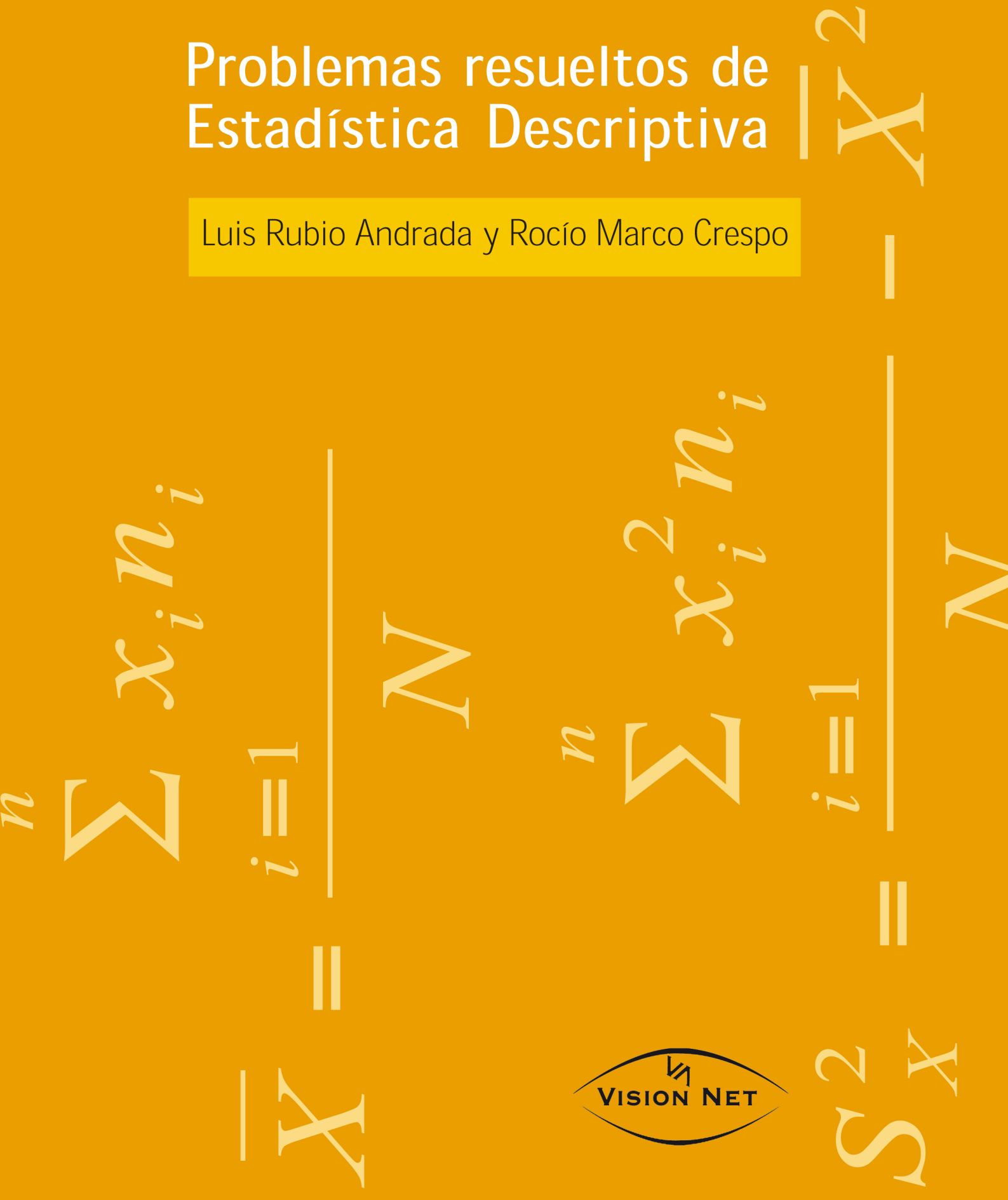
Problemas resueltos de Estadística Descriptiva

Luis Rubio Andrada y Rocío Marco Crespo



VISION NET

Problemas Resueltos de Estadística

Descriptiva



Problemas Resueltos de Estadística Descriptiva

Autores:

Luis Rubio Andrada, doctor en economía y profesor asociado del Departamento de Economía Aplicada de la Universidad Autónoma de Madrid.

