

Ricardo Michel MALLQUI BAÑOS

Elementos de la estadística

estadística descriptiva y probabilidades



Índice general



--- Índice de tablas





Índice de figuras



Resumen

La estadística es la ciencia que manipula datos los analiza e interpreta para poder sacar conclusiones razonables de ciertos fenomenos naturales. Esta ciencia puede ser dividido en dos: estadística descirptiva y estadística inferencial. En la estadística descriptiva se procesan datos de una manera teórica y utilitaria. Estos métodos consisten en la recolección, organización, resumen, descripcion y presenatacion de la información. Si la poblacion está disponible entonces la estadística descriptiva es suficiente para describir ciertos fenomenos. No obstante generalmente no se dispone de toda la población si no de una muestra de ella, es en este caso que se requieren usar técnicas más sofisticadas para tomar decisiones y generalizaciones acerca de la poblacion, desde una pequeña muestra de información. Es cuando entra en el juego la estadística inferencial.

La base teórica de la estadística son las matemáticas

Este libro se compone de dos partes, la primera parte trata sobre la estadística descirptiva y la segunda estadística inferencial. Cada una de ellas divididas en capítulos.



Parte I

Estadística descriptiva





2

Variables

Es una característica de personas cosas u objetos que son propensos a ser medidas o cualificadas

2.1. Variables cualitativas

Denotan cualidades de objetos personas o animales tales como características inherentes que no son medibles por números, tenemos dos casos de esta variable.

2.1.1. Nominales

Son características que simplemente nominan y están propensos a ser jerarquizados u ordenados tales como: El estado civil (soltero, casado, divorciado, viudo), Religión (católica, evangélico, judío, etc).

2.1.2. Ordinales

Son características que si están propensos a ser jerarquizados tales como: Nivel de instrucción (inicial, primaria, secundaria, superior).

2.2. Variables cuantitativas

Son aquellas variables que están propensos a ser medidas mediante números ya sean números enteros o reales.

2.2.1. Discretas

Aquellas que solo son medidos mediante números enteros por ejemplo: Número de hijos, número de habitaciones.

2.2.2. Continuas

Aquellas que solo son medidos mediante numeros reales es decir este incluye a los numeros racionales e irracionales. Estatura, volumen, peso.

2.3. Asignación

1. Reconosca 5 variables cualitativas de una persona, admosfera, una pintura

- Persona

- Color (moreno, blanco, trigueo)
- Religión (catolico, evangelico, pentecostal, etc)
- wwwwww
- www
- wwwwww

- Admosfera

- www
- wwwwww
- wwwwww
- www
- wwwwww

- Pintura

- www
- wwwwww
- wwwwww
- www
- wwwwww

2. Reconosca 5 variables cuantitativas de una video, tela, un celular.

- Video

- Duracion (x segundos)
- Numero video en youtube a la semana (n cantidades)
- wwwwww
- www
- wwwwww

- Tela

- www
- wwwwww
- wwwwww
- www
- wwwwww

- Celular

2.3 Asignación

7

- www
- wwwwww
- wwwwwwwww
- wwwwww
- wwwwwwwww



3

Organización de datos en tablas de frecuencias

3.1. Distribución de frecuencias

La tabulación es un proceso en el cual los datos son ordenados en grupos llamados clases para un análisis más eficaz de estos, los datos podrían estar clasificados mediante una variable cualitativa o cuantitativa en el caso de las variables cualitativas Y_i , se considera la siguiente Tabla ??

Tabla 3.1: Caption

Y_i	f_i	F_i	F_i^*	h_i	H_i	H_i^*	$h_i \%$	$H_i \%$	$H_i^* \%$
Y_1	f_1	F_1	F_1^*	$\frac{f_1}{n}$	$\frac{F_1}{n}$	$\frac{F_1^*}{n}$	h_1	H_1	H_1^*
Y_2	f_2	F_2	F_2^*	$\frac{f_2}{n}$	$\frac{F_2}{n}$	$\frac{F_2^*}{n}$	h_2	H_2	H_1^*
Y_3	f_3	F_3	F_3^*	$\frac{f_3}{n}$	$\frac{F_3}{n}$	$\frac{F_3^*}{n}$	h_3	H_3	H_1^*
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
Y_r	f_r	F_r	F_r^*	$\frac{f_r}{n}$	$\frac{F_r}{n}$	$\frac{F_r^*}{n}$	h_r	H_r	H_1^*

