

Villada Cantor, Diego Alexis

Elementos de estadística descriptiva y probabilidad / Diego Alexis Villada Cantor,
Oscar Javier Beltrán Cortés. Bogotá : Universidad Piloto de Colombia, 2020
159 páginas.

Incluye referencias bibliográficas

ISBN: 978-958-5106-25-3

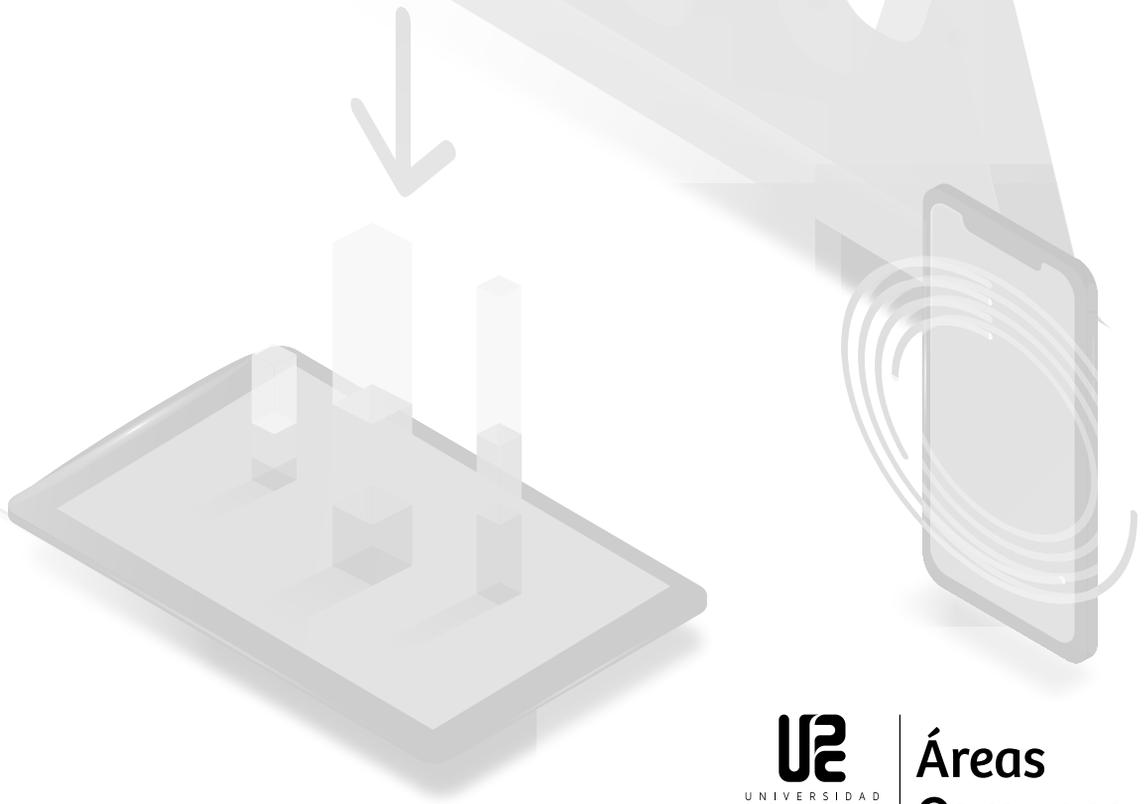
ESTADISTICA DESCRIPTIVA

ESTADISTICA -- PROBLEMAS, EJERCICIOS, ETC

CCD 519.53

Elementos de estadística descriptiva y probabilidad

Diego Alexis Villada Cantor
Oscar Javier Beltrán Cortés



Olinto Eduardo Quiñones Quiñones

Presidente

Ángela Bernal Medina

Rectora

Rodrigo Lobo-Guerrero Sarmiento

Director de Publicaciones y Comunicación Gráfica

Mauricio Hernández Tascón

Director de Investigaciones

Diego Ramírez Bernal

Coordinador General de Publicaciones

Área Común de Matemáticas

© **Elementos de estadística descriptiva y probabilidad**

Autores

Diego Alexis Villada C.

Oscar Javier Beltrán C.

ISBN: 978-958-5106-25-3

Primera edición 2020

Bogotá, Colombia

Coordinadora de la publicación

Catalina Moreno Correa

Diseño y diagramación

Diego Alexis Villada C.