Rocío Marco Crespo, doctora en economía y profesora asociada del Departamento de Economía Aplicada de la Universidad Autónoma de Madrid.

© Autores
© Editorial Vision Libros
Avenida de Asturias s/n Impares local 5
Semiesquina a Plaza Castilla 3
28029 Madrid (España)
Tel: 0034 91 3117696
url: www.visionlibros.com
ISBN PAPEL: 978-84-9770-003-0
ISBN Ebook: 978-84-9011-711-8
Descargado de: www.visionlibros.com
Diseño base cubierta: © Autores.

Producido por Grupo Corporativo Visionnet.
Cuatro Amigos 5 Posterior 28029 Madrid (España).

Impreso en España.

Reservados todos los derechos. Esta publicación no puede ser reproducida, ni en todo, ni en parte, ni registrada en o transmitida por un sistema de recuperación de información, de ninguna forma ni por ningún medio, sea mecánico, fotoquímico, electrónico, magnético, o por fotocopia, o cualquier otro sin el permiso previo por escrito de los autores.

"Necesitamos la estadística no sólo para explicar las cosas, sino para saber con precisión qué es lo que hay que explicar."

J.A. Schumpeter 1954

A Pilar

L.R.

A Santi

R.M.

PRÓLOGO

Este libro es el resultado de la experiencia acumulada por los autores a lo largo de más de dieciocho años de clases de estadística descriptiva, impartidas en las aulas de la Universidad Autónoma de Madrid y otros centros educativos de España. La interacción siempre fructífera con los estudiantes de las Licenciaturas de Administración de Empresas, Economía y Antropología Social, así como del Título Propio en Gestión Aeronáutica, de diversos cursos del Fondo Social Europeo, Master en Fundaciones, Master en Marketing y la Europäische Wirtschaftsakademie de Madrid, entre otros, ha sido la inspiración fundamental de estas páginas.

El libro se estructura en seis capítulos. En el primero y segundo se ven problemas para el estudio de una sola variable o atributo, considerando tanto los conceptos estadísticos básicos, como las distribuciones, gráficos y medidas necesarias para la descripción de un solo carácter. Así mismo, estos capítulos pueden ser empleados como repaso de la exploración de datos en cursos de probabilidad e inferencia estadística.

Los capítulos 3, 4 y 5 se han dedicado a problemas de relación entre dos o más caracteres, deteniéndose en el estudio de la asociación entre ellos, sean variables o atributos, relaciones lineales y no lineales entre dos variables y regresión múltiple.

En el capítulo 6 y último, se ven problemas esenciales para un estudiante de economía que serán utilizados en varias asignaturas a lo largo de sus estudios de grado, las tasas y los números índices.

Por lo tanto, los 67 problemas, algunos de ellos de desarrollo bastante extenso, son un buen complemento a las clases de teoría de un curso de estadística descriptiva. Permitiendo, dadas las explicaciones, fórmulas y comentarios contenidos en el texto, ser un instrumento eficaz para el repaso de conceptos básicos incluso en cursos de postgrado.

Al final del libro se ofrece una bibliografía de apoyo, dividida en teoría y problemas, que ha sido manejada extensamente por los autores a lo largo de estos lustros.

Por último, señalar que los problemas han sido resueltos con una hoja de cálculo: Excel XP, herramienta muy eficaz que permite una mejor comprensión de los conceptos estadísticos y que debe ser empleada en las prácticas de cualquier curso que se precie. Sin embargo, es posible que cuando los estudiantes resuelvan aquéllos con la calculadora, se encuentren con diferencias pequeñas en los resultados, en la mayoría de las ocasiones de milésimas, que les pueden perturbar en alguna medida, les pedimos disculpas por todo ello.

Capítulo 1

CONCEPTOS ESTADÍSTICOS, TABLAS Y REPRESENTACIONES GRÁFICAS

Problema 1.1.

Señala cuál es la población objeto de análisis en los estudios siguientes:

a) Grado de satisfacción con el servicio prestado por un determinado hotel.

b) Puntualidad en las llegadas de los aviones al Aeropuerto de Madrid-Barajas.

c) Resistencia de la carrocería de un coche, de una marca y un modelo determinado, ante un choque frontal.

d) Introducción de un nuevo refresco.

e) Hábitos culturales de los tharu.

f) Elecciones generales al parlamento.