En el caso de variables cuantitativas además si los datos son muy variados, que para se clasificados adecuadamente, necesitan generarse particiones de longitudes semejantes entonces se utiliza el siguiente proceso; el número de las particiones r se consideran de acuerdo a tres criterios

1. Criterio del investigador r no puede ser más de 20 ni menos de 5
2. $r = \sqrt{n}$ donde n es el número de datos
3. La regla de Starges que consiste en considerar la fórmula $r = 3,322 \cdot \log_{10} n$ Una vez establecido el número de particiones se procede a generar los límites laterales de cada una de las particiones, sea L la longitud de todo el conjunto es decir $L = x_{\max} - x_{\min}$ entonces la longitud de las particiones o amplitud intervállica se obtiene con $l = \frac{L}{r}$

Tabla 3.2: Datos cuantitativos (intervalos)

Clase	Y_i	f_i	F_i	F_i^*	h_i	H_i	H_i^*	$h_i\%$	$H_i\%$	$H_i^*\%$
$[y_1 - y_2)$	Y_1	f_1	F_1	F_1^*	$\frac{f_1}{n}$	$\frac{f_1}{n}$	$\frac{F_1^*}{n}$	$h_1\%$	$H_1\%$	$H_1^*\%$
$[y_2 - y_3)$	Y_2	f_2	F_2	F_2^*	$\frac{f_2}{n}$	$\frac{f_2}{n}$	$\frac{F_2^*}{n}$	$h_2\%$	$H_2\%$	$H_2^*\%$
...
$[y_{r-1} - y_r)$	Y_r	f_r	F_r	F_r^*	$\frac{f_r}{n}$	$\frac{f_r}{n}$	$\frac{F_r^*}{n}$	$h_r\%$	$H_r\%$	$H_r^*\%$

Tenga en cuenta que n es el número de datos, es decir $n = f_1 + f_2 + \dots + f_r = \sum_{i=1}^r f_i$ donde f_i es número de datos en la partición X_i , una de las r particiones del conjunto total de datos.

1. Las frecuencias absolutas f_i indican el número de datos con la característica X_i .
2. Las frecuencias absolutas acumuladas menor que F_i obedecen a la fórmula

$$F_m = f_1 + f_2 + \dots + f_m = \sum_{i=1}^m f_i$$

3. Las frecuencias absolutas acumuladas mayor que F_i^* obedecen a la fórmula

$$\begin{aligned}
 F_m^* &= f_m + f_{m+1} + \dots + f_r \\
 &= \sum_{i=m}^r f_i \\
 &= n - \sum_{i=1}^{m-1} f_i \\
 &= n - (f_1 + f_2 + \dots + f_{m-1})
 \end{aligned}$$

4. Las frecuencias absolutas relativas obedecen a la fórmula

$$h_m = \frac{f_m}{n}$$

5. Las frecuencias absolutas relativas menor que obedecen a la fórmula

$$H_m = \frac{f_m}{n}$$

6. Las frecuencias absolutas relativas mayor que obedecen a la fórmula

$$H_m^* = \frac{F_m^*}{n}$$

7. Las frecuencias absolutas relativas porcentuales obedecen a la fórmula $h_i \% = 100 \cdot h_i$
 8. Las frecuencias absolutas relativas menor que porcentuales obedecen a la fórmula $H_i \% = 100 \cdot H_i$
 9. Las frecuencias absolutas relativas mayor que porcentuales obedecen a la fórmula $H_i^* \% = 100 \cdot H_i^*$
-

3.2. Ejemplo sin intervalos

Ejercicio 3.1. Sean Los 16 tipos de personalidad en un grupo social encuestado.

1. ESTJ (Extraverted Sensing Thinking Judging) Personas a las que les gusta tener el control sobre lo que ocurre a su alrededor, siempre buscan la manera de que todo funcione como debe y, si es necesario, la implementan ellos mismos.

2. ESTP ((Extraverted Sensing Thinking Perceiving) Las personas que pertenecen a esta categoría son espontáneas, alegres y activas, pero al igual que lo que ocurre con los ESTJ, tienden a ejercer dominio sobre los demás, en este caso a través de su capacidad de observación y su carisma.

