialmente son determinados los índices de capacidad univariados tradicionales implementando la ecuación (87) para cada una de las variables. El peso de los atletas el C_p es de 0,5932 y el C_p del tiempo de recorrido del atleta de 0,6626.



Gráfica 26

Índice de capacidad para la variable pero del atleta $C_p = 0,5932.$

⊙ **Fuente:** elaboración propia de los autores.



Gráfica 27

Índice de capacidad para la variable tiempo de recorrido $C_p = 0,6626.$

⊙ **Fuente:** elaboración propia de los autores.

Los índices de capacidad que arrojaron las variables menores a la unidad, por lo que el proceso no cumple con las especificaciones de diseño.

El factor de correlación dio un valor muy bajo (0,0038) lo que indica que la relación entre las variables es débil. Por lo tanto, la ecuación de regresión línea de $\text{tiempo de recorrido} = 57.37 + 0.14\text{peso del atleta}$, o, en otras palabras, $\text{tiempo de recorrido} = 57.37 + 0.14\textbf{peso del atleta}$ no brindará un modelo eficiente para predecir la correspondencia entre estas variables.

El índice de capacidad lineal, referente a los límites de especificación superior e inferior (USL y LSL), así como los límites de proceso superior e inferior (UNTL Y LNTL), es determinada con los intervalos de confianza, los cuales se presentan en la Tabla 69:

Tabla 69

Intervalos de confianza, variable pH y concentración de un producto alimenticio

Parámetro	Estimación	Error Estándar	Límite Inferior	Límite Superior
Constante	57,3712	9,77517	36,2532	78,4892
Tiempo de recorrido	0,149211	0,665731	-1,28902	1,58744

⊙ **Fuente:** Elaboración propia de los autores.

Obteniendo las ecuaciones $\text{USL} = 78.48 + 1{,}58744\text{ peso del atleta}$ y $\text{LSL} = 36.253 - 1.289\text{peso del atleta}$, siendo μ la ecuación de la variable y, σ es la raíz del cuadrado medio del error la cual es 0.8717.

Los límites del proceso son: límite superior $\text{UNTL} = 78.48 + 0.1492\text{x}$ y el límite inferior $\text{LNTL} = 36.25 + 0.1492\text{x}$.

El índice de capacidad del proceso basado en las especificaciones de la variable de tiempo de recorrido es $LSL = 2.4$ y $USL = 3$, el índice de capacidad del proceso es:

$$C_{p(profile)} = \frac{\int_{2.4}^{3}[78.489 + 1.587x - 36.253 + 1.289x]\,dx}{\int_{2.4}^{3}[78.489 + 0.149x - 36.2532 - 0.149x]\,dx} = 1.183$$

indica un cumplimiento satisfactoriamente con respecto a las especificaciones establecidas. Es determinado el $C_{pk\,(profile)}$, con el fin de identificar hacia donde se encuentran desplazado los tiempos de recorrido.

$$C_{pk\,(profile)} = \min\left[C_{pu\,(profile)} = \left\{\frac{\int_{2,4}^{3}[78{,}4892+1{,}58744x - 57{,}3712-0{,}14211x]\,dx}{\int_{2,4}^{x_u}[78{,}4892+0{,}149211x-57{,}3712-0{,}14211x]\,dx}, C_{pl(profile)} = ,\ \frac{\int_{2,4}^{3}[57{,}3712-0{,}14211x-36{,}2532+1{,}28902x]\,dx}{\int_{2,4}^{3}[57{,}3712-0{,}14211x-36{,}2532-0{,}149211x.]\,dx}\right\}\right]$$

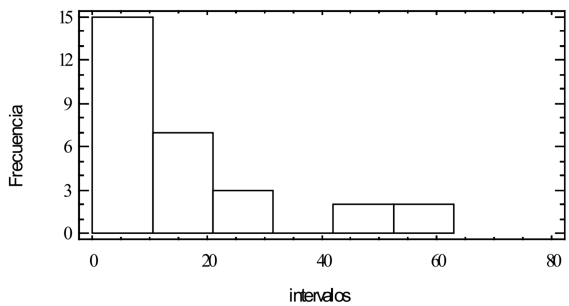
por tanto

$$C_{pk\,(profile)} = \min\{1.1; 1.1\}$$

El $C_{pk\,(profile)}$ indica un rendimiento ajustado a los lineamientos planificados en el entrenamiento.

ACTIVIDADES EVALUATIVAS

El siguiente histograma indica que el modelo se puede asumir como:



Gráfica 28
Histograma de frecuencia modelo supuesto

⊙ **Fuente:** elaboración realizada por el Autor

a. Normal
b. Variable
c. Exponencial
d. Bimodal

En la siguiente Tabla de frecuencia determine cuál es el valor de la frecuencia acumulada.

Tabla 70
Tabla de frecuencia

Intervalo	**Punto medio** x_i	**Frecuencia** f	**Frecuencia relativo** f_r	**Frecuencia acumulada** F	**Frecuencia relativa acumulada** F_r
10-15	12.5	2	0.20	2	0.20
15-20	17.5	5	0.50		0.70
20-25	22.5	3	0.30	10	1.00

⊙ **Fuente:** elaboración por el Autor

a. 10

b. 3

c. 8

d. 7

¿Cuál es el ángulo de conversión que se ajusta en la tercera categoría evaluada en la siguiente Tabla de frecuencia para construir un diagrama de torta?

Tabla 71

Ángulo de conversión

Categorías	Frecuencia	Ángulo
Excelente calidad	50	207
Buena calidad	30	
Calidad regular	5	21
Mala calidad	2	8
Total	87	

⊙ **Fuente:** elaboración por el Autor

a. 100

b. 124

c. 85

d. 167

En la siguiente Tabla se presenta la información de las características de calidad de una unas zapatillas deportivas. Se desea determinar el coeficiente de correlación entre estas variables.

Tabla 72

Información de las características de calidad de una unas zapatillas deportivas

Diámetro interno y	Resistencia x
1.24	145
1.25	143
1.23	144
1.24	143

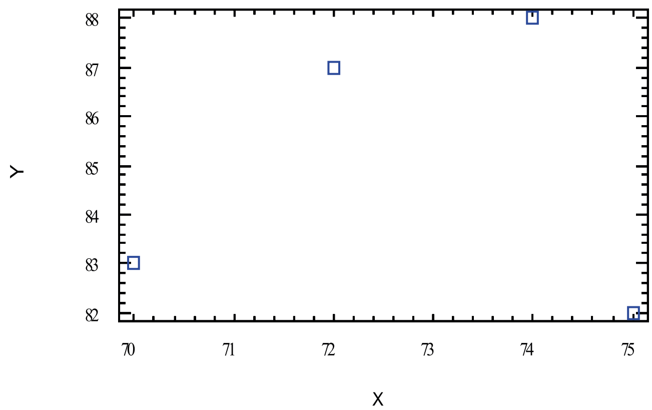
a. 0.50

b. -0.12

c. 0.85

d. -0.42

En el siguiente diagrama de correlación representa el tiempo y los kilometrajes de un corredor; determine el tipo de correlación existente.