Solución

Población, en sentido estadístico, es el conjunto de elementos que son objeto de estudio y que comparten una o varias características. Estos elementos son muy variados: campos de golf, bacterias, asalariados, neveras, niños…Por lo tanto, la solución de los casos propuestos es:

a) Los clientes que ha tenido el hotel.

b) Los vuelos que han llegado al Aeropuerto.

c) Los coches que se han fabricado de esa marca y modelo.

d) Los consumidores potenciales de esa bebida.

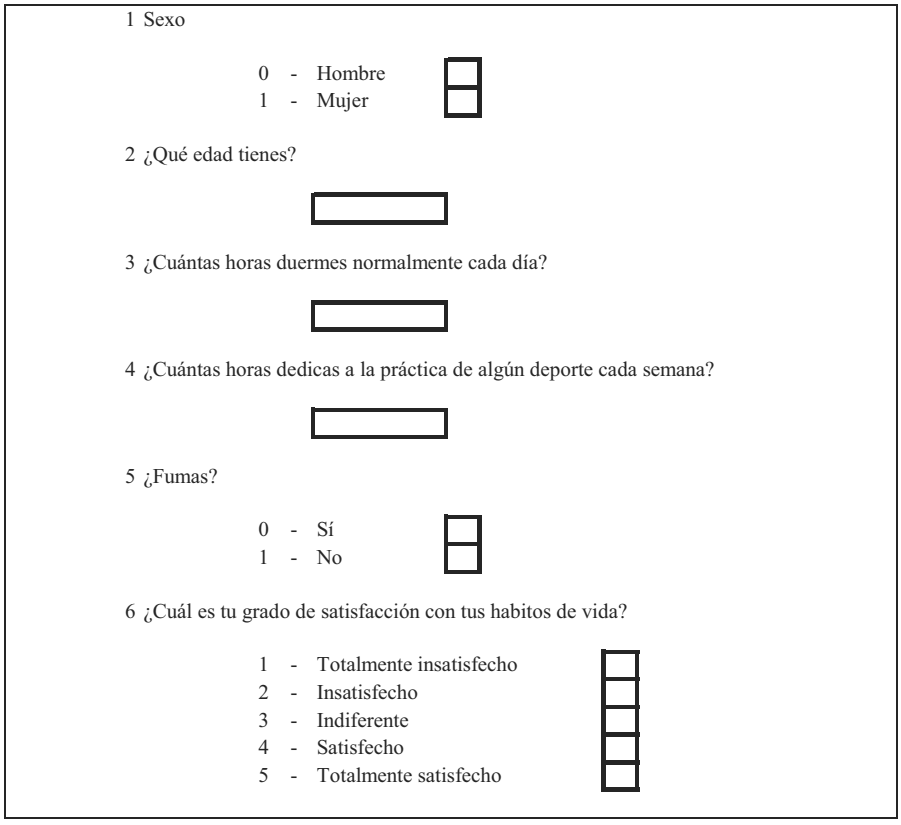
e) Los individuos que viven en la región de Tarai, situada al sur del Nepal.

f) Los votantes.

Problema 1.2.

El cuestionario es el instrumento de medida habitual que se emplea en las investigaciones sociales para recoger las respuestas de las unidades informantes a cada una de las características objeto de interés.

Con la finalidad de estudiar algunos de los hábitos de vida de los alumnos del Master de Marketing de la UAM, se distribuyó el siguiente cuestionario estructurado entre los mismos:



Considerando dicho cuestionario, señala la escala métrica de cada una de las preguntas recogidas en el mismo.

Solución

> Las preguntas 1 –sexo- y 5 -¿Fumas?- están en escala nominal, dado que son atributos en el que las categorías (respuestas) no son objeto de ordenación.