Diseño de portada

Juanita Castillo Aguirre

La obra literaria publicada expresa exclusivamente la opinión de sus respectivos autores, de manera que no representan el pensamiento de la Universidad Piloto de Colombia. Cada uno de los autores suscribió con la Universidad una autorización o contrato de sesión de derechos y una carta de originalidad sobre su aporte, por tanto, los autores asumen la responsabilidad del contenido de esta publicación.

Elementos de estadística descriptiva y probabilidad



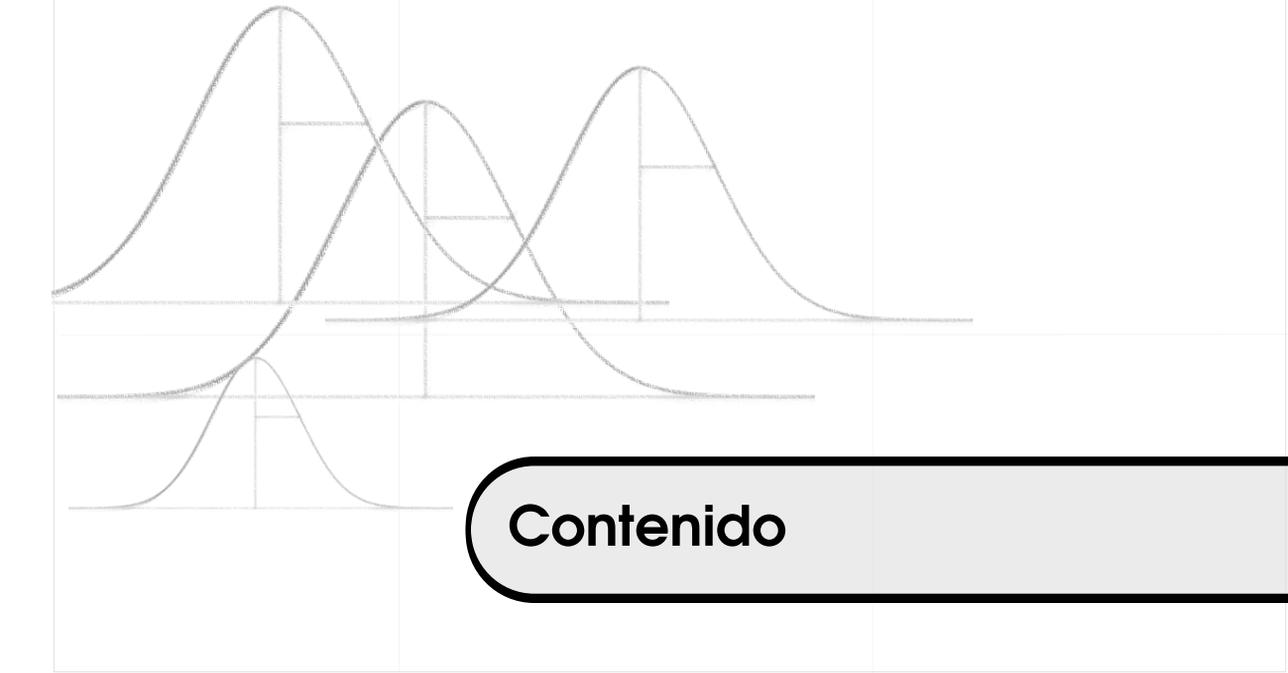
Diego Alexis Villada Cantor
Oscar Javier Beltrán Cortés

*A mis padres, a mis hijas, a mi esposa
esperanzas invariantes de mi vida.*

D. A. V. C.

A mi hijo Diego Alejandro.