Gráfica 29

Diagrama de correlación entre el tiempo y kilómetros recorrido

Fuente elaboración realizada por el Autor

a. Positiva

b. Negativa

c. No existe asociación entre las variables

La siguiente Tabla de frecuencia muestra el conteo de los kilómetros recorridos por un ciclista en un circuito contrarreloj. Determine qué porcentaje de los ciclistas posee un recorrido por debajo de los 206 kilómetros.

Tabla 73

Kilómetros recorridos por un ciclista en un circuito contrarreloj

Intervalo	Punto medio x_i	Frecuencia f	Frecuencia relativo f_r	Frecuencia acumulada F	Frecuencia relativa acumulada F_r
200-202	201	3	0.10	3	0.10
202-204	203	10	0.33	13	0.43
204-206	205	14	0.46	27	0.90
206-208	207	3	0.10	30	1.0

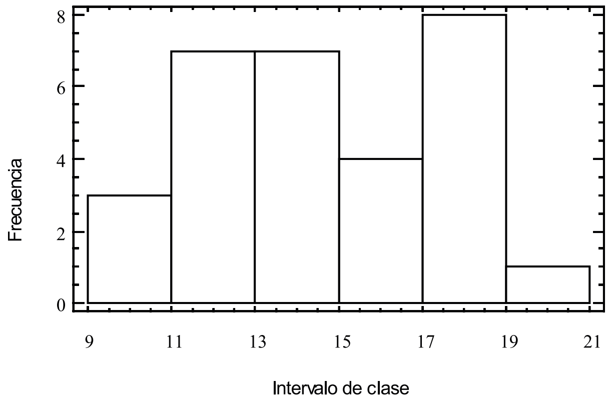
a. 50 %

b. 90 %

c. 10 %

d. 22 %

En el siguiente histograma indique cual es la frecuencia relativa del tercer intervalo de clase



Gráfica 30

Histograma de frecuencia de una información deportiva

⊙ **Fuente:** elaboración realizada por el Autor

a. 0.66

b. 0.23

c. 0.35

d. 0.40

Según se observa en la siguiente Tabla de frecuencia, cuál es el porcentaje de incidencia del número de errores de un equipo nacional de patinaje.

Tabla 74

Tabla de frecuencia de errores en un equipo de patinaje

Característica	Frecuencia
No hacer chequeos médicos regulares	25
Entrenar solo la resistencia, y no la fuerza	17
Entrenar siempre en circuitos grandes	2
No darle suficiente importancia a la técnica	6
Total	

⊙ **Fuente:** elaboración por el Autor

a. 25 %

b. 42 %

c. 18 %

d. 34 %

En la Tabla 75 determine ¿cuál de las siguientes categorías de carreras de automóvil posee un ángulo de conversión erróneamente calculado?

Tabla 75
Categorías de carreras de automóvil

Categorías	Frecuencia	Ángulo
Monoplazas	35	126
Rally	15	57
Carrera de montaña	50	180
Total	100	

⊙ **Fuente:** elaboración por el Autor

a. Monoplazas

b. Rally

c. Carrera de montaña

Se realizó el inventario de los productos alimenticios ofrecidos a una delegación deportiva departamental, encontrándose las siguientes fallas en el almacenamiento.

Tabla 76
Inventario de los productos alimenticios ofrecidos a una delegación deportiva departamental

Tipo de falla	Frecuencia	Precio de la falla por unidad
Mal almacenado	25	5
Caducidad de producto	32	10
Etiqueta borrosa	47	7
Dañó en el empaque	6	15
Total	110	

⊙ **Fuente:** elaboración por el Autor

Si se tiene en cuenta los costos de falla, qué porcentaje aporta la falla de la caducidad del producto.

a. 100

b. 124

c. 85

d. 167

En un campo deportivo se llevó a cabo un proyecto para mejorar el recorrido de los jugadores de un equipo de fútbol. Se desea monitorear estas actividades mediante un comportamiento gráfico utilizando el histograma de frecuencia.

Tabla 77

El recorrido en kilómetros de los jugadores de un equipo de fútbol

6.0	5.8	6.1
5.2	6.4	6.9
5.5	5.8	5.2
5.0	5.7	6.5
6.7	6.5	5.5
5.8	5.2	5.0
5.6	5.1	5.2
6.0	5.8	6.0
5.5	4.9	5.7

⊙ **Fuente:** elaboración por el Autor

¿Qué puede decir del comportamiento de los datos? Realice las conclusiones apropiadas. Asuma una distribución normal, evalué la probabilidad de que el recorrido de un jugador esté por encima de 7.0 kilómetro en el partido.

Esta información es recolectada de tres máquinas de bateo, en donde se mide la velocidad de la pelota.

Tabla 78

Información de la velocidad de las máquinas de bateo

	MAQUINA UNO			MAQUINA DOS			MAQUINA TRES		
1	253.921	241.582	265.27	257.596	249.071	251.981	248.401	247.706	251.055
2	252.737	242.684	244.039	260.321	249.671	259.723	248.867	249.572	249.377
3	247.741	243.876	258.582	256.257	256.636	254.387	248.121	252.331	252.495
4	243.375	251.334	244.508	255.115	254.207	256.244	249.999	247.5	250.469
5	246.965	250.114	251.109	246.667	257.146	258.925	246.548	249.736	250.381
6	243.089	232.618	254.782	258.166	251.914	250.263	253.332	250.562	246.194
7	252.151	247.59	249.155	248.152	252.222	255.578	248.035	247.238	249.464
8	244.049	252.354	260.716	257.866	256.078	258.456	248.891	253.992	253.231
9	244.555	244.04	247.375	253.519	258.513	254.674	252.393	251.984	248.945
10	260.236	246.303	242.79	256.352	255.899	255.949	247.557	249.112	248.679
11	255.823	238.399	241.24	253.885	254.637	259.436	249.728	249.663	242.471
12	252..224	244.322	246.965	255.772	256.373	248.522	253.755	252.5	250.843
13	245.397	257.707	256.807	254.653	255.751	259.114	254.005	249.664	253.774
14	256.565	242..563	247.886	254.801	248.828	254.219	252.992	247.634	254.412
15	247.419	249.912	253.94	253.157	258.125	257.273	251.226	250.458	249.637
16	249.012	251.358	243.46	254.182	249.722	248.727	249.64	246.266	249.702
17	246.681	258.316	256.192	258.434	250.191	257.825	246.469	248.299	246.918
18	257.847	258.524	255.222	256.088	256.444	257.666	249.714	254.477	249.03
19	246.312	245.88	245.628	251.6	253.84	256.207	253.169	254.387	244.158
20	245.487	236.964	249.258	252.379	254.646	250.341	248.627	249.927	247.703

⊙ **Fuente:** elaboración por el Autor

a. Construya la Tabla de frecuencia de cada una de las máquinas y obtenga el histograma de frecuencia. ¿Cuál de las maquinas posee un comportamiento tipo normal?

b. Elabore un intervalo de confianza para cada una de las máquinas de bateo, con nivel de significancia del 5 %.

c. Mediante un análisis de varianza, además utilizando la prueba LSD determine diferencia entre las máquinas de bateo.

La presente información son los costos asociados a la contratación de jugadores de béisbol en cuatro ciudades de la costa norte de Colombia.