3. ESFJ (Extraverted Sensing Feeling Judging) Se trata de personas muy volcadas en la atención de las necesidades de los demás, especialmente si forman parte de su círculo cercano: familia y amistades. Por eso siempre que pueden prestan su ayuda y procuran que sus círculos sociales cercanos permanezcan siempre estables y con buena salud. Por eso tienden a evitar que aparezcan conflictos fuertes y se muestran diplomáticas cuando hay choques de intereses.

4. ESFP (Extraverted Sensing Feeling Perceiving) Se trata de personas alegres y espontáneas que disfrutan entreteniéndose y entreteniéndolo a los demás. La diversión es uno de los pilares más importantes de sus vidas, y son de trato cercano y temperamento cálido. Aman la novedad y hablar acerca de experiencias personales.

5. ISTJ (Introverted Sensing Thinking Perceiving) Un tipo de personalidad definido por su fuerte sentido de la moralidad y del deber. Les gusta planear e implementar sistemas de reglas que permitan que equipos y organizaciones funcionen con una clara lógica y orden. Dan un gran valor a las normas y a la necesidad de que la realidad se corresponda con cómo deberían ser las cosas. Aunque son personas introvertidas, no rehuyen la interacción con los demás.

6. ISTP (Introverted Sensing Thinking Perceiving) Se trata de personas reservadas, orientadas a la acción y a las soluciones prácticas ante problemas del día a día. También son definidas por su tendencia hacia el pensamiento lógico y su espontaneidad y autonomía. Les gusta explorar entornos y descubrir modos en los que se puede interactuar con ellos.

7. ISFJ (Introverted Sensing Feeling Judging) Son personas definidas principalmente por sus ganas de proteger y ayudar a los demás y, en definitiva, de resultar confiables para los otros. Se esfuerzan por hacer todo lo que se espera de ellas, pero no tienen grandes aspiraciones ni se muestran muy ambiciosas. Tienden a pensar que es malo pedir compensaciones o aumentos a cambio de los sacrificios que realizan a la hora de trabajar, ya que este debería ser una meta en sí.

8. ISFP (Introverted Sensing Feeling Perceiving) Personas que viven totalmente en el aquí y el ahora, en constante búsqueda de la novedad y de las situaciones sensorialmente estimulantes. Son reservadas, pero también alegres, espontáneas y cálidas con sus amistades. Tienen un especial talento en el mundo de las artes.

9. ENTJ (Extraverted Intuitive Thinking Judging) Este es uno de los 16 tipos de personalidad más relacionados con el liderazgo y la asertividad. Las personas descritas por esta categoría son comunicativas, de pensamiento ágil y analítico y predispuestas a encabezar equipos y organizaciones. Se adaptan bien al cambio y hacen que sus estrategias también se amolden cada vez que el entorno varía. Además, casi siempre saben cómo explicar sus proyectos o historias de manera que resulten de interés para el resto, lo cual los convierte en comerciales muy aptos.

10. ENTP (Extraverted Intuitive Thinking Perceiving) Personas especialmente movidas por la curiosidad y por los retos que para ser resueltos requieren afrontar preguntas intelectualmente estimulantes. Su agilidad mental y su facilidad para detectar inconsistencias lógicas hace de ellas personas predispuestas a interesarse por la ciencia o la filosofía. Además, su tendencia a mostrarse competitivas las vuelve personas muy activas durante el día, siempre intentando llegar a soluciones innovadoras a problemas complejos.

11. ENFJ (Extraverted Intuitive Feeling Judging) Personas que aprenden constantemente acerca de todos los ámbitos del conocimiento (o una buena parte de ellas) y ayudan a aprender a las demás, guiándolas en su propia evolución. Les gusta ofrecer tutela y consejo, y son muy buenas influyendo en la conducta de los demás. Se centran en sus valores e ideales y hacen lo posible por mejorar

el bienestar del mayor número de personas a través de sus ideas y sus acciones.

12. ENFP (Extraverted Intuitive Feeling Perceiving) Uno de los 16 tipos de personalidad con mayor propensión al pensamiento creativo, las artes y a la sociabilidad. Son alegres, disfrutan de la interacción con otras personas, y actúan teniendo en mente su posición como parte de un “todo” formado por la humanidad, y no se muestran individualistas. De hecho, suelen involucrarse en tareas colectivas para ayudar a los demás, pensando en el impacto social de sus acciones. Sin embargo, también se distraen fácilmente y es frecuente que posterguen tareas que consideran aburridas o demasiado simples y rutinarias.

