

Estadística descriptiva para datos categóricos

Maribel Serna Rodríguez

Julián Alberto Pareja Vasseur



Serna Rodríguez, Maribel

Estadística descriptiva para datos categóricos / Maribel Serna Rodríguez, Julián Alberto Pareja Vasseur. -- Medellín: Editorial EAFIT, 2017

112 p. ; 27 cm. -- (Académica Z)

ISBN 978-958-720-483-4

1. Estadística. I. Pareja Vasseur, Julián Alberto. II. Tít. III. Serie

519.53 cd 23 ed.

S486

Universidad EAFIT - Centro Cultural Biblioteca Luis Echavarría Villegas

Estadística descriptiva para datos categóricos

Primera edición: diciembre de 2017

© Maribel Serna Rodríguez, Julián Alberto Pareja Vasseur

© Editorial EAFIT

Carrera 49 No. 7 sur - 50

Tel.: 261 95 23, Medellín

<http://www.eafit.edu.co/fondoeditorial>

Correo electrónico: fonedit@eafit.edu.co

ISBN: 978-958-720-483-4

Edición: Marcel René Gutiérrez

Diseño: Alina Giraldo Yepes

Diagramación: Ana María Giraldo G.

Imagen de carátula: 744664039, shutterstock.com

Universidad EAFIT | Vigilada Mineducación. Reconocimiento como Universidad: Decreto Número 759, del 6 de mayo de 1971, de la Presidencia de la República de Colombia. Reconocimiento personería jurídica: Número 75, del 28 de junio de 1960, expedida por la Gobernación de Antioquia. Acreditada institucionalmente por el Ministerio de Educación Nacional hasta el 2026, mediante Resolución 2158 emitida el 13 de febrero de 2018

Prohibida la reproducción total o parcial, por cualquier medio o con cualquier propósito, sin la autorización escrita de la editorial

Editado en Medellín, Colombia

*A mi maravillosa y siempre incondicional familia,
y de manera especial a Isaac*
Maribel Serna Rodríguez

*A la familia, madres y hermanas, que sirvieron de soporte
para la creación de este libro*
Julián Alberto Pareja Vasseur

Contenido

| | |
|---|----|
| Presentación | 19 |
| Prólogo | 21 |
| | |
| 1. Definiciones básicas en Estadística | 23 |
| Introducción..... | 25 |
| 1.1. Definición de la Estadística y sus ramas..... | 26 |
| 1.2. Definición de variable y tipos de variables..... | 26 |
| 1.3. Variables cuantitativas | 27 |
| 1.4. Variables cualitativas..... | 27 |
| 1.5. Definición de datos y su procedencia..... | 28 |
| 1.6. Otros conceptos estadísticos importantes..... | 30 |
| 1.7. Organización del texto | 30 |
| Referencias..... | 31 |
| | |
| 2. Organización de los datos en tablas y gráficas: caso univariado | 33 |
| Introducción..... | 35 |
| 2.1. Tabla de distribución de frecuencias para un dato categórico..... | 36 |
| 2.1.1. Categorías de la variable cualitativa..... | 37 |
| 2.1.2. Frecuencia absoluta | 37 |
| 2.1.3. Frecuencia relativa..... | 42 |
| 2.1.4. Frecuencias absoluta acumulada y relativa acumulada..... | 44 |
| 2.2. Gráficos de columnas, de barras y de pastel para un dato categórico..... | 45 |

