

Tomo I

Fórmulas estadísticas y distribuciones de probabilidad

MARÍA MARJORIE MORA VALVERDE

Editora



Tomo I

Fórmulas estadísticas y distribuciones de probabilidad

MARÍA MARJORIE MORA VALVERDE

Editora

The logo for Editorial UCR, featuring three horizontal white lines of increasing length above the text "EDITORIAL UCR" in a white sans-serif font.

**EDITORIAL
UCR**

Las opciones de resaltado del texto, anotaciones o comentarios dependerán de la aplicación y dispositivo en que se realice la lectura de este libro digital.

CC.SIBDI.UCR - CIP/4072

Nombres: Mora Valverde, María Marjorie, 1958- , editora.

Título: Fórmulas estadísticas y distribuciones de probabilidad / María Marjorie Mora Valverde, editora.

Descripción: Primera edición digital. | San José, Costa Rica : Editorial UCR, 2024.

Identificadores: **ISBN 978-9968-02-133-3** (PDF)

Materias: LEMB: Estadística matemática. | Matemáticas – Fórmulas.

| Probabilidades. | Estadística matemática – Procesamiento de datos.

Clasificación: CDD 519.5 –ed. 23

Edición aprobada por la Comisión Editorial de la Universidad de Costa Rica.

Primera edición digital (PDF): 2024

© Editorial Universidad de Costa Rica,

Ciudad Universitaria Rodrigo Facio. San José, Costa Rica.

Apdo.: 11501-2060 • Tel.: 2511 5310 • Fax: 2511 5257

administracion.siedin@ucr.ac.cr

www.editorial.ucr.ac.cr

Todos los derechos reservados. Prohibida la reproducción de la obra o parte de ella, bajo cualquier forma o medio, así como el almacenamiento en bases de datos, sistemas de recuperación y repositorios, sin la autorización escrita del editor.

Hecho el depósito de ley.

Contenido

Presentación	vii
I. Fórmulas estadísticas.....	1
1. Medidas de posición para datos sin agrupar.....	3
2. Medidas de posición para datos agrupados	4
3. Medidas de variabilidad para datos sin agrupar.....	6
4. Medidas de variabilidad para datos agrupados.....	8
5. Medidas de forma de la distribución.....	10
6. Proporción y variabilidad de variables dicotómicas.....	11
7. Variabilidad de una variable cualitativa	12
8. Índices de precios y de cantidades.....	12
9. Indicadores relativos.....	14
10. Modelos de crecimiento	15
11. Probabilidades	15
12. Distribuciones de probabilidad de variable discreta	17
13. Distribuciones de probabilidad de variable continua.....	18
14. Tamaño de muestra con muestreo simple al azar.....	19
15. Intervalos de confianza.....	20
Estimación de la media aritmética μ de una población	20
Estimación de la proporción P de una población.....	21
Estimación de la diferencia de medias de dos poblaciones ($\mu_1 - \mu_2$).....	22
Estimación de la diferencia de proporciones de dos poblaciones ($P_1 - P_2$).....	23
Estimación de la varianza poblacional.....	23
16. Estadísticos para pruebas de hipótesis.....	23
Media aritmética μ de una población con distribución normal	23
Diferencia de medias de dos poblaciones con distribución normal ($\mu_1 - \mu_2$).....	24
Observaciones pareadas con la distribución t de <i>Student</i>	25
Proporción P de una población con distribución normal.....	25
Diferencia de proporciones de dos poblaciones con distribución normal ($P_1 - P_2$)	26

