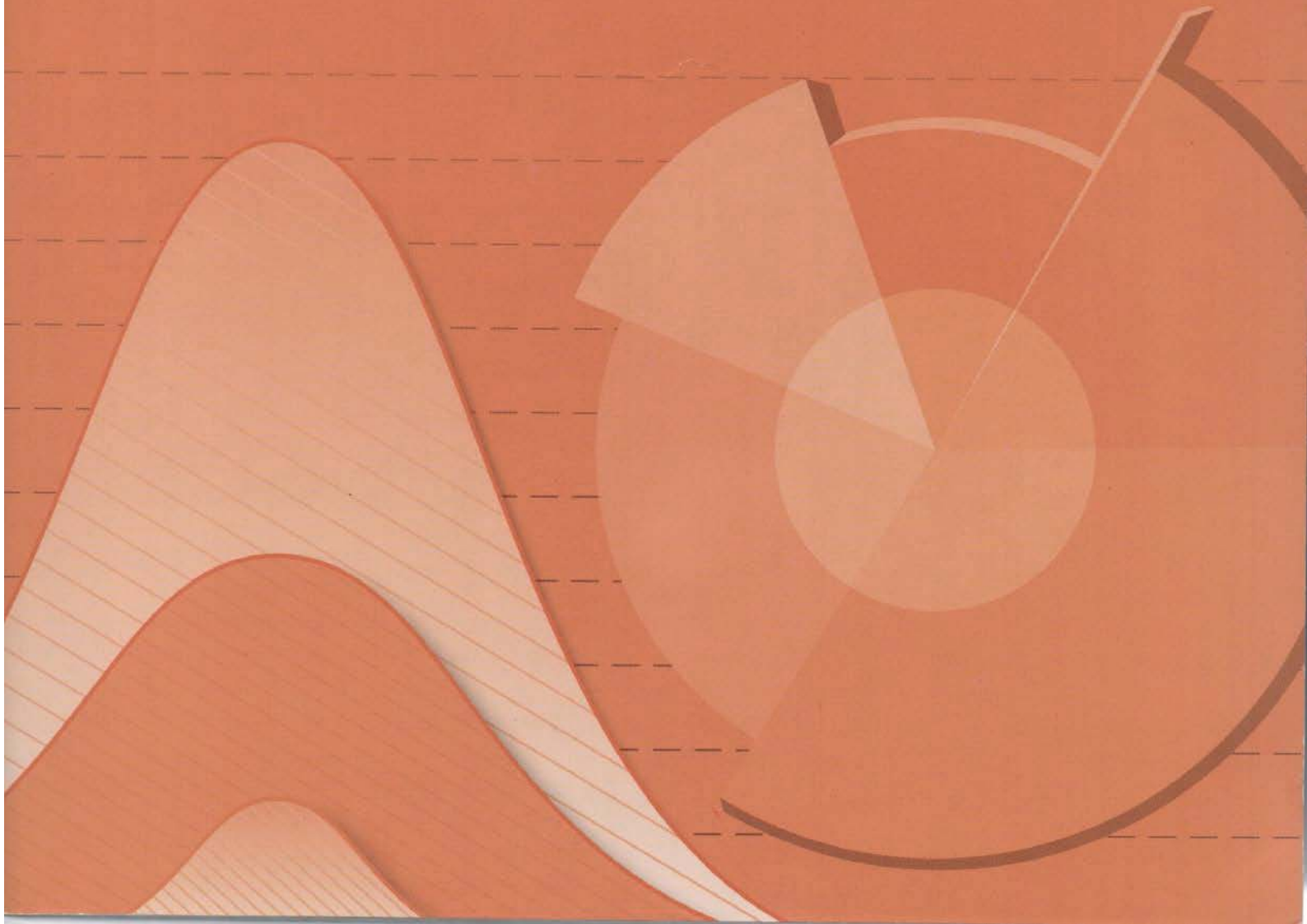




EUNED
EDITORIAL
UNIVERSIDAD
ESTATAL
A DISTANCIA

Miguel Gómez Barrantes

ELEMENTOS
DE **ESTADÍSTICA**
DESCRIPTIVA





PROMAI
Programa de Producción
de Material Impreso

Editora académica y asesora metodológica
Virginia Ramírez Cascante
Producción de Material Didáctico
Escrito, UNED