| | |
|--|--------|
| 2.2.1. Opción de gráficos | 46 |
| 2.2.2. Tabla dinámica..... | 48 |
| 2.3. Resumen | 51 |
| 2.4. Fórmulas | 51 |
| 2.5. Ejercicios..... | 51 |
| Referencias..... | 53 |
| 3. Organización de los datos en tablas y gráficas: caso bivariado | 55 |
| Introducción..... | 57 |
| 3.1. Tabla de contingencia en términos absolutos y figuras..... | 58 |
| 3.2. Figura de líneas para la tabla de contingencia en términos absolutos..... | 61 |
| 3.3. Tabla de contingencia en términos relativos y figuras | 64 |
| 3.3.1. Construcción de la tabla de perfiles | 64 |
| 3.3.2. Figura de los perfiles | 66 |
| 3.3.3. Construcción de la tabla de perfiles fila | 69 |
| 3.3.4. Figura de los perfiles fila | 71 |
| 3.3.5. Construcción de la tabla de perfiles columna | 72 |
| 3.3.6. Figura de los perfiles columna..... | 74 |
| 3.4. Resumen..... | 75 |
| 3.5. Fórmulas..... | 75 |
| 3.6. Ejercicios | 76 |
| Referencias..... | 79 |
| 4. Organización de los datos en tablas y gráficas: caso multivariado..... | 81 |
| Introducción..... | 83 |
| 4.1. Matriz o tabla de Burt | 85 |

| | |
|--|-----|
| 4.1.1. Conceptos básicos matriciales | 85 |
| 4.1.2. Ejercicio para la construcción de la tabla de Burt | 88 |
| 4.1.3. Metodologías para la construcción de la tabla de Burt | 89 |
| 4.2. Resumen | 105 |
| 4.3. Ejercicios..... | 105 |
| Referencias..... | 109 |
| Los autores | 111 |

Lista de Tablas

| | |
|--|----|
| Tabla 1. Aplicación de conceptos a diversas variables..... | 29 |
| Tabla 2. Otras fuentes de información secundaria | 30 |
| Tabla 3. Codificación de los niveles de la variable <i>color preferido</i> | 37 |
| Tabla 4. Preferencia de color por individuo | 38 |
| Tabla 5. Función condicional para preferencia de color..... | 39 |
| Tabla 6. Preferencia de color agrupada mediante función condicional..... | 39 |
| Tabla 7. Preferencia de color agrupada mediante función frecuencia..... | 41 |
| Tabla 8. Frecuencias relativas para preferencia de color en la tabla dinámica..... | 43 |
| Tabla 9. Distribución de frecuencias para preferencia de colores | 45 |
| Tabla 10. Distribución de la población mundial por regiones geográficas a mediados de 2013..... | 49 |
| Tabla 11. Distribución de frecuencias de la población mundial a mediados de 2013..... | 49 |
| Tabla 12. Frecuencia de ocurrencia de terremotos | 52 |
| Tabla 13. Estructura de una tabla de contingencia | 59 |
| Tabla 14. Tabla de contingencia para la estratificación socioeconómica de viviendas en Medellín año 2011..... | 60 |
| Tabla 15. Estructura de las tablas de contingencia..... | 64 |
| Tabla 16. Estructura de las tablas de contingencia según los perfiles..... | 65 |
| Tabla 17. Perfiles de estratificación socioeconómica de viviendas en Medellín año 2011 | 66 |
| Tabla 18. Estructura de las tablas de contingencia según los perfiles fila..... | 70 |

| | |
|--|-----|
| Tabla 19. Perfiles fila estratificación socioeconómica de viviendas en Medellín año 2011 | 70 |
| Tabla 20. Estructura de las tablas de contingencia según los perfiles columna | 73 |
| Tabla 21. Perfiles columna estratificación socioeconómica de viviendas en Medellín año 2011 | 73 |
| Tabla 22. Información de los asteroides con cercanía de impacto a la Tierra | 76 |
| Tabla 23. Información para la variable <i>diámetro</i> | 77 |
| Tabla 24. Información para la variable <i>Escala de Palermo</i> | 77 |
| Tabla 25. Total de personas de 12 años y más que vieron películas en video, según el género, en el último mes en las cabeceras municipales en Colombia para el año 2014..... | 78 |
| Tabla 26. Descripción de variables y niveles..... | 88 |
| Tabla 27. Descripción de los niveles de cada variable en la muestra | 89 |
| Tabla 28. Codificación de los niveles para cada variable para el primer individuo .. | 90 |
| Tabla 29. Codificación de los niveles de cada variable para los 20 individuos o matriz Z | 91 |
| Tabla 30. Matriz Z' | 92 |
| Tabla 31. Matriz de Burt o matriz B , sin las variables y sus respectivos niveles | 94 |
| Tabla 32. Matriz de Burt o matriz B , con las variables y sus respectivos niveles | 94 |
| Tabla 33. Identificación de las matrices que componen la diagonal y las matrices triangulares superior e inferior de la tabla de Burt..... | 95 |
| Tabla 34. Matriz de Burt correspondiente al ejercicio e identificación de su diagonal y sus matrices triangulares..... | 96 |
| Tabla 35. Forma estructural de la tabla de Burt | 97 |
| Tabla 36. Descripción de variables y niveles..... | 105 |
| Tabla 37. Descripción de los niveles de cada variable para cada observación | 106 |
| Tabla 38. Descripción de variables y niveles..... | 107 |
| Tabla 39. Descripción de los niveles de cada variable para cada observación | 108 |