Tabla 79

Costos asociados a la contratación de jugadores de béisbol en cuatro ciudades de la costa norte de Colombia

Ciudad	Ventas en millones de dólares
Montería	1.25
Barranquilla	2.56
Cartagena	3.53
Sincelejo	1.82

⊙ **Fuente:** elaboración por el Autor

¿Construya un diagrama de torta en el cual se establezca el porcentaje de venta de cada una de las ciudades?

Un Estudio de Bienestar Universitario desea evaluar la presencia o no de contusiones físicas de dos de los deportes más comunes en el campus: Futbol y el Baloncesto, con el propósito de realizar una predicción de los estudiantes que sufrirán lesiones y permita planificar la atención médica.

Tabla 80

Contusiones de los estudiantes universitarios en Fútbol y Baloncesto

		Presencia de contusión		
		Si	No	Total
Deporte	Fútbol	11	213	224
	Baloncesto	6	85	91
Total		17	298	315

⊙ **Fuente:** elaboración por el Autor

Elabore un modelo logístico para la predicción de las contusiones de los estudiantes universitarios inscritos en fútbol y baloncesto.

Referencias Bibliográficas

STURGES, H. (1926). The choice of a class-interval. J. Amer. Statist. Assoc., 21, 65–66.

SCOTT, D.W. (1979). On optimal and data-based histograms. Biometrika, 66, 605–610.

WAND, M.P. (1995). Data-based choice of histogram bin-width. Technical report, Australian Graduate School of Management, University of NSW.

BAYES, (1763). An Essay towards solving a Problem in the Doctrine of Chances. Philosophical Transactions of the Royal Society of London 53: 370-418. doi:10.1098/rstl.1763.0053

ANDERSON, D. SWEENEY D. y Williams, T. (1982, 2005). Estadística para administración y economía. México: Thomson editores.

ASCHNER P. Y RUIZ Á. (2012). Prevalencia de Obesidad Abdominal y Factores de Riesgo asociados en atención primaria en Colombia. Biomédica, Vol. 12, núm. 4.

BLANCO C. L. (2004). Probabilidad. Universidad Nacional de Colombia. Bogotá: UNIBIBLOS.

BLANCO CATAÑEDA, L. (2004). Probabilidad. Colección Texto. Universidad Nacional de Colombia, Bogotá.

CANAVOS G. (1993). Probabilidad y Estadística, Aplicaciones y métodos. México: McGraw-Hill

CHISTENSEN, H. (1990). Estadística paso a paso. México: Trillas 3era edición.

DE LA HORRA, J. (2003). Estadística aplicada. Ediciones Díaz de Santos.

DEVORE, JAY L. (2001). Probabilidad y estadística para ingeniería y ciencias. Internacional Thompson. International Thomson Editores, S. A. de C. V.

ÁVILA BARAY, H.L. (2006) Introducción a la metodología de la investigación. Edición electrónica. Texto completo en: http://www.eumed.net/libros-gratis/2006c/203/index.htm

GARZO, F. Y GARCíA, F. (1988). Estadística. España: McGraw-Hill Interamericana.

HILDEBRAND, D. K. & OTT, LYMAN R. (1997). Estadística aplicada a la administración y la economía. México: Addison-Wesley Iberoamericana.

JOHNSON. R. (1999). Estadística elemental: Lo esencial. Segunda Edición. México: Thomson.

KING H, AUBERT RE, HERMAN W. (1998). Global Burden of Diabetes 1995-2025. Diabetes Care, Vol. (21): p. 1414-1431.

LEVINE. DAVID M. (2006). Estadística para administración. Séptima Edición. Pearson: México.

LIND, D. A.; MARCHAL, W. G. & MASON, R. D. (2004). Estadística para administración y economía. Alfaomega Colombiana, S.A. Colombia.

LIND. MARCHAL Y WATHEN. (2001). Estadística aplicada a los negocios y a la economía. 12º Edición. México: Mc Graw Hill.

LINDSTROM J. Y TUOMILEHTO J. (2008). Determints for the Effectives of Lifestyle Intervention in the Finnish Diabetes Preventions study. Diabetes Care. 31(5). 857-862.

MENDENHALL, WILLIAM. (1990). Estadística para Administración. México: Grupo Editorial Iberoamérica.

MILTON. J.S. (1994). Estadística para biología y ciencias de la salud. P.C. Puig (trad.). 2da Ed. Madrid Editorial Interamericana-McGraw-Hill.

MONTGOMERY DOUGLAS. (2002). Diseño y Análisis de Experimentos. México: Editorial Limusa.

PAGANO, ROBERT R. (1999). Estadística para las ciencias del comportamiento. Séptima edición. Thomson.

PEÑA SANCHEZ DE RIVERA, DANIEL. (1988). Estadística Modelos y Métodos. México: Editorial Alianza Editorial textos.

ROBERTO J., HERRERA ACOSTA T., & FONTALVO HERRERA, J. (2011). Seis Sigma: un enfoque práctico. Corporación para la gestión del conocimiento. Asesores del 2000. Bogotá 129. ISBN: 9789589973714.

D.C. SPIEGEL. M.R. (1991). Estadística. R. Hernández (trad.). Madrid: Editorial McGraw-Hill-Interamericana.

SOTE, A. (2005) Principios de Estadística. Caracas: Panapo de Venezuela.

WALPOLE, RONALD E. & MYERS, RAYMOND H. (1998). Probabilidad y estadística. México: Prentice Hall Latinoamericana.

SAHRIARI, H. y SARAFIAN, M. (2009). Evaluación del proceso - Evaluación capacidad en lineal perfiles lineas, Internacional de Ingeniería Industrial Conferencia, Teherán, Irán.

EBADI M, SHAHRIARI H., (2012). A process capability index for simple linear profile. International Journal of Advanced Manufacturing Technology; doi:10.1007/s00170-012-4066-7

MESTEK, O., PAVLIK, J. and SUCHANEK, M. (1994). Multivariate control charts: Control charts for calibration curves. *Journal of Analytical Chemistry, 350*(6), pp. 344-351.

HOSSEINIFARD, S. y Abbasi, B. (2012). Evaluación de índices de capacidad de con lineal perfiles. *Revista Internacional de Gestión de Calidad y Confiabilidad, 29*(2), pp. 162-176.

BURR, W. (1973). Parameters for a general system of distribution to match a grid of α3 a α4. Commun Stat. 2:1, 1-21.