- © MIGUEL GÓMEZ BARRANTES
- © Sobre la presente edición
Editorial Universidad Estatal a Distancia

Revisión filológica
María Benavides González
Coordinación de producción académica
Ana Beatriz Picado González,
Gustavo Hernández Castro
Encargado de cátedra
Greibin Villegas Barahona
Especialista
Breda Muñoz Hernández
Elisa Sánchez Godínez
Ilustraciones
Luis Fernando Quirós Abarca
Diagramación y artes finales
Luis Fernando Quirós Abarca
Diseño de portada
Carlos Villalobos Argüello
Coordinador de producción
Daniel Villalobos Gamboa
Imposición digital
Christian Cascante Brenes

CUARTA EDICIÓN
Editorial Universidad Estatal a Distancia
San José, Costa Rica, 2012

ISBN 978-9968-31-934-8

310
G633e4 Gómez Barrantes, Miguel
Elementos de estadística descriptiva /
Miguel Gómez Barrantes.--San José, C.R. :
EUNED, 2012.
664 p.

ISBN 978-9968-31-934-8

1. ESTADISTICA. I. Título.

Impreso en Costa Rica.
Reservados todos los derechos.
Prohibida la reproducción no autorizada
por cualquier medio, mecánico o electrónico,
del contenido total o parcial de esta publicación.
Hecho el depósito de ley.

CONTENIDO

Miguel Gómez Barrantes y el estudio de la Estadística	IX
Presentación	XIII
Capítulo 1 NATURALEZA DE LA ESTADÍSTICA	
Objetivos	2
Resumen	3
1.1. INTRODUCCIÓN	3
1.2. DATOS ESTADÍSTICOS O ESTADÍSTICAS	3
1.3. LA ESTADÍSTICA COMO DISCIPLINA CIENTÍFICA	5
1.4. UNIDAD ESTADÍSTICA, CARACTERÍSTICA Y OBSERVACIÓN	6
1.5. EL CONCEPTO DE POBLACIÓN	8
1.6. USO DEL MUESTREO	11
1.7. SELECCIÓN DE MUESTRA	13
1.8. LA TABLA DE NÚMEROS AL AZAR	17
1.9. ATRIBUTOS Y VARIABLES	21
1.10. ESTADÍSTICA DESCRIPTIVA Y ESTADÍSTICA INFERENCIAL	24
Ejercicios de autoevaluación	27
Respuesta a los ejercicios de autoevaluación	30
Capítulo 2 FUENTES DE DATOS Y TÉCNICAS DE RECOLECCIÓN	
Objetivos	34
Resumen	35
2.1. INTRODUCCIÓN	35
2.2. LA OBTENCIÓN DE LA INFORMACIÓN ESTADÍSTICA: DATOS EXISTENTES Y NO EXISTENTES	36
2.3. FUENTES PRIMARIAS Y FUENTES SECUNDARIAS	37
2.4. LA NECESIDAD DE EVALUAR LAS FUENTES	41
2.5. TÉCNICAS PARA LA RECOLECCIÓN DE DATOS NO EXISTENTES	43
2.5.1. Experimento	43
2.5.2. Observación	44
2.5.3. Entrevista cara a cara	45

2.5.4.	Entrevista telefónica	45
2.5.5.	Correo	46
2.5.6.	Encuesta por internet.....	47
2.5.7.	Registro.....	48
2.5.8.	Encuestas autoadministradas a personas o grupos	49
2.6.	FASES EN UNA INVESTIGACIÓN ESTADÍSTICA.....	50
2.6.1.	Definición del problema.....	51
2.6.2.	Establecimiento de los propósitos específicos del estudio.....	52
2.6.3.	Preparación de un plan general de trabajo	52
2.6.4.	Construcción del cuestionario.....	53
2.6.5.	Diseño y selección de la muestra.....	54
2.6.6.	Preparación y ejecución del trabajo de campo	54
2.6.7.	Elaboración del plan de análisis	55
2.6.8.	El procesamiento de la información.....	56
2.6.9.	Análisis e interpretación de los datos	56
2.6.10.	Preparación del informe.....	56
	Ejercicios de autoevaluación.....	57
	Respuesta a los ejercicios de autoevaluación.....	61