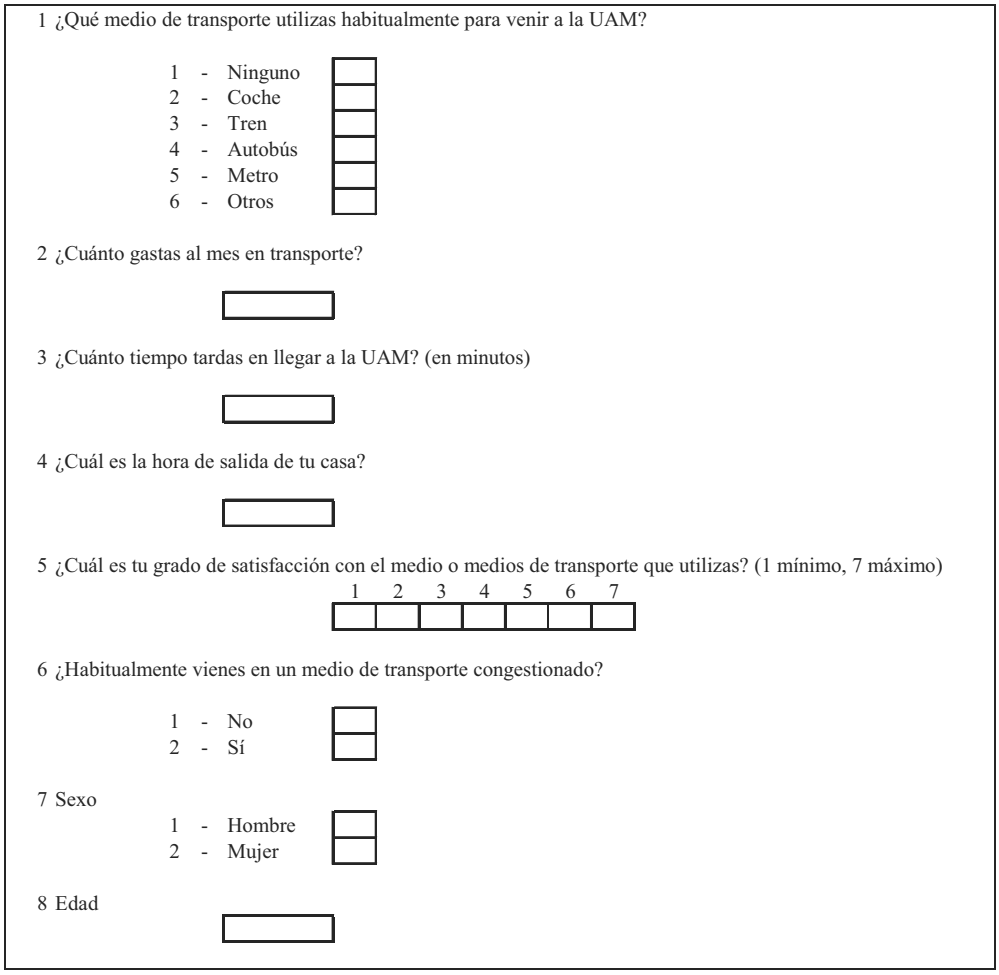
> Las preguntas 2 -¿Qué edad tienes?-, 3 -¿Cuántas horas duermes cada día?- y 4 -¿Cuántas horas dedicas a la práctica de algún deporte cada semana?-, son variables en escala proporción, dado que tienen un punto cero real en su origen y la proporción de un valor a otro cualesquiera de la escala es independiente de la unidad de medida con la que se trabaje.

> Y la pregunta 6 -¿Cuál es tu grado de satisfacción con tus hábitos de vida?- está en escala ordinal dado que es un atributo en el que las categorías se pueden ordenar.

Problema 1.3.

Durante el curso 2002-2003, en el grupo de tercero de la Licenciatura de Antropología Social y Cultural, se realizó un cuestionario sobre el medio de transporte utilizado para venir a la Universidad.

Considerando dicho cuestionario, señala la escala métrica de cada una de las preguntas recogidas en el mismo.


1 ¿Qué medio de transporte utilizas habitualmente para venir a la UAM?

1 - Ninguno
2 - Coche
3 - Tren
4 - Autobús
5 - Metro
6 - Otros

2 ¿Cuánto gastas al mes en transporte?

3 ¿Cuánto tiempo tardas en llegar a la UAM? (en minutos)

4 ¿Cuál es la hora de salida de tu casa?