ANEXOS

Anexo 1. Distribucion normal estándar acumulada

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-4.5	0,000003	0,0000032	0,0000031	0,0000029	0,0000028	0,0000027	0,0000026	0,0000024	0,0000023	0,0000022
-4.4	0,000004	0,0000052	0,0000049	0,0000047	0,0000045	0,0000043	0,0000041	0,0000039	0,0000037	0,0000036
-4.3	0,000008	0,0000082	0,0000078	0,0000075	0,0000071	0,0000068	0,0000065	0,0000062	0,0000059	0,0000057
-4.2	0,000013	0,0000128	0,0000122	0,0000117	0,0000112	0,0000107	0,0000102	0,0000098	0,0000093	0,0000089
-4.1	0,000020	0,0000198	0,0000189	0,0000181	0,0000174	0,0000166	0,0000159	0,0000152	0,0000146	0,0000139
-4.0	0,000031	0,0000304	0,0000291	0,0000279	0,0000267	0,0000256	0,0000245	0,0000235	0,0000225	0,0000216
-3.9	0,000048	0,0000461	0,0000443	0,0000425	0,0000407	0,0000391	0,0000375	0,0000359	0,0000345	0,0000330
-3.8	0,000072	0,0000695	0,0000667	0,0000641	0,0000615	0,0000591	0,0000567	0,0000544	0,0000522	0,0000501
-3.7	0,000107	0,0001036	0,0000996	0,0000957	0,0000920	0,0000884	0,0000850	0,0000816	0,0000784	0,0000753
-3.6	0,000159	0,0001531	0,0001473	0,0001417	0,0001363	0,0001311	0,0001261	0,0001213	0,0001166	0,0001121
-3.5	0,000232	0,0002241	0,0002158	0,0002078	0,0002001	0,0001926	0,0001854	0,0001785	0,0001718	0,0001653
-3.4	0,000336	0,0003248	0,0003131	0,0003018	0,0002909	0,0002803	0,0002701	0,0002602	0,0002507	0,0002415
-3.3	0,000483	0,0004665	0,0004501	0,0004342	0,0004189	0,0004041	0,0003897	0,0003758	0,0003624	0,0003495
-3.2	0,000687	0,0006637	0,0006410	0,0006190	0,0005976	0,0005770	0,0005571	0,0005377	0,0005190	0,0005009
-3.1	0,000967	0,0009354	0,0009043	0,0008740	0,0008447	0,0006992	0,0007888	0,0007622	0,0007364	0,0007114
-3.0	0,001349	0,0013062	0,0012639	0,0012228	0,0011829	0,0011442	0,0011067	0,0010703	0,0010350	0,0010008
-2.9	0,001865	0,0018071	0,0017502	0,0016948	0,0016411	0,0015889	0,0015382	0,0014890	0,0014412	0,0013949
-2.8	0,002555	0,0024771	0,0024012	0,0023274	0,0022557	0,0021860	0,0021182	0,0020524	0,0019884	0,0019262
-2.7	0,003467	0,0033642	0,0032641	0,0031667	0,0030720	0,0029798	0,0028901	0,0028028	0,0027179	0,0026354
-2.6	0,004661	0,0045271	0,0043965	0,0042692	0,0041453	0,0040246	0,0039070	0,0037926	0,0036811	0,0035726
-2.5	0,006209	0,0060366	0,0058677	0,0057031	0,0055426	0,0053861	0,0052336	0,0050849	0,0049400	0,0047988
-2.4	0,008197	0,0079763	0,0077603	0,0075494	0,0073436	0,0071428	0,0069469	0,0067557	0,0065691	0,0063872
-2.3	0,010724	0,0104441	0,0101704	0,0099031	0,0096419	0,0093867	0,0091375	0,0088940	0,0086563	0,0084242
-2.2	0,013903	0,0135526	0,0132094	0,0128737	0,0125455	0,0122245	0,0119106	0,0116038	0,0113038	0,0110107
-2.1	0,017864	0,0174292	0,0170030	0,0165858	0,0161774	0,0157776	0,0153863	0,0150034	0,0146287	0,0142621
-2.0	0,022750	0,0222156	0,0216917	0,0211783	0,0206752	0,0201822	0,0196993	0,0192262	0,0187628	0,0183089
-1.9	0,028716	0,0280666	0,0274289	0,0268034	0,0261898	0,0255881	0,0249979	0,0244192	0,0238518	0,0232955
-1.8	0,035930	0,0351479	0,0343795	0,0336250	0,0328841	0,0321568	0,0314428	0,0307419	0,0300540	0,0293790
-1.7	0,044565	0,0436329	0,0427162	0,0418151	0,0409295	0,0400592	0,0392039	0,0383636	0,0375380	0,0367270