Capítulo 3 EL CUESTIONARIO

	Objetivos.....	66
	Resumen	67
3.1.	EL CUESTIONARIO: CARACTERÍSTICAS Y FUNCIONES.....	67
3.1.1.	Las funciones del cuestionario.....	68
3.1.2.	Preguntas de la investigación y del cuestionario	70
3.2.	LA FORMA DE LAS PREGUNTAS: ABIERTAS Y CERRADAS.....	72
3.2.1.	Preguntas cerradas.....	73
3.2.2.	Batería de preguntas.....	74
3.2.3.	Preguntas abiertas.....	76
3.2.4.	Pregunta semiabierta.....	79
3.3.	VENTAJAS Y DESVENTAJAS COMPARATIVAS DE LAS PREGUNTAS ABIERTAS Y CERRADAS	80
3.4.	TIPOS DE PREGUNTAS SEGÚN INFORMACIÓN QUE INDAGAN.....	81
3.4.1.	Preguntas de hechos.....	82
3.4.2.	Preguntas de acciones o conductas pasadas	83
3.4.3.	Preguntas de intención.....	83
3.4.4.	Preguntas de opinión	85
3.4.5.	Preguntas de actitudes	85
3.5.	DETALLES ACERCA DE LA CONSTRUCCIÓN DE CUESTIONARIOS.....	87
3.5.1.	El orden de las preguntas	87
3.5.2.	Preguntas filtro	89
3.5.3.	El número de preguntas y la duración del cuestionario	89
3.5.4.	La modulación de los cuestionarios.....	90
3.5.5.	Algunos problemas frecuentes en la redacción de las preguntas	91
3.5.6.	Instrucciones para el entrevistador	94

3.5.7.	La prueba del cuestionario	94
3.6.	PROCESAMIENTO DE LOS DATOS	95
3.6.1.	Revisión de los cuestionarios	96
3.6.2.	Codificación	97
3.6.3.	Digitación o captura de los datos	98
3.7	EJEMPLO DE UN CUESTIONARIO	102
	Ejercicios de autoevaluación.....	114
	Respuesta a los ejercicios de autoevaluación.....	119
Capítulo 4	LOS NÚMEROS RELATIVOS	
	<i>Objetivos</i>	126
	<i>Resumen</i>	127
4.1.	LA NECESIDAD DE RESUMIR Y COMPARAR LA INFORMACIÓN: LOS NÚMEROS RELATIVOS	127
4.2.	RAZONES, PROPORCIONES Y PORCENTAJES: CÁLCULO E INTERPRETACIÓN	128
4.2.1.	Razón	128
4.2.2.	Proporción.....	129
4.2.3.	Porcentajes	129
4.2.4.	Amplificación de razones y proporciones.....	130
4.2.5.	La importancia de la base en la razón o proporción amplificada	131
4.3.	LAS TASAS: CONCEPTO, CÁLCULO E INTERPRETACIÓN	134
4.3.1.	Las tasas vitales	134
4.3.2.	Tasas de crecimiento	135
4.3.3.	El cálculo de las tasas de crecimiento	138
4.3.4.	Comparación de las tasas de crecimiento.....	141
4.4.	ALGUNOS NÚMEROS RELATIVOS DE USO FRECUENTE	142
4.4.1.	Porcentaje de población urbana.....	143
4.4.2.	Densidad de población.....	143
4.4.3.	Producto interno bruto per cápita	143
4.4.4.	Tasa de participación en actividad económica o tasa de actividad	144
4.4.5.	Tasa de desempleo o porcentaje de desocupados	145
4.4.6.	Tasa de mortalidad infantil.....	146
4.4.7.	Índice de masa corporal	147
4.5.	ERRORES FRECUENTES EN EL USO DE NÚMEROS RELATIVOS.....	148
4.5.1.	Errores originados en la forma de presentación o anotación	148
4.5.2.	Confusión con respecto a la base	148
4.5.3.	Cálculo de proporciones o razones basadas en un número pequeño de casos	149
4.5.4.	Errores al promediar números relativos	151
4.6.	LOS ÍNDICES: CONCEPTO, TIPOS, INTERPRETACIÓN	152
4.6.1.	Índice de desarrollo humano (IDH).....	154
4.6.2.	Índices simples y compuestos.....	155
4.7.	LOS ÍNDICES DE PRECIOS.....	160
4.7.1.	Agregado simple.....	160