O. J. B. C.

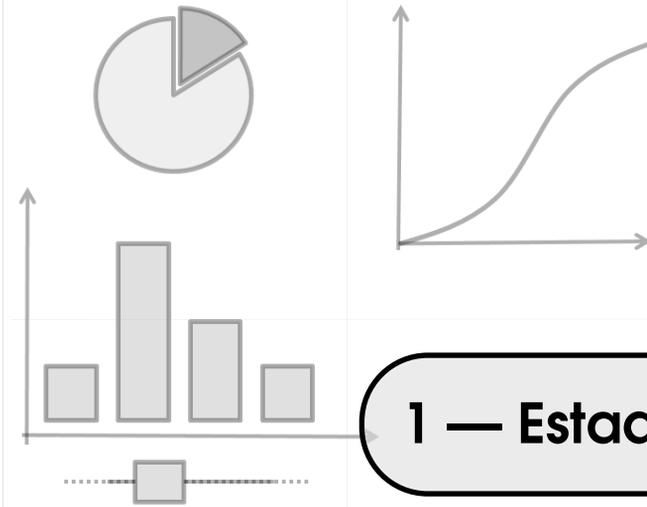


Contenido

1	Estadística descriptiva	12
1.1	Definiciones iniciales	13
1.1.1	Términos clave	14
1.2	Tablas de frecuencias	17
1.2.1	Tablas simples	17
1.2.2	Tablas formuladas	18
1.2.3	Otras frecuencias	21
1.3	Métodos gráficos	25
1.3.1	Histograma	25
1.3.2	Gráfico circular	26
1.3.3	Ojiva	26
1.4	Medidas de resumen	27
1.4.1	Medidas de centralización	27
1.4.2	Medidas de centralización en datos agrupados	30
1.4.3	Medidas de dispersión	34
1.4.4	Cálculo de la dispersión para datos agrupados	39
1.4.5	Medidas de localización	41
1.4.6	Medidas de forma	47

2	Conjuntos y conteo	52
2.1	Elementos básicos de conjuntos	53
2.1.1	Operaciones	53
2.1.2	Otras propiedades	55
2.2	Cardinalidad	57
2.3	Técnicas de conteo	66
2.3.1	Principios de conteo	66
2.3.2	Factoriales	67
2.3.3	Permutación	70
2.3.4	Combinación	71
2.3.5	Tabla de resumen	73
3	Probabilidad	78
3.1	Probabilidad	79
3.2	Probabilidad condicional	86
3.2.1	Definición de probabilidad condicional	86
3.2.2	Tablas de contingencia	89
3.2.3	Diagramas de árbol	92
3.3	Teorema de probabilidad total	94
3.3.1	Regla de Bayes	96
3.4	Independencia de eventos	100
3.4.1	Modelo de urnas	101
4	Variables aleatorias	106
4.1	Variables aleatorias	107
4.2	Variables aleatorias discretas	108
4.2.1	Valor esperado de una variable discreta	112
4.2.2	Propiedades de la esperanza	112
4.2.3	Varianza de variables discretas	114
4.2.4	Propiedades de la varianza	115
4.3	Variables aleatorias continuas	116
4.3.1	Valor esperado de una variable continua	119
4.3.2	Varianza	119
5	Distribuciones discretas	124
5.1	Distribución de Bernoulli	125
5.2	Distribución binomial	125

5.3	Distribución binomial negativa	131
5.3.1	Fórmula para el cálculo de probabilidades	131
5.4	Distribución geométrica	133
5.5	Distribución hipergeométrica	135
5.6	Distribución de Poisson	138
6	Distribuciones continuas	142
6.1	Cálculo de probabilidades para variables aleatorias continuas	143
6.2	Distribución uniforme	143
6.3	Distribución normal	146
6.4	Distribución Gamma	152
6.4.1	Distribución de probabilidad Gamma	152
6.5	Distribución exponencial	154
7	Referencias	158



1 — Estadística descriptiva

Presentación del capítulo

El análisis exploratorio de datos se puede definir como un compendio de métodos que ayudan a la descripción de un conjunto de observaciones registradas antes, durante y después de un estudio o investigación. La mayoría de fórmulas que se encuentran en la estadística descriptiva han sido obtenidas empíricamente, es decir, han sido deducidas, sintetizadas y probadas después de largos años tras la aplicación de dichos métodos por parte de una gran cantidad de investigadores en diferentes ramas del conocimiento, que han necesitado procedimientos con fundamento científico para clasificar, resumir o describir el comportamiento de un fenómeno de particular interés. La información que se obtiene al realizar este tipo de estudios, será usada en etapas posteriores de una investigación para conjeturar e incluso probar una ley propia de las ciencias naturales, económicas o sociales, entre otras.