Prueba de independencia χ^2 para tablas de contingencia.....	26
Prueba exacta de Fisher para tablas de contingencia.....	27
Prueba χ^2 de homogeneidad para las proporciones de dos o más poblaciones.....	27
Prueba χ^2 para bondad de ajuste (muestreo simple al azar y muestras independientes) .	27
Prueba para la varianza de la población	28
17. Probabilidades y riesgo relativo en tablas de contingencia 2x2.....	28
18. Estadísticos de correlación y asociación lineal	29
19. Estadísticos de regresión lineal.....	31
20. Modelo de regresión logística simple (logaritmo de odds).....	33
21. Muestreo estratificado con muestreo simple azar dentro de estratos.....	33
Selección de la muestra con asignación proporcional	33
Selección de la muestra con asignación de varianza mínima (Neyman)	33
Estimaciones para la media aritmética de una muestra estratificada.....	34
Estimaciones para proporciones en una muestra estratificada.....	35
22. Muestreo simple al azar dentro de conglomerados.....	36
23. Análisis de varianza (ANOVA o ANDEVA).....	37
Análisis de varianza para un diseño completamente al azar (un factor)	37
Análisis de varianza para un diseño de bloques al azar (dos factores).....	38
24. Pruebas de la diferencia entre pares de medias.....	39
Referencias bibliográficas capítulo I.....	41
II. Distribuciones de probabilidad.....	43
Tabla 1. Números aleatorios.....	45
Tabla 2. Distribución acumulada de probabilidades binomiales.....	46
Tabla 3. Distribución acumulada de probabilidades de Poisson.....	62
Tabla 4. Distribución acumulada de probabilidades hipergeométricas	67
Tabla 5. Distribución acumulada de probabilidades normal estándar ($F(z)$)	77
Tabla 6. Valores críticos de la distribución t de <i>Student</i>	79
Tabla 7. Valores críticos de la distribución χ^2	81
Tabla 8. Valores críticos de la distribución F	83
Tabla 9. Valores críticos de la distribución q para la prueba de Tukey.....	99
Tabla 10. Valores críticos del coeficiente de correlación de Pearson (r).....	101
Tabla 11. Características generales de algunas distribuciones de probabilidad	102
Referencias bibliográficas capítulo II.....	105
Notas y créditos	107
Acerca de la editora	109

I

Fórmulas estadísticas

1. Medidas de posición para datos sin agrupar

Moda (M_o)	
$M_o = X_f$	X_f : valor más frecuente en un conjunto de datos
Mediana (M_e)	
$M_e = X_{\left(\frac{n+1}{2}\right)}$ término	n : número total de observaciones o datos $\left(\frac{n+1}{2}\right)$: posición de la mediana en un conjunto de datos ordenados de menor a mayor
Percentil (P_m)	
$P_m = X_{\left[\frac{m(n+1)}{100}\right]}$ término $\rightarrow$ posición del percentil en un conjunto ordenado de menor a mayor	
m : percentil de interés $1 \leq m \leq 99$ n : número total de observaciones o datos	
Aproximación del valor del percentil: ¹ $P_m = X_i + (\text{parte decimal}) \cdot (X_{i+1} - X_i)$ X_i : valor numérico en un conjunto que corresponde a la parte entera de la posición del percentil X_{i+1} : valor numérico siguiente al de la parte entera de la posición del percentil	
Media aritmética simple de la población (μ)	
$\mu = \frac{1}{N} \sum_{i=1}^N X_i$	X_i : valor del elemento i $i = 1, 2, 3, \dots, N$ N : número total de observaciones o datos en la población
Media aritmética simple de la muestra ($\bar{X}$)	
$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$	X_i : valor del elemento i $i = 1, 2, 3, \dots, n$ n : número total de observaciones o datos en la muestra
Media aritmética con ponderador absoluto ($\bar{X}_w$)	
$\bar{X}_w = \frac{\sum_{i=1}^k w_i \cdot X_i}{\sum_{i=1}^k w_i}$	X_i : valor del elemento i $i = 1, 2, 3, \dots, n$ w_i : ponderador absoluto del elemento i k : grupos de ponderadores
Media aritmética con ponderador relativo ($\bar{X}_p$)	
$\bar{X}_p = \sum_{i=1}^k p_i \cdot X_i$	X_i : valor del elemento i $i = 1, 2, 3, \dots, n$ p_i : ponderador relativo del elemento i k : grupos de ponderadores

¹ Se aplica cuando el valor del percentil está entre dos valores enteros. La mayoría de los programas estadísticos redondean el término percentil al entero más próximo. Excel presenta las dos opciones: 1. INC redondea, 2. EXC aproxima con la fórmula.

Media geométrica ($\bar{X}_g$)	
$\bar{X}_g = \sqrt[n]{X_1 \cdot X_2 \cdot X_3 \cdot \dots \cdot X_n}$	n: número de pares de períodos (años, meses, ...) $X_i = \frac{\text{Valor del período actual (t)}}{\text{Valor del período anterior (t - 1)}}$

Otras medias		
Media armónica ($\bar{X}_a$)	Media cuadrática ($\bar{X}_c$)	Media móvil simple
$\bar{X}_a = \frac{n}{\sum \frac{1}{x}}$	$\bar{X}_c = \sqrt{\frac{\sum x^2}{n}}$	$\bar{X}_m = \frac{\text{Suma de medias en } X \text{ períodos}}{\text{Total de períodos}}$