-1.6	0,054799	0,0536989	0,0526161	0,0515507	0,0505026	0,0494715	0,0484572	0,0474597	0,0464787	0,0455140
-1.5	0,066807	0,0655217	0,0642555	0,0630084	0,0617802	0,0605708	0,0593799	0,0582076	0,0570534	0,0559174
-1.4	0,080756	0,0792698	0,0778038	0,0763585	0,0749337	0,0735293	0,0721450	0,0707809	0,0694366	0,0681121
-1.3	0,096800	0,0950979	0,0934175	0,0917591	0,0901227	0,0885080	0,0869150	0,0853435	0,0837933	0,0822644
-1.2	0,115069	0,1131394	0,1112324	0,1093486	0,1074877	0,1056498	0,1038347	0,1020423	0,1002726	0,0985253
-1.1	0,135666	0,1334995	0,1313569	0,1292381	0,1271432	0,1250719	0,1230244	0,1210005	0,1190001	0,1170232
-1.0	0,158655	0,1562476	0,1538642	0,1515050	0,1491700	0,1468591	0,1445723	0,1423097	0,1400711	0,1378566
-0.9	0,184060	0,1814113	0,1787864	0,1761855	0,1736088	0,1710561	0,1685276	0,1660232	0,1635431	0,1610871
-0.8	0,211855	0,2089701	0,2061081	0,2032694	0,2004542	0,1976625	0,1948945	0,1921502	0,1894297	0,1867329
-0.7	0,241963	0,2388521	0,2357625	0,2326951	0,2296500	0,2266274	0,2236273	0,2206499	0,2176954	0,2147639
-0.6	0,274253	0,2709309	0,2676289	0,2643473	0,2610863	0,2578461	0,2546269	0,2514289	0,2482522	0,2450971
-0.5	0,308537	0,3050257	0,3015318	0,2980560	0,2945985	0,2911597	0,2877397	0,2843388	0,2809573	0,2775953
-0.4	0,344578	0,3409030	0,3372427	0,3335978	0,3299686	0,3263552	0,3227581	0,3191775	0,3156137	0,3120669
-0.3	0,382088	0,3782805	0,3744842	0,3707000	0,3669283	0,3631693	0,3594236	0,3556912	0,3519727	0,3482683
-0.2	0,420740	0,4168338	0,4129356	0,4090459	0,4051651	0,4012937	0,3974319	0,3935801	0,3897388	0,3859081
-0.1	0,460172	0,4562047	0,4522416	0,4482832	0,4443300	0,4403823	0,4364405	0,4325051	0,4285763	0,4246546
-0.0	0,500000	0,4960106	0,4920217	0,4880335	0,4840466	0,4800612	0,4760778	0,4720968	0,4681186	0,4641436
0.0	0,500000	0,5039894	0,5079783	0,5119665	0,5159534	0,5199388	0,5239222	0,5279032	0,5318814	0,5358564
0.1	0,539827	0,5437953	0,5477584	0,5517168	0,5556700	0,5596177	0,5635595	0,5674949	0,5714237	0,5753454
0.2	0,579259	0,5831662	0,5870644	0,5909541	0,5948349	0,5987063	0,6025681	0,6064199	0,6102612	0,6140919
0.3	0,617911	0,6217195	0,6255158	0,6293000	0,6330717	0,6368307	0,6405764	0,6443088	0,6480273	0,6517317
0.4	0,655421	0,6590970	0,6627573	0,6664022	0,6700314	0,6736448	0,6772419	0,6808225	0,6843863	0,6879331
0.5	0,691462	0,6949743	0,6984682	0,7019440	0,7054015	0,7088403	0,7122603	0,7156612	0,7190427	0,7224047
0.6	0,725746	0,7290691	0,7323711	0,7356527	0,7389137	0,7421539	0,7453731	0,7485711	0,7517478	0,7549029
0.7	0,758036	0,7611479	0,7642375	0,7673049	0,7703500	0,7733726	0,7763727	0,7793501	0,7823046	0,7852361
0.8	0,788144	0,7910299	0,7938919	0,7967306	0,7995458	0,8023375	0,8051055	0,8078498	0,8105703	0,8132671
0.9	0,815939	0,8185887	0,8212136	0,8238145	0,8263912	0,8289439	0,8314724	0,8339768	0,8364569	0,8389129
1.0	0,841344	0,8437524	0,8461358	0,8484950	0,8508300	0,8531409	0,8554277	0,8576903	0,8599289	0,8621434
1.1	0,864333	0,8665005	0,8686431	0,8707619	0,8728568	0,8749281	0,8769756	0,8789995	0,8809999	0,8829768
1.2	0,884930	0,8868606	0,8887676	0,8906514	0,8925123	0,8943502	0,8961653	0,8979577	0,8997274	0,9014747
1.3	0,903199	0,9049021	0,9065825	0,9082409	0,9098773	0,9114920	0,9130850	0,9146565	0,9162067	0,9177356
1.4	0,919243	0,9207302	0,9221962	0,9236415	0,9250663	0,9264707	0,9278550	0,9292191	0,9305634	0,9318879
1.5	0,933192	0,9344783	0,9357445	0,9369916	0,9382198	0,9394292	0,9406201	0,9417924	0,9429466	0,9440826
1.6	0,945200	0,9463011	0,9473839	0,9484493	0,9494974	0,9505285	0,9515428	0,9525403	0,9535213	0,9544860
1.7	0,955434	0,9563671	0,9572838	0,9581849	0,9590705	0,9599408	0,9607961	0,9616364	0,9624620	0,9632730
1.8	0,964069	0,9648521	0,9656205	0,9663750	0,9671159	0,9678432	0,9685572	0,9692581	0,9699460	0,9706210
1.9	0,971283	0,9719334	0,9725711	0,9731966	0,9738102	0,9744119	0,9750021	0,9755808	0,9761482	0,9767045
2.0	0,977249	0,9777844	0,9783083	0,9788217	0,9793248	0,9798178	0,9803007	0,9807738	0,9812372	0,9816911
2.1	0,982135	0,9825708	0,9829970	0,9834142	0,9838226	0,9842224	0,9846137	0,9849966	0,9853713	0,9857379
2.2	0,986096	0,9864474	0,9867906	0,9871263	0,9874545	0,9877755	0,9880894	0,9883962	0,9886962	0,9889893
2.3	0,989275	0,9895559	0,9898296	0,9900969	0,9903581	0,9906133	0,9908625	0,9911060	0,9913437	0,9915758
2.4	0,991802	0,9920237	0,9922397	0,9924506	0,9926564	0,9928572	0,9930531	0,9932443	0,9934309	0,9936128
2.5	0,993790	0,9939634	0,9941323	0,9942969	0,9944574	0,9946139	0,9947664	0,9949151	0,9950600	0,9952012
2.6	0,995338	0,9954729	0,9956035	0,9957308	0,9958547	0,9959754	0,9960930	0,9962074	0,9963189	0,9964274
2.7	0,996533	0,9966358	0,9967359	0,9968333	0,9969280	0,9970202	0,9971099	0,9971972	0,9972821	0,9973646

2.8	0,997444	0,9975229	0,9975988	0,9976726	0,9977443	0,9978140	0,9978818	0,9979476	0,9980116	0,9980738
2.9	0,998134	0,9981929	0,9982498	0,9983052	0,9983589	0,9984111	0,9984618	0,9985110	0,9985588	0,9986051
3.0	0,998650	0,9986938	0,9987361	0,9987772	0,9988171	0,9988558	0,9988933	0,9989297	0,9989650	0,9989992
3.1	0,999032	0,9990646	0,9990957	0,9991260	0,9991553	0,9993008	0,9992112	0,9992378	0,9992636	0,9992886
3.2	0,999312	0,9993363	0,9993590	0,9993810	0,9994024	0,9994230	0,9994429	0,9994623	0,9994810	0,9994991
3.3	0,999516	0,9995335	0,9995499	0,9995658	0,9995811	0,9995959	0,9996103	0,9996242	0,9996376	0,9996505
3.4	0,999663	0,9996752	0,9996869	0,9996982	0,9997091	0,9997197	0,9997299	0,9997398	0,9997493	0,9997585
3.5	0,999767	0,9997759	0,9997842	0,9997922	0,9997999	0,9998074	0,9998146	0,9998215	0,9998282	0,9998347
3.6	0,999840	0,9998469	0,9998527	0,9998583	0,9998637	0,9998689	0,9998739	0,9998787	0,9998834	0,9998879
3.7	0,999892	0,9998964	0,9999004	0,9999043	0,9999080	0,9999116	0,9999150	0,9999184	0,9999216	0,9999247
3.8	0,999927	0,9999305	0,9999333	0,9999359	0,9999385	0,9999409	0,9999433	0,9999456	0,9999478	0,9999499
3.9	0,999951	0,9999539	0,9999557	0,9999575	0,9999593	0,9999609	0,9999625	0,9999641	0,9999655	0,9999670
4.0	0,999968	0,9999696	0,9999709	0,9999721	0,9999733	0,9999744	0,9999755	0,9999765	0,9999775	0,9999784
4.1	0,999979	0,9999802	0,9999811	0,9999819	0,9999826	0,9999834	0,9999841	0,9999848	0,9999854	0,9999861
4.2	0,999986	0,9999872	0,9999878	0,9999883	0,9999888	0,9999893	0,9999898	0,9999902	0,9999907	0,9999911
4.3	0,999991	0,9999918	0,9999922	0,9999925	0,9999929	0,9999932	0,9999935	0,9999938	0,9999941	0,9999943
4.4	0,999995	0,9999948	0,9999951	0,9999953	0,9999955	0,9999957	0,9999959	0,9999961	0,9999963	0,9999964
4.5	0,999996	0,9999968	0,9999969	0,9999971	0,9999972	0,9999973	0,9999974	0,9999976	0,9999977	0,9999978