5 ¿Cuál es tu grado de satisfacción con el medio o medios de transporte que utilizas? (1 mínimo, 7 máximo)

1 2 3 4 5 6 7

6 ¿Habitualmente vienes en un medio de transporte congestionado?

1 - No
2 - Sí

7 Sexo

1 - Hombre
2 - Mujer

8 Edad

Solución

> Las preguntas 1 -¿Qué medio de transporte utilizas habitualmente para venir?- 6 -¿Habitualmente vienes en un medio de transporte congestionado?- y 7 –sexo- están en escala nominal porque son atributos en los que las categorías no se pueden ordenar.

> La pregunta 5 -¿Cuál es tu grado de satisfacción con el medio o medios de transporte que utilizas?- está en escala ordinal dado que es un atributo en el que las categorías se pueden ordenar.

> Las preguntas 2 -¿Cuánto gastas al mes en transporte?-, 3 -¿Cuánto tiempo tardas en llegar a la UAM?-, 4 -¿Cuál es la hora de salida de tu casa?- y 8 –Edad- son variables en

escala proporción, dado que tienen un punto cero real en su origen y la proporción de un valor a otro cualesquiera de la escala es independiente de la unidad de medida con la que se trabaje.

Problema 1.4.

Indica para los caracteres siguientes:

a) Nivel de estudios.
b) Talla de la ropa.
c) Estado civil.
d) Fecha de inicio de las vacaciones estivales.
e) Número de televisores en tu vivienda.
f) Altura de una persona.

Si son variables o atributos, la escala de medida a la que pertenecen y si son variables discretas o continuas.

Solución

a) Es un atributo en escala ordinal.

b) Es una variable discreta en escala de intervalos. Es de esa escala porque el punto cero y la unidad de medida son arbitrarios.

c) Es un atributo en escala nominal.

d) Es una variable discreta en escala de intervalos.

e) Es una variable discreta en escala proporción.

f) Es una variable continua en escala proporción.

Problema 1.5.

Indica para los caracteres siguientes:

a) Distrito postal.
b) Precio de un coche.
c) Idiomas que habla.
d) Fecha de nacimiento.
e) Beneficio neto de una empresa.
f) Número de vuelo.

Si son variables o atributos, la escala de medida a la que pertenecen y si son variables discretas o continuas.

Solución

a) Es un atributo en escala nominal al que se le asigna un código numérico de identificación.
b) Es una variable continua en escala proporción.
c) Es un atributo en escala nominal.
d) Es una variable discreta en escala de intervalos.
e) Es una variable continua en escala proporción.
f) Es un atributo en escala nominal al que se le asigna un código alfanumérico.

Problema 1.6.

El cuestionario descrito en el problema 1.2. ha sido cumplimentado por 27 personas y sus respuestas aparecen en la tabla siguiente, siendo: CASO el número de identificación; SEXO la característica que indica el sexo de los entrevistados que se ha codificado como "0" si hombre y "1" si mujer; EDAD la pregunta ¿Qué edad tienes?; DORMIR la pregunta ¿Cuántas horas duermes normalmente cada día?; DEPORTE la pregunta ¿Cuántas horas dedicas a la práctica de algún deporte cada semana?; FUMAS la pregunta de ¿Fumas? siendo "0" el código para la categoría sí y "1"

el código para la categoría no; y SATISHAB la pregunta ¿Cuál es tu grado de satisfacción con tus hábitos de vida? siendo "1" totalmente insatisfecho, "2" insatisfecho, "3" indiferente, "4" satisfecho y "5" totalmente satisfecho.

CASO	SEXO	EDAD	DORMIR	DEPORTE	FUMAS	SATISHAB
1	0	31	6	1	0	3
2	1	22	9	0	0	2
3	1	25	8	0	1	4
4	1	24	10	0	0	4
5	1	24	9	3	1	3
6	0	28	9	3	1	3
7	1	28	5	0	1	2
8	1	28	6	0	1	4
9	1	26	10	4	1	4
10	1	26	8	0	1	4
11	1	25	6	0	0	4
12	0	26	6	0	0	4
13	1	24	8	1	0	3
14	1	24	8	3	1	4
15	1	23	8	3	1	4
16	1	25	9	0	1	3
17	1	24	8	0	0	3
18	1	25	7	0	1	2
19	1	24	8	0	1	2
20	0	26	7	6	1	4
21	1	24	8	3	1	4
22	1	28	8	0	1	2
23	1	24	10	1	0	2
24	1	27	10	0	1	4
25	1	24	6	0	0	3
26	1	24	10	0	0	4
27	0	28	7	4	1	2

A la luz de la información expuesta, realiza las distribuciones de frecuencias de cada uno de los caracteres.