Lista de Figuras

| | |
|---|----|
| Figura 1. Mapa conceptual de conceptos generales de la Estadística Descriptiva . | 25 |
| Figura 2. Fuentes y procedencia de los datos..... | 28 |
| Figura 3. Relación entre los objetivos y los capítulos del texto | 31 |
| Figura 4. Mapa conceptual de la Estadística Descriptiva para datos categóricos.... | 35 |
| Figura 5. Mapa conceptual de la Estadística Descriptiva para un dato categórico | 36 |
| Figura 6. Rango de la tabla dinámica para preferencia de color..... | 39 |
| Figura 7. Celda de salida de la tabla dinámica para preferencia de color..... | 40 |
| Figura 8. Preferencia de color agrupada en la tabla dinámica | 40 |
| Figura 9. Función FRECUENCIA para preferencia de color..... | 41 |
| Figura 10. Campo de valor para frecuencias relativas para preferencia de color en la tabla dinámica..... | 43 |
| Figura 11. Mostrar valor de frecuencias relativas para preferencia de color en la tabla dinámica..... | 43 |
| Figura 12. Opción de gráficos para elaborar gráfico de pastel..... | 46 |
| Figura 13. Opción de gráficos para elaborar un gráfico de columnas..... | 47 |
| Figura 14. Opción de gráficos para elaborar un gráfico de barras | 47 |
| Figura 15. Gráfico de columnas a partir de una tabla dinámica..... | 48 |
| Figura 16. Gráfico de pastel a partir de una tabla dinámica..... | 48 |
| Figura 17. Gráfico de columnas de la distribución porcentual de la población mundial por continente a mediados de 2013 | 50 |
| Figura 18. Gráfico de pastel de la distribución porcentual de la población mundial por continente a mediados de 2013 | 50 |
| Figura 19. Mapa conceptual de la Estadística Descriptiva para datos categóricos ... | 57 |

| | |
|---|-----|
| Figura 20. Mapa conceptual de la Estadística Descriptiva de datos categóricos para dos variables..... | 58 |
| Figura 21. Instrucciones para elaborar una gráfica de líneas en Excel a partir de una tabla de contingencia..... | 61 |
| Figura 22. Gráfica en líneas de la distribución de viviendas de la comuna 16..... | 62 |
| Figura 23. Distribución de viviendas para el estrato 3..... | 62 |
| Figura 24. Distribución de viviendas de los seis estratos | 63 |
| Figura 25. Perfil para la comuna 7..... | 67 |
| Figura 26. Perfil del estrato 3..... | 67 |
| Figura 27. Masa de las filas..... | 68 |
| Figura 28. Masa de las columnas | 69 |
| Figura 29. Perfil de la comuna 7..... | 71 |
| Figura 30. Perfil de fila medio | 72 |
| Figura 31. Perfil columna del estrato 2 | 74 |
| Figura 32. Perfil columna medio | 75 |
| Figura 33. Mapa conceptual de la Estadística Descriptiva para n datos categóricos | 83 |
| Figura 34. Mapa conceptual de la Estadística Descriptiva de datos categóricos para n variables..... | 84 |
| Figura 35. Multiplicación de Z y Z' | 93 |
| Figura 36. Tabla dinámica de la variable <i>género</i> | 98 |
| Figura 37. Registro de la información que suministra la tabla dinámica en la tabla de Burt | 98 |
| Figura 38. Tabla dinámica de la variable <i>publicaciones</i> | 99 |
| Figura 39. Tabla dinámica de la variable <i>categoría</i> | 99 |
| Figura 40. Tabla dinámica de la variable <i>categoría</i> según orden personalizado | 100 |
| Figura 41. Tabla dinámica para las variables <i>género</i> y <i>publicaciones</i> | 101 |
| Figura 42. Registro de la información que suministra la tabla dinámica en la tabla de Burt..... | 101 |