4.7.2.	Promedio simple de relativos.....	161
4.7.3.	Índices de precios ponderados.....	163
4.8.	ÍNDICES DE PRECIOS EN COSTA RICA.....	166
4.8.1.	Selección del periodo base.....	167
4.8.2.	Selección de los artículos que deben ser incluidos en el índice	167
4.8.3.	Elección de la fórmula de cálculo del índice.....	168
4.8.4.	Ponderaciones o pesos que recibirán los artículos.....	168
4.8.5.	Procedimiento de recolección de la información para el cálculo regular del índice.....	168
4.8.6.	Cambios de calidad y productos nuevos	169
4.9.	USO DE LOS ÍNDICES DE PRECIOS.....	170
4.9.1.	Medición y seguimiento del poder adquisitivo de la moneda.....	171
4.9.2.	Cálculo de valores reales o deflactados	172
4.9.3.	Ajustes salariales y de contratos	176
4.9.4.	Ajustes de series de ventas y de valores de inventarios.....	177
4.10.	EL ÍNDICE DE PRECIOS AL CONSUMIDOR (IPC) DE COSTA RICA.....	177
	Ejercicios de autoevaluación.....	181
	Respuesta a los ejercicios de autoevaluación	190

Capítulo 5 CONSTRUCCIÓN Y ANÁLISIS DE CUADROS ESTADÍSTICOS

Objetivos.....	200	
Resumen	201	
5.1.	DEFINICIONES Y CLASIFICACIONES ESTADÍSTICAS	201
5.2.	LAS SERIES ESTADÍSTICAS	203
5.2.1.	Series cuantitativas	203
5.2.2.	Series cualitativas.....	204
5.2.3.	Series geográficas.....	205
5.2.4.	Series cronológicas o de tiempo.....	206
5.3.	FORMAS DE PRESENTACIÓN DE LA INFORMACIÓN ESTADÍSTICA	207
5.3.1.	Presentación textual.....	208
5.3.2.	Presentación semitabular.....	209
5.4.	PRESENTACIÓN TABULAR: LOS COMPONENTES DEL CUADRO	210
5.4.1.	Los componentes del cuadro.....	212
5.5.	CUADROS GENERALES DE REFERENCIA Y DE RESUMEN O DERIVADOS.	215
5.5.1.	Cuadros generales o de referencia.....	215
5.5.2.	Cuadros de resumen o derivados.....	216
5.6.	FORMAS USUALES DE ORDENAMIENTO DE LAS CATEGORÍAS EN LA COLUMNA MATRIZ.....	217
5.6.1.	Alfabético	217
5.6.2.	Magnitud.....	217
5.6.3.	Cronológico.....	217
5.6.4.	Geográfico	218
5.6.5.	Usual.....	218
5.6.6.	Progresivo	218
5.7.	DETALLES SOBRE LA CONSTRUCCIÓN DE CUADROS	218

5.7.1.	Necesidad de destacar ciertas cifras.....	219
5.7.2.	Comparaciones.....	219
5.7.3.	Tamaño y forma del cuadro.....	220
5.7.4.	Rayado.....	221
5.7.5.	Indicación de ceros y cifras faltantes.....	222
5.8.	EJEMPLOS CONCRETOS SOBRE LA UTILIDAD DE LOS CUADROS.....	223
5.9.	EL CUADRO COMO INSTRUMENTO ANALÍTICO: CRITERIOS PARA EL CÁLCULO DE PORCENTAJE EN LOS CUADROS.....	227
5.10	EJERCICIO ILUSTRATIVO.....	230
	Ejercicios de autoevaluación.....	233
	Respuesta a los ejercicios de autoevaluación.....	239