En este capítulo trataremos los métodos descriptivos de datos más importantes y reconocidos:

Los temas serán los siguientes:

1. Resumen por tablas de frecuencias
2. Resumen por métodos gráficos
3. Medidas de resumen
 - a) Medidas de centralización
 - b) Medidas de localización
 - c) Medidas de dispersión
 - d) Medidas de forma

1.1 Definiciones iniciales

La estadística (*statistics*, término original en inglés) se puede definir, en su origen, como una derivación de la teoría de la probabilidad, la que a su vez se enmarca dentro de la teoría de la medida; todo esto desde el contexto matemático del asunto. Quienes estudian la estadística con más dedicación la definen como una ciencia por sí misma, debido a que se compone de métodos demostrativos y de comprobación propios. Sin embargo, más allá del marco epistemológico a la que se pueda llevar esta dualidad, la estadística es considerada como una herramienta fundamental en la comprobación de resultados en distintas ramas del conocimiento que carecían de un método científico de validación (Martínez, 2000) (Blanco, 2004) (Lastra, 2004) (Walpole, 2015).

Debido a la incontable variedad de saberes que hacen uso de los métodos que se desarrollan, la estadística es usada, por ejemplo, para identificar patrones en campos que pueden llegar a ser muy complejos tales como la administración, medicina, sociología, ingeniería y economía, solo por mencionar algunas de las áreas más permeadas por esta herramienta que tiene su base en la matemática.

En medicina por ejemplo, el comportamiento de una enfermedad, o el diagnóstico de un paciente, se puede establecer mediante un modelo probabilístico que necesitó de méos estadísticos para establecerse. En administración, desde la distribución de los afiliados a una caja de compensación por estratos o por salario, hasta la conformación del grupo familiar, representan ejemplos de aplicación de la estadística. En el sector financiero se puede usar para monitorear el comportamiento de una divisa a través del tiempo. En ingeniería es común usarla para modelar y pronosticar el tiempo de vida de los componentes de un sistema o controlar del funcionamiento de un elemento sometido a esfuerzos o trabajos continuos.

Es así como éste compendio de conocimientos al que llamamos Estadística, repleto de algoritmos y procedimientos numéricos, se puede clasificar de acuerdo a la finalidad, la naturaleza de los métodos usados, la interpretación de la probabilidad y el número de variables examinadas. Dicha clasificación se explica a continuación:

Según la finalidad

Descriptiva

Permite la descripción, resumen y visualización de los grupos en que se subdivide una población o muestra bajo estudio.

Inferencial

Se encarga de hacer conjeturas sobre el comportamiento general de una variable en estudio, de acuerdo con los datos recolectados en el proceso descriptivo del mismo. Es función de la inferencia estadística dar validez a dichas conjeturas por medio de la teoría de probabilidad.

Según los métodos de tratamiento de los datos

Paramétrica

Basa sus apreciaciones e inferencias directamente en los valores de los datos. Esta característica hace que esta metodología se exclusiva de los datos numéricos. Las mediciones obtenidas de este tipo de análisis se contrastan con modelos probabilísticos generados a partir de los patrones naturales de los mismos.

No paramétrica

Las inferencias se extraen a partir del análisis de los rangos de los datos, es decir, a partir de la posición de cada uno de estos una vez son ordenados ascendentemente y no en los valores de estos como en la estadística paramétrica, de modo que al final, los resultados se corroboran de acuerdo a estándares previamente construidos, y aunque también son comparadas con algunas funciones de probabilidad, éstas, aparecen en una etapa posterior del estudio.

Según sus principios de análisis

Clásica

Toma datos y posteriormente aplica métodos descriptivos de inferencia para obtener conclusiones y modelos sobre una situación en particular.

Bayesiana

Investiga la naturaleza de una situación y propone distribuciones de probabilidad que puedan acomodarse ésta. No es primordial tomar datos al respecto, pero se hace para corroborar las formulaciones.

Según el número de variables

Univariada

Cuando el estudio se concentra en el comportamiento de una única característica o variable.

Multivariada

Debido a la presencia de varias variables en una situación, se hace necesario el examen de la asociación que pueda presentarse entre éstas y las relaciones causales que estas asociaciones puedan dar a lugar.