2. Medidas de posición para datos agrupados

Media aritmética de la población (μ)	
$\mu = \frac{1}{N} \sum_{i=1}^K X_i \cdot f_i$ $N = \sum_{i=1}^K f_i$ $X_i = \frac{L_i + L_s}{2}$	X_i: punto medio de la clase i $i = 1, 2, 3, \dots, N$ f_i: número de observaciones en la clase i K: número de clases N: número de observaciones o datos en la población L_i: límite inferior de la clase L_s: límite superior de la clase
Media aritmética de la muestra ($\bar{X}$)	
$\bar{X} = \frac{1}{n} \sum_{i=1}^k X_i \cdot f_i$ $n = \sum_{i=1}^k f_i$ $X_i = \frac{L_i + L_s}{2}$	X_i: punto medio de la clase i $i = 1, 2, 3, \dots, n$ f_i: número de observaciones en la clase i k: número de clases n: número de observaciones o datos en la muestra L_i: límite inferior de la clase L_s: límite superior de la clase

Mediana (M_e)	
$M_e = L_{med} + \left(\frac{\left(\frac{n}{2} \right) - F_{amed}}{f_{med}} \right) \cdot I_{med}$	L_{med} : límite inferior de la clase mediana f_{med} : frecuencia absoluta simple de la clase mediana F_{amed} : frecuencia absoluta acumulada de la clase mediana I_{med} : tamaño del intervalo de la clase mediana n : número total de observaciones o datos
Moda (M_o)	
$M_o = L_{mo} + \left(\frac{d_1}{d_1 + d_2} \right) \cdot I_{mo}$ $d_1 = f_{mo} - f_{amo}$ $d_2 = f_{mo} - f_{pmo}$	L_{mo} : límite inferior de la clase modal f_{mo} : frecuencia absoluta de la clase modal f_{amo} : frecuencia absoluta de la clase anterior a la clase modal f_{pmo} : frecuencia absoluta de la clase posterior a la clase modal d_1 : diferencia entre clase modal y la clase anterior a la clase modal d_2 : diferencia entre clase modal y la clase posterior a la clase modal I_{mo} : tamaño del intervalo de la clase modal
Percentil (P_m)	
Ubicación de la clase del percentil de interés $P_m = L_p + \left(\frac{\left(\frac{m}{100} \cdot n \right) - F_{ap}}{f_p} \right) \cdot I_p$	L_p : límite inferior de la clase percentil $\left(\frac{m}{100} \cdot n \right)$: ubicación de la clase del percentil de interés m : percentil de interés $1 \leq m \leq 99$ n : número total de observaciones o datos F_{ap} : frecuencia absoluta acumulada anterior a la clase percentil f_p : frecuencia absoluta de la clase percentil I_p : tamaño del intervalo de la clase percentil

3. Medidas de variabilidad para datos sin agrupar

Variabilidad de la media aritmética	
Varianza de la población (σ^2)	
$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2$ Fórmula alternativa de cálculo $\sigma^2 = \frac{1}{N} \left[\sum_{i=1}^N X_i^2 - \frac{1}{N} \left(\sum_{i=1}^N X_i \right)^2 \right]$	N: número total de observaciones o datos en la población X_i: valor de un elemento i de la población $i = 1, 2, 3, \dots, N$ μ: media aritmética de la población
Desviación estándar de la población (σ)	
$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2}$ Fórmula alternativa de cálculo $\sigma = \sqrt{\frac{1}{N} \left[\sum_{i=1}^N X_i^2 - \frac{1}{N} \left(\sum_{i=1}^N X_i \right)^2 \right]}$	N: número total de observaciones o datos en la población X_i: valor de un elemento i de la población $i = 1, 2, 3, \dots, N$ μ: media aritmética de la población
Varianza de la muestra (s^2)	
$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ Fórmula alternativa de cálculo $s^2 = \frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 - \frac{1}{n} \left(\sum_{i=1}^n X_i \right)^2 \right]$	n: número total de observaciones o datos en la muestra X_i: valor de un elemento i de la muestra $i = 1, 2, 3, \dots, n$ $\bar{X}$: media aritmética de la muestra