Capítulo 6 **CONSTRUCCIÓN Y ANÁLISIS DE GRÁFICOS ESTADÍSTICOS**

	Objetivos.....	246
	Resumen.....	247
6.1.	UTILIDAD Y CARACTERÍSTICAS DESEABLES DE LOS GRÁFICOS ESTADÍSTICOS.....	247
6.1.1.	Características deseables en los gráficos.....	248
6.1.2.	Una observación importante.....	250
6.1.3.	Tipos de gráficos.....	250
6.2.	LOS GRÁFICOS DE BARRAS.....	252
6.2.1.	Gráfico de barras simples.....	256
6.2.2.	Gráfico de barras compuestas.....	257
6.2.3.	Gráfico de barras comparativa.....	258
6.2.4.	Gráfico de barras verticales.....	259
6.2.5.	Gráfico de barras de dos direcciones.....	260
6.3.	GRÁFICO CIRCULAR Y BARRA 100%.....	261
6.4.	GRÁFICOS LINEALES.....	266
6.5.	GRÁFICOS LINEALES ARITMÉTICOS.....	267
6.5.1.	Interpretación de los gráficos lineales aritméticos.....	270
6.5.2.	Importancia de la base 0.....	272
6.5.3.	Gráfico radial o "de araña".....	274
6.6.	GRÁFICO SEMILOGARÍTMICO.....	275
6.6.1.	Series aritméticas y geométricas.....	275
6.6.2.	Construcción de un gráfico semilogarítmico.....	278
6.6.3.	Interpretación y uso de los gráficos semilogarítmico.....	279
6.6.4.	Gráfico aritmético con dos escalas verticales.....	283
6.6.5.	Algunas guías para interpretar los gráficos semilogarítmicos.....	285
6.7.	PICTOGRAMAS.....	286
6.8.	MAPAS ESTADÍSTICOS.....	288
6.8.1.	Mapas sombreados.....	288
6.8.2.	Mapas punteados.....	288
	Ejercicios de autoevaluación.....	291
	Respuesta a los ejercicios de autoevaluación.....	298

Capítulo 7 DISTRIBUCIONES DE FRECUENCIAS

Objetivos	310
Resumen	311
7.1. NECESIDAD DE RESUMIR LA INFORMACIÓN CUANTITATIVA: LA DISTRIBUCIÓN DE FRECUENCIAS.....	311
7.2. DISTRIBUCIÓN DE FRECUENCIAS DE VARIABLES DISCRETAS.....	313
7.3. LA MEDICIÓN DE LAS VARIABLES CONTINUAS Y EL PROBLEMA DEL REDONDEO.....	317
7.3.1. Redondeo a la unidad más próxima	317
7.3.2. Redondeo hacia abajo.....	319
7.3.3. Redondeo hacia arriba	320
7.3.4. Algunos ejemplos de redondeos.....	320
7.4. DISTRIBUCIÓN DE FRECUENCIAS DE VARIABLES CONTINUAS.....	321
7.5. LÍMITES REALES Y LÍMITES INDICADOS, INTERVALO DE CLASE Y PUNTO MEDIO	325
7.5.1. Límites de clase	325
7.5.2. Intervalo de clase	327
7.5.3. Punto medio.....	328
7.5.4. Clases abiertas	329
7.6. FRECUENCIAS ABSOLUTAS Y RELATIVAS, SIMPLES Y ACUMULADAS	329
7.6.1. Frecuencia absoluta	329
7.6.2. Frecuencia relativa	329
7.6.3. Frecuencias acumuladas	330
7.7. LA REPRESENTACIÓN GRÁFICA DE LAS DISTRIBUCIONES DE FRECUENCIAS: HISTOGRAMAS, POLÍGONOS Y OJIVAS	331
7.7.1. Histograma	331
7.7.2. Polígono de frecuencias	333
7.7.3. Representación de distribuciones con intervalos desiguales	334
7.7.4. Las "ojivas" o polígonos de frecuencias acumuladas.....	336
Ejercicios de autoevaluación.....	339
Respuesta a los ejercicios de autoevaluación.....	343

Capítulo 8 MEDIDAS DE TENDENCIA CENTRAL

Objetivos	350
Resumen	351
8.1. LAS MEDIDAS DE TENDENCIA CENTRAL: IDEAS BÁSICAS.....	351
8.2. SÍMBOLO DE SUMATORIA.....	353
8.3. MODA, MEDIANA Y MEDIA ARITMÉTICA EN DATOS NO AGRUPADOS....	355
8.3.1. La moda (M_o)	355
8.3.2. La mediana (M_e)	357
8.3.3. La media aritmética (\bar{X})	358
8.3.4. La media aritmética simple	358
8.3.5. Media aritmética ponderada	359
8.3.6. Media aritmética global de varios conjuntos de datos	363