| | |
|--|-----|
| Figura 43. Tabla dinámica para las variables <i>género</i> y <i>categoría</i> | 102 |
| Figura 44. Tabla dinámica para las variables <i>publicaciones</i> y <i>categoría</i> | 102 |
| Figura 45. Proceso para transponer la información de la figura 42, mediante tablas dinámicas | 103 |
| Figura 46. Proceso para transponer la información de la figura 42, mediante tablas dinámicas..... | 104 |
| Figura 47. Proceso para transponer la información de la figura 42, mediante la función TRANSPONER | 105 |

Lista de Ecuaciones

| | |
|-------------------|----|
| Ecuación 1 | 38 |
| Ecuación 2 | 42 |
| Ecuación 3 | 42 |
| Ecuación 4 | 44 |
| Ecuación 5 | 44 |
| Ecuación 6 | 44 |
| Ecuación 7 | 45 |
| Ecuación 8 | 65 |
| Ecuación 9 | 69 |
| Ecuación 10 | 72 |

Lista de Ejemplos

| | |
|---|----|
| Ejemplo 1. Encuesta a 10 personas para preguntar por su color preferido. | 37 |
| Ejemplo 2. Distribución de la población mundial por regiones geográficas a mediados de 2013 | 49 |
| Ejemplo 3. Estratificación socioeconómica de viviendas en Medellín año 2011 | 60 |
| Ejemplo 4. Muestra de 20 profesores para la construcción de la tabla de Burt | 88 |

Lista de Ejercicios

| | |
|---|-----|
| Ejercicio 1. Análisis de los terremotos en el mundo. | 51 |
| Ejercicio 2. Análisis del consumo mundial de alcohol desde el año 2000 hasta la fecha | 52 |
| Ejercicio 3. Análisis de riesgo del potencial de impacto de un asteroide contra la Tierra..... | 76 |
| Ejercicio 4. Análisis del género de películas vistas en video por personas de 12 años y más en diciembre de 2014 en Colombia en las cabeceras municipales | 78 |
| Ejercicio 5. Análisis de las características de una muestra de estudiantes. | 105 |
| Ejercicio 6. Análisis de las características sociodemográficas de una muestra de estudiantes | 107 |

Presentación

La necesidad que posee todo profesional relacionada con la recolección de datos, su organización, la presentación, el análisis, la interpretación y la toma de decisiones es lo que motivó la escritura de este texto por parte de los autores, con fin de llenar el vacío en manuscritos que abarcan este tipo de temas. Se hace especial énfasis en el manejo tanto de los datos categóricos como de un *software*, el Excel, tema que cada vez se convierte en una herramienta esencial para todo tipo de público.

Prólogo

El objetivo principal que se persigue con este libro es ofrecer al público en general una guía para abordar la Estadística Descriptiva para datos categóricos utilizando como herramienta fundamental el *software* Excel. A lo largo del texto se documenta la recolección, organización, presentación e interpretación de los datos categóricos para la toma de decisiones. Por tanto, el principal objetivo del mismo corresponde a la presentación formal de una serie de herramientas que permiten resumir la información referente a este tipo de variables.

El contenido del texto se ha organizado en cuatro capítulos. En el primer capítulo, el objetivo es definir una serie de conceptos que permiten entender lo correspondiente a la Estadística Descriptiva. Por su parte, los capítulos 2, 3 y 4, abordan la construcción de tablas y gráficas para datos categóricos de tipo univariado (una variable), bivariado (dos variables) y multivariado (más de dos variables) respectivamente.

Medellín, diciembre de 2017