Desviación estándar de la muestra (<i>s</i>)			
$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$ Fórmula alternativa de cálculo $s = \sqrt{\frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 - \frac{1}{n} \left(\sum_{i=1}^n X_i \right)^2 \right]}$		<i>n</i>: número total de observaciones o datos en la muestra <i>X_i</i>: valor de un elemento <i>i</i> de la muestra <i>i</i> = 1, 2, 3, ..., <i>n</i> $\bar{X}$: media aritmética de la muestra	
Desviación media absoluta de la población (<i>DMA</i>)			
$DMA = \frac{1}{N} \sum_{i=1}^N X_i - \mu $		<i>N</i>: número total de observaciones o datos de la población <i>X_i</i>: valor de un elemento <i>i</i> de la población <i>i</i> = 1, 2, 3, ..., <i>N</i> μ: media aritmética de la población	
Desviación media absoluta de la muestra (<i>DMA</i>)			
$DMA = \frac{1}{n} \sum_{i=1}^n X_i - \bar{X} $		<i>n</i>: número total de observaciones o datos de la muestra <i>X_i</i>: valor de un elemento de la muestra <i>i</i> = 1, 2, 3, ..., <i>n</i> $\bar{X}$: media aritmética de la muestra	
Coeficiente de variación de la población (<i>CV</i>)			
$CV = \frac{\sigma}{\mu} \cdot 100$		σ: desviación estándar de la población μ: media aritmética de la población	
Coeficiente de variación de la muestra (<i>CV</i>)			
$CV = \frac{s}{\bar{X}} \cdot 100$		<i>s</i>: desviación estándar de la muestra $\bar{X}$: media aritmética de la muestra	
Variabilidad de la mediana			
Intervalo intercuartil: P ₇₅ – P ₂₅		P ₇₅ : percentil 75 P ₂₅ : percentil 25	
Rango: valor máximo – valor mínimo			
Intervalo interdecil			
P ₉₀ – P ₁₀		P ₉₀ : percentil 90 P ₁₀ : percentil 10	

4. Medidas de variabilidad para datos agrupados

Variabilidad de la media aritmética	
Varianza de la población (σ^2)	
$\sigma^2 = \frac{1}{N} \sum_{i=1}^K (X_i - \mu)^2 \cdot f_i$ Fórmula alternativa de cálculo $\sigma^2 = \frac{1}{N} \left[\sum_{i=1}^K X_i^2 \cdot f_i - \frac{1}{N} \left(\sum_{i=1}^K X_i \cdot f_i \right)^2 \right]$	N: número total de observaciones o datos en la población K: número de clases o grupos en la población X_i: punto medio de la clase i $i = 1, 2, 3, \dots, K$ μ: media aritmética de la población f_i: número de observaciones de la clase i
Desviación estándar de la población (σ)	
$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^K (X_i - \mu)^2 \cdot f_i}$ Fórmula alternativa de cálculo $\sigma = \sqrt{\frac{1}{N} \left[\sum_{i=1}^K X_i^2 \cdot f_i - \frac{1}{N} \left(\sum_{i=1}^K X_i \cdot f_i \right)^2 \right]}$	N: número total de observaciones o datos en la población K: número de clases o grupos en la población X_i: punto medio de la clase i $i = 1, 2, 3, \dots, K$ μ: media aritmética de la población f_i: número de observaciones de la clase i
Varianza de la muestra (s^2)	
$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^k (X_i - \bar{X})^2 \cdot f_i \right)$ Fórmula alternativa de cálculo $s^2 = \frac{1}{n-1} \left[\sum_{i=1}^k X_i^2 \cdot f_i - \frac{1}{n} \left(\sum_{i=1}^k X_i \cdot f_i \right)^2 \right]$	n: número total de observaciones o datos en la muestra k: número de clases en la muestra X_i: punto medio de la clase i $i = 1, 2, 3, \dots, k$ $\bar{X}$: media aritmética de la muestra f_i: número de observaciones de la clase i

Desviación estándar de la muestra (s)	
$s = \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^k (X_i - \bar{X})^2 \cdot f_i \right)}$ Fórmula alternativa de cálculo $s = \sqrt{\frac{1}{n-1} \left[\sum_{i=1}^k X_i^2 \cdot f_i - \frac{1}{n} \left(\sum_{i=1}^k X_i \cdot f_i \right)^2 \right]}$	n: número total de observaciones o datos en la muestra k: número de clases o grupos en la muestra X_i: punto medio de la clase i $i = 1, 2, 3, \dots, k$ $\bar{X}$: media aritmética de la muestra f_i: número de observaciones de la clase i
Varianza entre grupos (σ_e^2)	
$\sigma_e^2 = \frac{1}{N} \sum_{i=1}^K N_i \cdot (\mu_i - \mu)^2$	N: número total de observaciones o datos en la población N_i: número total de observaciones o datos en la población del grupo i $i = 1, 2, 3, \dots, K$ μ: media aritmética de la población μ_i: media aritmética de la población en el grupo i K: número de grupos en la población
Varianza dentro de grupos (σ_d^2)	
$\sigma_d^2 = \frac{1}{N} \sum_{i=1}^K N_i \cdot \sigma_i^2$	N: número total de observaciones o datos en la población N_i: número total de observaciones o datos en la población del grupo i σ_i^2: varianza de la población en el grupo i K: número de grupos en la población
Desviación media absoluta ² (DMA)	
$DMA = \frac{1}{n} \left(\sum_{i=1}^k X_i - \bar{X} \cdot f_i \right)$	n: total de observaciones o datos en la muestra X_i: punto medio de la clase i $i = 1, 2, 3, \dots, k$ $\bar{X}$: media aritmética de la muestra f_i: número de observaciones de la clase i K: número de grupos en la población