Solución

> La distribución de frecuencias de SEXO es:

A_i	n_i	N_i	f_i%	F_i%
Hombre (0)	5	5	18,519	18,519
Mujer (1)	22	27	81,481	100,000
Total	27		100,000	

Comentario: las frecuencias acumuladas, tanto absolutas como relativas, pueden calcularse pero no tienen sentido estadístico dado que dependen del orden, que es subjetivo, de colocación de cada una de las categorías. Esto ocurre en todas las distribuciones de frecuencias de características en escala nominal.

> La distribución de frecuencias de EDAD (en años) es:

X_i	n_i	N_i	f_i%	F_i%
22	1	1	3,704	3,704
23	1	2	3,704	7,407
24	10	12	37,037	44,444
25	4	16	14,815	59,259
26	4	20	14,815	74,074
27	1	21	3,704	77,778
28	5	26	18,519	96,296
31	1	27	3,704	100,000
Total	27		100,000	

> La distribución de frecuencias de DORMIR (en horas) es:

X_i	n_i	N_i	f_i%	F_i%
5	1	1	3,704	3,704
6	5	6	18,519	22,222
7	3	9	11,111	33,333
8	9	18	33,333	66,667
9	4	22	14,815	81,481
10	5	27	18,519	100,000
Total	27		100.000	

> La distribución de frecuencias de DEPORTE (en horas) es:

X_i	n_i	N_i	f_i%	F_i%
0	16	16	59.259	59.259
1	3	19	11,111	70,370
3	5	24	18,519	88,889
4	2	26	7,407	96,296
6	1	27	3,704	100,000
Total	27		100.000	

> La distribución de frecuencias de FUMAS es:

A_i	n_i	N_i	f_i%	F_i%
Sí(0)	10	10	37,037	37,037
No(1)	17	27	62,963	100,000
Total	27		100,000	

> La distribución de frecuencias de SATISHAB es:

A_i	n_i	N_i	f_i%	F_i%
Insatisfecho(2)	7	7	25,926	25,926
Indiferente (3)	7	14	25,926	51,852
Satisfecho (4)	13	27	48,148	100,000
Total	27		100,000	

Problema 1.7.

37 sujetos cumplimentaron un cuestionario sobre teléfonos móviles, recogiéndose sus respuestas en la tabla siguiente. Para su interpretación, CASO es el número de identificación de cada uno de los cuestionarios recogidos; SEXO un atributo en escala nominal que se le ha asignado el código "0" si hombre y "1" si mujer; EDAD una variable discreta en escala proporción que señala la edad de los sujetos; GASMOV una variable continua en escala proporción que informa sobre el gasto mensual en teléfono móvil; COMPA un atributo en escala nominal que indica la compañía suministradora del servicio con códigos "0" si Movistar, "1" si Vodafone y "2" si Amena; MODAL un atributo en escala nominal que señala la modalidad de pago, con códigos "0" si pago con tarjeta y "1" si pago a través de contrato y SATISF un atributo en escala ordinal que señala el grado de satisfacción con el servicio que presta el móvil, siendo "1" si totalmente insatisfecho, "2" si insatisfecho, "3" si indiferente, "4" si satisfecho y "5" si totalmente satisfecho.