⊙ **Fuente:** elaboración por parte del Autor

Anexo 2. Distribucion ji-cuadrada acumulada

	0,005	0,01	0,025	0,05	0,1	0,25	0,5	0,75	0,9	0,95	0,975	0,99	0,995	0,999
1	0,000	0,000	0,001	0,004	0,016	0,102	0,455	1,323	2,706	3,841	5,024	6,635	7,879	10,828
2	0,010	0,020	0,051	0,103	0,211	0,575	1,386	2,773	4,605	5,991	7,378	9,210	10,597	13,816
3	0,072	0,115	0,216	0,352	0,584	1,213	2,366	4,108	6,251	7,815	9,348	11,345	12,838	16,266
4	0,207	0,297	0,484	0,711	1,064	1,923	3,357	5,385	7,779	9,488	11,143	13,277	14,860	18,467
5	0,412	0,554	0,831	1,145	1,610	2,675	4,351	6,626	9,236	11,070	12,833	15,086	16,750	20,515
6	0,676	0,872	1,237	1,635	2,204	3,455	5,348	7,841	10,645	12,592	14,449	16,812	18,548	22,458
7	0,989	1,239	1,690	2,167	2,833	4,255	6,346	9,037	12,017	14,067	16,013	18,475	20,278	24,322
8	1,344	1,646	2,180	2,733	3,490	5,071	7,344	10,219	13,362	15,507	17,535	20,090	21,955	26,124
9	1,735	2,088	2,700	3,325	4,168	5,899	8,343	11,389	14,684	16,919	19,023	21,666	23,589	27,877
10	2,156	2,558	3,247	3,940	4,865	6,737	9,342	12,549	15,987	18,307	20,483	23,209	25,188	29,588
11	2,603	3,053	3,816	4,575	5,578	7,584	10,341	13,701	17,275	19,675	21,920	24,725	26,757	31,264
12	3,074	3,571	4,404	5,226	6,304	8,438	11,340	14,845	18,549	21,026	23,337	26,217	28,300	32,909
13	3,565	4,107	5,009	5,892	7,042	9,299	12,340	15,984	19,812	22,362	24,736	27,688	29,819	34,528
14	4,075	4,660	5,629	6,571	7,790	10,165	13,339	17,117	21,064	23,685	26,119	29,141	31,319	36,123
15	4,601	5,229	6,262	7,261	8,547	11,037	14,339	18,245	22,307	24,996	27,488	30,578	32,801	37,697
16	5,142	5,812	6,908	7,962	9,312	11,912	15,338	19,369	23,542	26,296	28,845	32,000	34,267	39,252
17	5,697	6,408	7,564	8,672	10,085	12,792	16,338	20,489	24,769	27,587	30,191	33,409	35,718	40,790
18	6,265	7,015	8,231	9,390	10,865	13,675	17,338	21,605	25,989	28,869	31,526	34,805	37,156	42,312
19	6,844	7,633	8,907	10,117	11,651	14,562	18,338	22,718	27,204	30,144	32,852	36,191	38,582	43,820
20	7,434	8,260	9,591	10,851	12,443	15,452	19,337	23,828	28,412	31,410	34,170	37,566	39,997	45,315

21	8,034	8,897	10,283	11,591	13,240	16,344	20,337	24,935	29,615	32,671	35,479	38,932	41,401	46,797
22	8,643	9,542	10,982	12,338	14,041	17,240	21,337	26,039	30,813	33,924	36,781	40,289	42,796	48,268
23	9,260	10,196	11,689	13,091	14,848	18,137	22,337	27,141	32,007	35,172	38,076	41,638	44,181	49,728
24	9,886	10,856	12,401	13,848	15,659	19,037	23,337	28,241	33,196	36,415	39,364	42,980	45,559	51,179
25	10,520	11,524	13,120	14,611	16,473	19,939	24,337	29,339	34,382	37,652	40,646	44,314	46,928	52,620
26	11,160	12,198	13,844	15,379	17,292	20,843	25,336	30,435	35,563	38,885	41,923	45,642	48,290	54,052
27	11,808	12,879	14,573	16,151	18,114	21,749	26,336	31,528	36,741	40,113	43,195	46,963	49,645	55,476
28	12,461	13,565	15,308	16,928	18,939	22,657	27,336	32,620	37,916	41,337	44,461	48,278	50,993	56,892
29	13,121	14,256	16,047	17,708	19,768	23,567	28,336	33,711	39,087	42,557	45,722	49,588	52,336	58,301
30	13,787	14,953	16,791	18,493	20,599	24,478	29,336	34,800	40,256	43,773	46,979	50,892	53,672	59,703
31	14,458	15,655	17,539	19,281	21,434	25,390	30,336	35,887	41,422	44,985	48,232	52,191	55,003	61,098
32	15,134	16,362	18,291	20,072	22,271	26,304	31,336	36,973	42,585	46,194	49,480	53,486	56,328	62,487
33	15,815	17,074	19,047	20,867	23,110	27,219	32,336	38,058	43,745	47,400	50,725	54,776	57,648	63,870
34	16,501	17,789	19,806	21,664	23,952	28,136	33,336	39,141	44,903	48,602	51,966	56,061	58,964	65,247
35	17,192	18,509	20,569	22,465	24,797	29,054	34,336	40,223	46,059	49,802	53,203	57,342	60,275	66,619
36	17,887	19,233	21,336	23,269	25,643	29,973	35,336	41,304	47,212	50,998	54,437	58,619	61,581	67,985
37	18,586	19,960	22,106	24,075	26,492	30,893	36,336	42,383	48,363	52,192	55,668	59,893	62,883	69,346
38	19,289	20,691	22,878	24,884	27,343	31,815	37,335	43,462	49,513	53,384	56,896	61,162	64,181	70,703
39	19,996	21,426	23,654	25,695	28,196	32,737	38,335	44,539	50,660	54,572	58,120	62,428	65,476	72,055
40	20,707	22,164	24,433	26,509	29,051	33,660	39,335	45,616	51,805	55,758	59,342	63,691	66,766	73,402
41	21,421	22,906	25,215	27,326	29,907	34,585	40,335	46,692	52,949	56,942	60,561	64,950	68,053	74,745
42	22,138	23,650	25,999	28,144	30,765	35,510	41,335	47,766	54,090	58,124	61,777	66,206	69,336	76,084
43	22,859	24,398	26,785	28,965	31,625	36,436	42,335	48,840	55,230	59,304	62,990	67,459	70,616	77,419
44	23,584	25,148	27,575	29,787	32,487	37,363	43,335	49,913	56,369	60,481	64,201	68,710	71,893	78,750
45	24,311	25,901	28,366	30,612	33,350	38,291	44,335	50,985	57,505	61,656	65,410	69,957	73,166	80,077
46	25,041	26,657	29,160	31,439	34,215	39,220	45,335	52,056	58,641	62,830	66,617	71,201	74,437	81,400
47	25,775	27,416	29,956	32,268	35,081	40,149	46,335	53,127	59,774	64,001	67,821	72,443	75,704	82,720
48	26,511	28,177	30,755	33,098	35,949	41,079	47,335	54,196	60,907	65,171	69,023	73,683	76,969	84,037
49	27,249	28,941	31,555	33,930	36,818	42,010	48,335	55,265	62,038	66,339	70,222	74,919	78,231	85,351
50	27,991	29,707	32,357	34,764	37,689	42,942	49,335	56,334	63,167	67,505	71,420	76,154	79,490	86,661
60	35,534	37,485	40,482	43,188	46,459	52,294	59,335	66,981	74,397	79,082	83,298	88,379	91,952	99,607
70	43,275	45,442	48,758	51,739	55,329	61,698	69,334	77,577	85,527	90,531	95,023	100,425	104,215	112,317
80	51,172	53,540	57,153	60,391	64,278	71,145	79,334	88,130	96,578	101,879	106,629	112,329	116,321	124,83
100	67,328	70,065	74,222	77,929	82,358	90,133	99,334	109,141	118,498	124,342	129,561	135,807	140,169	149,44
200	152,241	156,432	162,72	168,279	174,835	186,172	199,334	213,102	226,021	233,994	241,058	249,445	255,26	267,54
300	240,66	245,97	253,91	260,87	269,06	283,135	299,334	316,138	331,789	341,395	349,874	359,906	366,84	381,42
400	330,903	337,15	346,48	354,64	364,20	380,57	399,33	418,697	436,649	447,632	457,305	468,724	476,60	493,13
500	422,303	429,38	439,93	449,147	459,926	478,323	499,333	520,950	540,930	553,127	563,852	576,493	585,20	603,44