2 El valor absoluto de un número expresa su valor positivo, por ejemplo: $|-5| = 5$; $|-4| = 4$

5. Medidas de forma de la distribución

Coeficientes de asimetría	
Coeficiente de asimetría de Pearson (A_{SP})³ $A_{SP} = \frac{\bar{x} - Moda}{s}$ $A_{SP} = \frac{3(\bar{x} - Mediana)}{s}$ Coeficiente de asimetría de Fisher Datos sin agrupar $A_{SF} = \frac{1}{ns^3} \sum_{i=1}^n (x_i - \bar{x})^3$ Datos agrupados $A_{SF} = \frac{1}{ns^3} \sum_{i=1}^k (x_i - \bar{x})^3 f_i$ Coeficiente de asimetría de Bowley $A_{SB} = \frac{P_{25} + P_{75} - 2P_{50}}{P_{75} - P_{25}}$	Si $As = 0$ la distribución es simétrica Si $As > 0$ la distribución es asimétrica positiva Si $As < 0$ la distribución es asimétrica negativa $\bar{x}$: media aritmética de la muestra s: desviación estándar de la muestra n: número total de observaciones o datos x_i: valor de un elemento de la muestra $i = 1, 2, 3, \dots, n$ x_i: punto medio de la clase i $i = 1, 2, 3, \dots, k$ f_i: número de observaciones de la clase i k: número de clases o grupos en la muestra P_{25}: percentil 25 P_{50}: percentil 50 = mediana P_{75}: percentil 75
Coeficiente de curtosis de Fisher	
Datos sin agrupar $C_F = \frac{1}{ns^4} \sum_{i=1}^n (x_i - \bar{x})^4 - 3$ Datos agrupados $C_F = \frac{1}{ns^4} \sum_{i=1}^k (x_i - \bar{x})^4 f_i - 3$	Mesocúrtica si $C = 0$ (concentración de los datos similar a la normal) Leptocúrtica si $C > 0$ (mayor concentración de los datos alrededor de la media, su forma es más puntiaguda) Platicúrtica si $C < 0$ (menor concentración de los datos alrededor de la media, su forma es más achatada)

3 Se utiliza en distribuciones uniformes, unimodales y moderadamente asimétricas.

Esta es una
muestra del libro
en la que se despliega
un número limitado de páginas.

Adquiera el libro completo en la
[Librería UCR Virtual.](#)

LIBRERÍA
UCR

VIRTUAL

Acerca de la editora

María Marjorie Mora Valverde

Nació en Costa Rica en 1958. Obtuvo la Licenciatura en Estadística de la Universidad de Costa Rica (UCR), la Maestría en Tecnología Educativa con Énfasis en Medios Instruccionales de la Universidad Estatal a Distancia y el Doctorado Académico en Educación de la UCR. Durante más de 30 años se desempeñó como profesora de Estadística de la UCR, también fue docente e investigadora en la Escuela de Bibliotecología, Documentación e Información (EBDI) de la Universidad Nacional, Costa Rica.

Corrección filológica: *Isaac Marín M.* y *Pamela Bolaños A.* • Revisión de pruebas: *Sofía Conejo A.*
Diseño de portada: *Abraham Ugarte S.* • Diseño de contenido, diagramación, realización del libro digital
y control de calidad de la versión digital: *Raquel Fernández C.*

Editorial UCR es miembro del Sistema Editorial Universitario Centroamericano (SEDUCA),
perteneciente al Consejo Superior Universitario Centroamericano (CSUCA).

Edición digital de la Editorial Universidad de Costa Rica. Fecha de creación: marzo, 2024.

Esta es una nueva versión, actualizada y ampliada, de fórmulas estadísticas y distribuciones de probabilidad estudiadas en los cursos de Estadística; muy útil para apoyar la enseñanza y el aprendizaje de esta disciplina. Se presenta en dos tomos:

Tomo I. Fórmulas estadísticas y distribuciones de probabilidad. Grupo de fórmulas que orientan el cálculo de las medidas cuantitativas para el análisis de conjuntos de datos, junto a un grupo de tablas con las distribuciones de probabilidad más usadas.

Tomo II. Métodos estadísticos no paramétricos. Para un grupo seleccionado de métodos presenta la descripción de la prueba, las fórmulas y las probabilidades correspondientes.