CASO	SEXO	EDAD	GASMOV	COMPA	MODAL	SATISF
1	1	18	100	0	0	2
2	0	20	120	0	1	3
3	0	17	6	0	0	3
4	0	27	35	2	1	4
5	0	18	10	0	0	4
6	1	18	40	1	1	3
7	1	18	30	0	0	3
8	1	19	20	0	0	2
9	1	18	35	2	1	2
10	0	18	20	0	0	4
11	1	18	12	0	1	4
12	1	18	10	0	0	4
13	1	18	15	0	0	4
14	1	18	40	1	0	4
15	1	18	14	2	1	2
16	1	18	5	0	0	4
17	1	19	60	1	0	3
18	1	18	5	2	0	4
19	1	20	8	0	0	4
20	1	18	20	0	0	4
21	1	19	40	0	0	3
22	0	18	45	0	1	3
23	1	25	10	2	0	3
24	1	18	4	0	0	2
25	1	19	25	0	0	3
26	1	18	30	0	1	4
27	1	18	25	0	0	4
28	1	17	25	2	0	5
29	1	17	18	0	1	4
30	1	17	15	1	0	4
31	1	20	30	2	0	4
32	1	17	25	0	0	3
33	1	18	20	2	0	3
34	1	18	50	0	0	4
35	0	18	9	0	0	3
36	0	18	8	0	0	3
37	0	20	10	1	0	3

Considerando los datos de la tabla anterior, realiza lo siguiente:

a) Las distribuciones de frecuencias de cada uno de los caracteres señalados.

b) La distribución agrupada en intervalos para GASMOV.

c) Los diagramas de barras para SEXO, EDAD y COMPA, considerando en el eje de ordenadas las frecuencias absolutas, y para GASMOV, MODAL y SATISF, situando las frecuencias relativas en el eje de ordenadas.

d) Los polígonos de frecuencias para EDAD, GASMOV y SATISF.

e) Los diagramas en escalera para EDAD, GASMOV y SATISF.

Solución

a)

> La distribución de frecuencias de SEXO es:

A_i	n_i	N_i	f_i%	F_i%
Hombre (0)	9	9	24,324	24,324
Mujer (1)	28	37	75,676	100,000
Total	37		100,000	

> La distribución de frecuencias de EDAD (en años) es:

X_i	n_i	N_i	f_i%	F_i%
17	5	5	13,514	13,514
18	22	27	59,459	72,973
19	4	31	10,811	83,784
20	4	35	10,811	94,595
25	1	36	2,703	97,297
27	1	37	2,703	100,000
Total	37		100,000	

> La distribución de frecuencias de COMPA es:

A_i	n_i	N_i	f_i%	F_i%
Movistar (0)	24	24	64,865	64,865
Vodafone (1)	5	29	13,514	78,378
Amena (2)	8	37	21,622	100,000
Total	37		100,000	

> La distribución de frecuencias de GASMOV (en euros) es:

X_i	n_i	N_i	f_i%	F_i%
4	1	1	2,703	2,703
5	2	3	5,405	8,108
6	1	4	2,703	10,811
8	2	6	5,405	16,216
9	1	7	2,703	18,919
10	4	11	10,811	29,730
12	1	12	2,703	32,432
14	1	13	2,703	35,135
15	2	15	5,405	40,541
18	1	16	2,703	43,243
20	4	20	10,811	54,054
25	4	24	10,811	64,865
30	3	27	8,108	72,973
35	2	29	5,405	78,378
40	3	32	8,108	86,486
45	1	33	2,703	89,189
50	1	34	2,703	91,892
60	1	35	2,703	94,595
100	1	36	2,703	97,297
120	1	37	2,703	100,000
Total	37		100.000	

> La distribución de frecuencias de MODAL es:

A_i	n_i	N_i	f_i%	F_i%
Tarieta (0)	28	28	75.676	75.676
Contrato (1)	9	37	24,324	100,000
Total	37		100.000	

> La distribución de frecuencias de SATISF es:

A_i	n_i	N_i	f_i%	F_i%
Insatisfecho (2)	5	5	13.514	13.514
Indiferente (3)	14	19	37,838	51,351
Satisfecho (4)	17	36	45,946	97,297
Totalmente satisfecho (5)	1	37	2,703	100,000
Total	37		100.000	

b)

La distribución agrupada en intervalos de GASMOV es:

L_{i-1}-L_i	n_i	N_i	f_i%	F_i%
0-10	11	11	29,730	29,730
10-20	9	20	24,324	54,054
20-30	7	27	18,919	72,973
30-40	5	32	13,514	86,486
40-50	2	34	5,405	91,892
50-100	2	36	5,405	97,297
100-150	1	37	2,703	100,000
Total	37		100,000	