⊙ **Fuente:** elaboración por parte del Autor

Anexo 3. Tabla de la distribución t de students acumulada

ν	0,005	0,01	0,025	0,05	0,1	0,25	0,5	0,75	0,9	0,95	0,975	0,99	0,995	0,999
1	-63,656	-31,820	-12,706	-6,3138	-3,0777	-1,0000	0,0000	1,0000	3,0777	6,3138	12,7062	31,8205	63,656	318,308
2	-9,9248	-6,9646	-4,3027	-2,9200	-1,8856	-0,8165	0,0000	0,8165	1,8856	2,9200	4,3027	6,9646	9,9248	22,3271
3	-5,8409	-4,5407	-3,1824	-2,3534	-1,6377	-0,7649	0,0000	0,7649	1,6377	2,3534	3,1824	4,5407	5,8409	10,2145
4	-4,6041	-3,7469	-2,7764	-2,1318	-1,5332	-0,7407	0,0000	0,7407	1,5332	2,1318	2,7764	3,7469	4,6041	7,1732
5	-4,0321	-3,3649	-2,5706	-2,0150	-1,4759	-0,7267	0,0000	0,7267	1,4759	2,0150	2,5706	3,3649	4,0321	5,8934
6	-3,7074	-3,1427	-2,4469	-1,9432	-1,4398	-0,7176	0,0000	0,7176	1,4398	1,9432	2,4469	3,1427	3,7074	5,2076
7	-3,4995	-2,9980	-2,3646	-1,8946	-1,4149	-0,7111	0,0000	0,7111	1,4149	1,8946	2,3646	2,9980	3,4995	4,7853
8	-3,3554	-2,8965	-2,3060	-1,8595	-1,3968	-0,7064	0,0000	0,7064	1,3968	1,8595	2,3060	2,8965	3,3554	4,5008
9	-3,2498	-2,8214	-2,2622	-1,8331	-1,3830	-0,7027	0,0000	0,7027	1,3830	1,8331	2,2622	2,8214	3,2498	4,2968
10	-3,1693	-2,7638	-2,2281	-1,8125	-1,3722	-0,6998	0,0000	0,6998	1,3722	1,8125	2,2281	2,7638	3,1693	4,1437
11	-3,1058	-2,7181	-2,2010	-1,7959	-1,3634	-0,6974	0,0000	0,6974	1,3634	1,7959	2,2010	2,7181	3,1058	4,0247
12	-3,0545	-2,6810	-2,1788	-1,7823	-1,3562	-0,6955	0,0000	0,6955	1,3562	1,7823	2,1788	2,6810	3,0545	3,9296
13	-3,0123	-2,6503	-2,1604	-1,7709	-1,3502	-0,6938	0,0000	0,6938	1,3502	1,7709	2,1604	2,6503	3,0123	3,8520
14	-2,9768	-2,6245	-2,1448	-1,7613	-1,3450	-0,6924	0,0000	0,6924	1,3450	1,7613	2,1448	2,6245	2,9768	3,7874
15	-2,9467	-2,6025	-2,1314	-1,7531	-1,3406	-0,6912	0,0000	0,6912	1,3406	1,7531	2,1314	2,6025	2,9467	3,7328
16	-2,9208	-2,5835	-2,1199	-1,7459	-1,3368	-0,6901	0,0000	0,6901	1,3368	1,7459	2,1199	2,5835	2,9208	3,6862
17	-2,8982	-2,5669	-2,1098	-1,7396	-1,3334	-0,6892	0,0000	0,6892	1,3334	1,7396	2,1098	2,5669	2,8982	3,6458
18	-2,8784	-2,5524	-2,1009	-1,7341	-1,3304	-0,6884	0,0000	0,6884	1,3304	1,7341	2,1009	2,5524	2,8784	3,6105
19	-2,8609	-2,5395	-2,0930	-1,7291	-1,3277	-0,6876	0,0000	0,6876	1,3277	1,7291	2,0930	2,5395	2,8609	3,5794
20	-2,8453	-2,5280	-2,0860	-1,7247	-1,3253	-0,6870	0,0000	0,6870	1,3253	1,7247	2,0860	2,5280	2,8453	3,5518
21	-2,8314	-2,5176	-2,0796	-1,7207	-1,3232	-0,6864	0,0000	0,6864	1,3232	1,7207	2,0796	2,5176	2,8314	3,5272
22	-2,8188	-2,5083	-2,0739	-1,7171	-1,3212	-0,6858	0,0000	0,6858	1,3212	1,7171	2,0739	2,5083	2,8188	3,5050
23	-2,8073	-2,4999	-2,0687	-1,7139	-1,3195	-0,6853	0,0000	0,6853	1,3195	1,7139	2,0687	2,4999	2,8073	3,4850
24	-2,7969	-2,4922	-2,0639	-1,7109	-1,3178	-0,6848	0,0000	0,6848	1,3178	1,7109	2,0639	2,4922	2,7969	3,4668
25	-2,7874	-2,4851	-2,0595	-1,7081	-1,3163	-0,6844	0,0000	0,6844	1,3163	1,7081	2,0595	2,4851	2,7874	3,4502
26	-2,7787	-2,4786	-2,0555	-1,7056	-1,3150	-0,6840	0,0000	0,6840	1,3150	1,7056	2,0555	2,4786	2,7787	3,4350
27	-2,7707	-2,4727	-2,0518	-1,7033	-1,3137	-0,6837	0,0000	0,6837	1,3137	1,7033	2,0518	2,4727	2,7707	3,4210
28	-2,7633	-2,4671	-2,0484	-1,7011	-1,3125	-0,6834	0,0000	0,6834	1,3125	1,7011	2,0484	2,4671	2,7633	3,4082
29	-2,7564	-2,4620	-2,0452	-1,6991	-1,3114	-0,6830	0,0000	0,6830	1,3114	1,6991	2,0452	2,4620	2,7564	3,3962
30	-2,7500	-2,4573	-2,0423	-1,6973	-1,3104	-0,6828	0,0000	0,6828	1,3104	1,6973	2,0423	2,4573	2,7500	3,3852
31	-2,7440	-2,4528	-2,0395	-1,6955	-1,3095	-0,6825	0,0000	0,6825	1,3095	1,6955	2,0395	2,4528	2,7440	3,3749
32	-2,7385	-2,4487	-2,0369	-1,6939	-1,3086	-0,6822	0,0000	0,6822	1,3086	1,6939	2,0369	2,4487	2,7385	3,3653
33	-2,7333	-2,4448	-2,0345	-1,6924	-1,3077	-0,6820	0,0000	0,6820	1,3077	1,6924	2,0345	2,4448	2,7333	3,3563
34	-2,7284	-2,4411	-2,0322	-1,6909	-1,3070	-0,6818	0,0000	0,6818	1,3070	1,6909	2,0322	2,4411	2,7284	3,3479
35	-2,7238	-2,4377	-2,0301	-1,6896	-1,3062	-0,6816	0,0000	0,6816	1,3062	1,6896	2,0301	2,4377	2,7238	3,3400
36	-2,7195	-2,4345	-2,0281	-1,6883	-1,3055	-0,6814	0,0000	0,6814	1,3055	1,6883	2,0281	2,4345	2,7195	3,3326
37	-2,7154	-2,4314	-2,0262	-1,6871	-1,3049	-0,6812	0,0000	0,6812	1,3049	1,6871	2,0262	2,4314	2,7154	3,3256
38	-2,7116	-2,4286	-2,0244	-1,6860	-1,3042	-0,6810	0,0000	0,6810	1,3042	1,6860	2,0244	2,4286	2,7116	3,3190
39	-2,7079	-2,4258	-2,0227	-1,6849	-1,3036	-0,6808	0,0000	0,6808	1,3036	1,6849	2,0227	2,4258	2,7079	3,3128
40	-2,7045	-2,4233	-2,0211	-1,6839	-1,3031	-0,6807	0,0000	0,6807	1,3031	1,6839	2,0211	2,4233	2,7045	3,3069
60	-2,6603	-2,3901	-2,0003	-1,6706	-1,2958	-0,6786	0,0000	0,6786	1,2958	1,6706	2,0003	2,3901	2,6603	3,2317
80	-2,6387	-2,3739	-1,9901	-1,6641	-1,2922	-0,6776	0,0000	0,6776	1,2922	1,6641	1,9901	2,3739	2,6387	3,1953
100	-2,6259	-2,3642	-1,9840	-1,6602	-1,2901	-0,6770	0,0000	0,6770	1,2901	1,6602	1,9840	2,3642	2,6259	3,1737
200	-2,6006	-2,3451	-1,9719	-1,6525	-1,2858	-0,6757	0,0000	0,6757	1,2858	1,6525	1,9719	2,3451	2,6006	3,1315