Más allá de la división que se quiera hacer de la estadística debido al contexto en el que esté siendo usada, la finalidad, los métodos y la fundamentación, harán que se generen nuevas ramas de estudio y, de esta manera, que la toma de decisiones basada en evidencia estadística, sea una práctica usual en los distintos escenarios del conocimiento en la que es usada.

1.1.1 Términos clave

Para entrar en el contexto de los procesos estadísticos, debemos definir algunos términos que hacen parte del día a día en el trabajo con datos.

Definición 1.1

Dato: Unidad mínima de información tomada de la realización de un experimento aleatorio y cuyo resultado es registrado como un hecho o acontecimiento. En conjunto con otros, los datos conforman vectores, muestras y poblaciones.

Valor: Asignación numérica o de carácter que contiene el dato.

Población: Conjunto de todos los elementos con una característica en común que puede ser observada.

Muestra: Subconjunto de la población. Se llama *muestra representativa* a aquella que contiene información de la población de manera proporcional.

Variable: Característica observable en cada muestra o población.

Unidad experimental: Elemento básico de los estudios comparativos y encuestas en donde se recogen los datos para el estudio.

Distribución: Resumen del comportamiento de los valores de los datos tomados en una muestra o población.

Sesgo: Influencia ejercida por valores fuera de lo común en una distribución.

No todas las variables estadísticas se pueden analizar de la misma forma. La naturaleza de las variables y la forma en que se ha recogido la información las clasifican en dos tipos:

Cualitativas**Nominal**

Refiere a variables que no se pueden ordenar debido a la ausencia de alguna de forma de medibilidad, nombres, colores, lugares, entre otras.

Ordinal

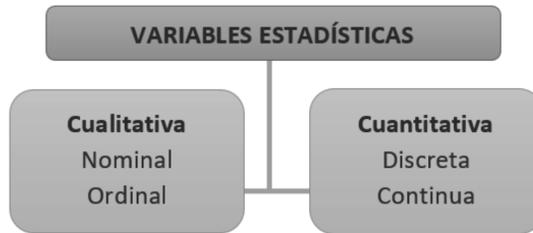
Característica que a pesar de no ser numérica se puede ordenar teniendo en cuenta su asociación con algún tipo de jerarquía. Las apreciaciones sobre la calidad del servicio, la frecuencia en la realización de un evento (nunca, casi nunca, a veces, casi siempre, siempre) o los niveles de complejidad, hacen parte de este tipo de variables.

Cuantitativas**Discreta**

Variable que se caracteriza por ser numérica y por usar, en la mayoría de las ocasiones, una cantidad entera de posibles valores. Se usa para conteos.

Continua

Variable numérica que toma su valor de los números reales. Se usa en casos donde se deben hacer mediciones, índices, porcentajes, entre otras medidas que permiten una alta exactitud.



Ejemplo 1.1

Para caracterizar la condición física de los usuarios de un gimnasio se prepara un formulario de inscripción en el que las preguntas llenarán la base de datos de afiliados y cuyas variables son y se clasifican de la siguiente manera:

1. Sexo: cualitativa nominal
2. Edad (años): cuantitativa discreta
3. Peso (Kg): cuantitativa continua
4. Talla (cm): cuantitativa continua
5. Tiempo en 1609 metros (en minutos): cuantitativa continua
6. Cantidad de abdominales en 30 segundos: cuantitativa discreta
7. Valor en la prueba flexión de tronco al frente (cm): cuantitativa continua
8. BMI (body mass index / índice de masa corporal): cuantitativa continua

Ejercicios 1.1

1.1 Indicar cuáles de las siguientes variables son cualitativas y cuáles son cuantitativas:

- a.) Comida favorita
- b.) Profesión de agrado
- c.) Número de goles marcados por el equipo favorito en la última temporada
- d.) Número de alumnos de un instituto
- e.) El color de los ojos de los compañeros de clase
- f.) Coeficiente intelectual de los compañeros de clase

1.2 Clasificar las siguientes variables en cualitativas nominales u ordinales y en cuantitativas discretas o continuas según corresponda:

- a.) La nacionalidad de las personas
- b.) Litros de agua contenidos en un depósito
- c.) Número de libros en un estante de librería
- d.) Suma de puntos obtenidos en el lanzamiento de un par de dados
- e.) La profesión de las personas
- f.) El área de las distintas baldosas de un edificio