13. INTJ (Introverted Intuitive Thinking Judging) Un tipo de personalidad orientado hacia la resolución de problemas específicos a partir del razonamiento analítico. Las descritas por esta categoría son personas muy centradas en sus propias ideas y teorías acerca del funcionamiento del mundo, lo cual significa que analizan su entorno centrándose en sus ideas sobre cómo opera este. Son conocedoras de sus propias capacidades y confían en su propio criterio, aunque este vaya en contra de algunos superiores.

Es muy frecuente que lleguen a ser expertas en un ámbito de conocimiento muy específico, ya que les gusta tener el suficiente conocimiento sobre algo como para poder tener en cuenta todos los factores que entran en juego en su funcionamiento y, a partir de ahí, saber lo que se puede hacer o lo que pasará en el futuro.

14. INTP (Introverted Intuitive Thinking Perceiving) Uno de los 16 tipos de personalidad más definido por la propensión a la reflexión. A estas personas les gustan las teorías con capacidad para explicar todo lo que puede ocurrir en un sistema, y su tendencia hacia el perfeccionismo hace que corrijan a los demás en múltiples ocasiones. Valoran más la exactitud en términos teóricos que el pragmatismo y la resolución de problemas concretos.

15. INFJ (Introverted Intuitive Feeling Judging) Personas muy sensibles, reservadas y movidas por unos ideales muy definidos y que, además, sienten la necesidad de hacer que los demás también se beneficien de estos ideales. Esto hace que sean propensas tanto a la reflexión como a la acción, lo cual puede llegar a suponer tanto trabajo que se sobrecargan por tener demasiadas responsabilidades. Muestran una gran capacidad para interpretar exitosamente los estados mentales de los demás y tratan de utilizar esta información para ayudarlas antes de que la otra persona se lo pida.

16. INFP (Introverted Intuitive Feeling Perceiving) Menos moralistas que los INFJ, los INFP también se preocupan mucho por ayudar a los demás desde su posición de personas reservadas. Muestran una sensibilidad estética y artística que las vuelve creativas.

Solución. Se tiene 16 características con los siguientes datos

Personalidad	Cantidad
ESTJ	1
ESTJ	2
ESTP	3
ESFJ	4
ESFP	5
ISTJ	6
ISTP	7
ISFJ	8
ISFP	9
ENTJ	10
ENTP	6
ENFJ	5
ENFP	3
INTJ	2
INTP	1
INFJ	1
INFP	1

Tabla 3.4: Datos cualitativos

Y_i	f_i	F_i	F_i^*	h_i	H_i	H_i^*	$h_i \%$	$H_i \%$	$H_i^* \%$
ESTJ	1	1	75	0.01	0.01	1.00	1.33	1.33	100.00
ESTJ	2	3	150	0.03	0.04	2.00	2.67	4.00	200.00
ESTP	3	6	150	0.04	0.08	2.00	4.00	8.00	200.00
ESFJ	4	10	150	0.05	0.13	2.00	5.33	13.33	200.00
ESFP	6	16	150	0.08	0.21	2.00	8.00	21.33	200.00
ISTJ	6	22	155	0.08	0.29	2.07	8.00	29.33	206.67
ISTP	7	29	157	0.09	0.39	2.09	9.33	38.67	209.33
ISFJ	8	37	158	0.11	0.49	2.11	10.67	49.33	210.67
ISFP	9	46	166	0.12	0.61	2.21	12.00	61.33	221.33
ENTJ	10	56	166	0.13	0.75	2.21	13.33	74.67	221.33
ENTP	6	62	166	0.08	0.83	2.21	8.00	82.67	221.33
ENFJ	5	67	166	0.07	0.89	2.21	6.67	89.33	221.33
ENFP	3	70	166	0.04	0.93	2.21	4.00	93.33	221.33
INTJ	2	72	167	0.03	0.96	2.23	2.67	96.00	222.67

INTP	1	73	169	0.01	0.97	2.25	1.33	97.33	225.33
INFJ	1	74	172	0.01	0.99	2.29	1.33	98.67	229.33
INFP	1	75	176	0.01	1.00	2.35	1.33	100.00	234.67
TOTAL	75						100.00		