⊙ **Fuente:** elaboración por parte del Autor

Anexo 4. Tabla de la distribución f acumulada

							v_1									
v_2	1	2	3	4	5	6	7	8	9	10	12	14	15	16	18	19	20
1	161,4	199,5	215,7	224,5	230,1	233,9	236,7	238,8	240,5	241,8	243,9	245,3	245,9	246,4	247,3	247,6	248,0
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,42	19,43	19,43	19,44	19,44	19,45
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,71	8,70	8,69	8,67	8,67	8,66
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,87	5,86	5,84	5,82	5,81	5,80
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,64	4,62	4,60	4,58	4,57	4,56
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,96	3,94	3,92	3,90	3,88	3,87
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,53	3,51	3,49	3,47	3,46	3,44
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,24	3,22	3,20	3,17	3,16	3,15
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,03	3,01	2,99	2,96	2,95	2,94
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,86	2,85	2,83	2,80	2,79	2,77
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,74	2,72	2,70	2,67	2,66	2,65
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,64	2,62	2,60	2,57	2,56	2,54
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,55	2,53	2,51	2,48	2,47	2,46
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,48	2,46	2,44	2,41	2,40	2,39
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,42	2,40	2,38	2,35	2,34	2,33
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,37	2,35	2,33	2,30	2,29	2,28
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,33	2,31	2,29	2,26	2,24	2,23
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,29	2,27	2,25	2,22	2,20	2,19
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,26	2,23	2,21	2,18	2,17	2,16
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,23	2,20	2,18	2,15	2,14	2,12
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,20	2,18	2,16	2,12	2,11	2,10
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,17	2,15	2,13	2,10	2,08	2,07
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,15	2,13	2,11	2,08	2,06	2,05
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,13	2,11	2,09	2,05	2,04	2,03
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,11	2,09	2,07	2,04	2,02	2,01
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,09	2,07	2,05	2,02	2,00	1,99
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,08	2,06	2,04	2,00	1,99	1,97
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,06	2,04	2,02	1,99	1,97	1,96
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,05	2,03	2,01	1,97	1,96	1,94
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,04	2,01	1,99	1,96	1,95	1,93
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,95	1,92	1,90	1,87	1,85	1,84
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,95	1,89	1,87	1,85	1,81	1,80	1,78
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,86	1,84	1,82	1,78	1,76	1,75
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,88	1,82	1,79	1,77	1,73	1,72	1,70
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94	1,86	1,80	1,78	1,76	1,72	1,70	1,69
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,85	1,79	1,77	1,75	1,71	1,69	1,68
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,80	1,74	1,72	1,69	1,66	1,64	1,62
400	3,86	3,02	2,63	2,39	2,24	2,12	2,03	1,96	1,90	1,85	1,78	1,72	1,69	1,67	1,63	1,61	1,60
600	3,86	3,01	2,62	2,39	2,23	2,11	2,02	1,95	1,90	1,85	1,77	1,71	1,68	1,66	1,62	1,60	1,59
800	3,85	3,01	2,62	2,38	2,23	2,11	2,02	1,95	1,89	1,84	1,76	1,70	1,68	1,66	1,62	1,60	1,58
1000	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84	1,76	1,70	1,68	1,65	1,61	1,60	1,58

⊙ **Fuente:** elaboración por parte del Auto

Glosario

A continuación, se presenta algunas de las definiciones y anotaciones expuestas con el objetivo de facilitar la comprensión e interpretación de los diferentes aspectos desarrollados en el texto:

Medida de localización o posición: es el valor que resume un conjunto de observaciones.

Medida de variabilidad: es el grado de dispersión que posee la información de interés.

Medidas de forma: permiten verificar si una distribución empírica posee características de simetría, asimetría y apuntamiento.

Muestra: elementos representativos seleccionados aleatoriamente de una población de interés.

Media aritmética o promedio: es una medida de posición o centramiento que está definida como la suma de las observaciones por su frecuencia.

Media armónica: medida de posición o centramiento que se define como la razón entre la cantidad de la muestra y la suma del recíproco de cada uno de los valores de la variable aleatoria por su frecuencia.

Media geométrica: es una medida de posición o localización que se determina realizando la productoria de los valores de las variables y posteriormente calculando la raíz enésima (dependiendo de la cantidad de datos) de este producto.

Varianza: es una medida de variabilidad o dispersión que se obtiene de la diferencia al cuadrado entre el valor de la variable aleatoria y la media aritmética o promedio.

Desviación estándar: medida de variabilidad o dispersión calculada a partir de la raíz de la varianza, se expresa con las mismas unidades de los valores de la variable evaluada.

Rango : medida de variabilidad o dispersión definida como la diferencia entre el máximo valor de la variable aleatoria y el mínimo valor seleccionado de la muestra.