3.3. Example intervalos

Solución. Edades de cierta comunidad

25 35 38 45 47 48 51 52 53 55 60 62 63 66 67 70 71 72 75 77 78 81 88 89 90 99

Clase	f_i
$[20 - 30 >$	1
$[30 - 40 >$	2
$[40 - 50 >$	3
$[50 - 60 >$	4
$[60 - 70 >$	5
$[70 - 80 >$	6
$[80 - 90 >$	3
$[90 - 100]$	2

Tabulando

Tabla 3.6: Datos cuantitativos (intervalos)

Clase	f_i	F_i	F_i^*	h_i	H_i	H_i^*	$h_i\%$	$H_i\%$	$H_i^*\%$
$[20 - 30)$	1	1	26	0.02	0.01	0.35	2.38	1.33	34.67
$[30 - 40)$	2	3	68	0.05	0.04	0.91	4.76	4.00	90.67
$[40 - 50)$	3	6	68	0.07	0.08	0.91	7.14	8.00	90.67
$[50 - 60)$	4	10	68	0.10	0.13	0.91	9.52	13.33	90.67
$[60 - 70)$	5	15	68	0.12	0.20	0.91	11.90	20.00	90.67
$[70 - 80)$	6	21	84	0.14	0.28	1.12	14.29	28.00	112.00
$[80 - 90)$	3	24	84	0.07	0.32	1.12	7.14	32.00	112.00
$[90 - 100]$	2	26	84	0.05	0.35	1.12	4.76	34.67	112.00
TOTAL	42						61.90		



4

Gráficos estadísticos

4.1. Histograma de frecuencias

4.2. Circulares

4.3. Histograma de frecuencias



5

Medidas de tendencia central

Son aquellas medidas que buscan un dato representativo central de un conjunto de datos tales como la media, la moda y la mediana.

5.1. La media (\bar{x})

A veces llamada promedio aritmético, es la medida de tendencia central que pondera los datos.

5.1.1. Media de datos no agrupados

Los datos no están agrupados cuando no están ordenados sobre una tabla de distribución de frecuencias. Sean los n datos x_1, x_2, \dots, x_n entonces la media o promedio aritmético se define como

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

5.1.2. Media de datos agrupados

Considérese la siguiente tabla de distribución de frecuencias entonces el promedio es

$$\bar{x} = \frac{y_1 f_1 + y_2 f_2 + \dots + y_n f_n}{n} = \frac{1}{n} \sum_{i=1}^n y_i f_i$$

Tabla 5.1: Captionww

Clase	Y_i	f_i	F_i	...	H_i^*	%
$[y_1, y_2)$	y_1	f_1	H_1^*	%
$[y_2, y_3)$	y_2	f_2	H_1^*	%
$[y_3, y_4)$	y_3	f_3	H_1^*	%
\vdots	\vdots	\vdots	\vdots	\vdots

Clase	Y_i	f_i	F_i	\dots	H_i^*	%
$[y_{r-1}, y_r]$	y_r	f_r	\dots	\dots	H_1^*	%

Ejercicio 5.1. Si el promedio de n datos es \bar{x} entonces el promedio del conjunto inicial más un dato adicional x_{n+1} es

$$\bar{x}' = \frac{n\bar{x} + x_{n+1}}{n + 1}$$

en general si se adicionan r datos y_1, y_2, \dots, y_r entonces el nuevo promedio será

$$\bar{x}' = \frac{n\bar{x} + y_1 + y_2 + \dots + y_r}{n + r}$$

Solución. En efecto sea el promedio

$$\begin{aligned} \bar{x}' &= \frac{x_1 + x_2 + \dots + x_{n+1}}{n + 1} \\ &= \frac{n \frac{x_1 + x_2 + \dots + x_n}{n} + x_{n+1}}{n + 1} \\ &= \frac{n\bar{x} + x_{n+1}}{n + 1} \end{aligned}$$

5.2. La moda (Mo)

El dato que mas se repite

5.2.1. Moda de datos no tabulados

En este caso es dato que más repite en un conjunto de datos dados.

La moda es el dato que más se repite por ejemplo sea el conjunto de datos x_1, x_2, x_2, x_2, x_3 entonces la moda es $Mo = x_2$

5.2.2. Moda de datos tabulados

La moda es el dato que más se repite por ejemplo sea el conjunto de datos tabulados entonces la moda es

$$Mo = Li + \frac{Li - Ls}{(Li + Ls) - (L + Ls)}r$$