8.4.	PROPIEDADES DE LA MEDIA ARTIMÉTICA.....	363
8.5.	EL CÁLCULO DE LAS MEDIDAS DE POSICIÓN EN DATOS AGRUPADOS...	365
8.5.1.	La moda (Mo).....	365
8.5.2.	La mediana (Me).....	366
8.5.3.	La media aritmética (\bar{X}).....	368
8.6.	USO DE LAS MEDIDAS DE TENDENCIA CENTRAL.....	370
8.7.	LA MEDIA GEOMÉTRICA Y LA MEDIA ARMÓNICA.....	374
8.7.1.	Media geométrica.....	374
8.7.2.	Media armónica.....	377
8.7.3.	Relación entre la media aritmética, la geométrica y la armónica.....	378
	Ejercicios de autoevaluación.....	379
	Respuesta a los ejercicios de autoevaluación.....	382
Capítulo 9 MEDIDAS DE VARIABILIDAD O DISPERSIÓN		
	Objetivos.....	388
	Resumen.....	389
9.1.	LA VARIABILIDAD Y SU IMPORTANCIA.....	389
9.2.	LA MEDICIÓN DE LA VARIABILIDAD. EL RECORRIDO Y LA DESVIACIÓN MEDIA.....	392
9.2.1.	El recorrido o amplitud.....	393
9.2.2.	La desviación media.....	393
9.3.	LA DESVIACIÓN ESTÁNDAR Y LA VARIANCIA: CONCEPTO, DEFINICIÓN Y CÁLCULO EN DATOS SIMPLES Y AGRUPADOS.....	394
9.3.1.	El cálculo de la variancia en datos sin agrupar.....	396
9.3.2.	El cálculo de la variancia en datos agrupados en una distribución de frecuencias.....	397
9.3.3.	La variancia de un conjunto formado por la combinación de dos o más grupos de datos.....	399
9.3.4.	Variancia dentro y entre grupos.....	400
9.4.	VARIABILIDAD RELATIVA. EL COEFICIENTE DE VARIACIÓN.....	402
9.5.	ESTANDARIZACIÓN DE NOTAS Y PUNTAJES.....	404
9.6.	MEDIA Y VARIANCIA DE VARIABLES DICOTÓMICAS.....	406
9.7.	ILUSTRACIÓN INTEGRADA DE DISTRIBUCIONES DE FRECUENCIAS Y MEDIDAS DE TENDENCIA CENTRAL Y VARIABILIDAD.....	408
9.8.	MEDIDAS DE POSICIÓN O CUANTILIOS.....	410
9.8.1.	Recorrido intercuartil (RIC).....	414
9.9.	DIAGRAMA DE CAJA.....	415
	Ejercicios de autoevaluación.....	417
	Respuesta a los ejercicios de autoevaluación.....	421

Capítulo 10 INTRODUCCIÓN A LAS PROBABILIDADES

Objetivos	426
Resumen	427
10.1. INFERENCIA ESTADÍSTICA Y PROBABILIDAD	427
10.2. EL CONCEPTO DE PROBABILIDAD	428
10.3. EVENTOS Y ESPACIO MUESTRAL	430
10.4. LA DEFINICIÓN DE PROBABILIDAD. ENFOQUE CLÁSICO	434
10.5. PROPIEDADES BÁSICAS DE LAS PROBABILIDADES	437
10.6. LA LEY DE LA SUMA	439
10.6.1. El evento contrario	441
10.7. PROBABILIDAD CONDICIONAL	442
10.8. LA REGLA DEL PRODUCTO	442
10.8.1. Eventos mutuamente independientes	443
10.9. PROBABILIDAD ESTADÍSTICA O FRECUENCIAL Y PERSONALISTA	445
10.9.1. Limitaciones de la definición clásica	445
10.9.2. Enfoque estadístico o frecuencial	446
10.9.3. El enfoque personalista o subjetivo	447
Ejercicios de autoevaluación	449
Respuesta a los ejercicios de autoevaluación	452

Capítulo 11 LA CURVA NORMAL

Objetivos	456
Resumen	457
11.1. VARIABLE ALEATORIA	457
11.1.1. Variable aleatoria discreta	457
11.1.2. Valores esperados	461
11.2. LA VARIABLE ALEATORIA CONTINUA	463
11.3. CURVA NORMAL	467
11.3.1. Características	468
11.3.2. Importancia	468
11.4. NORMAL ESTÁNDAR	468
11.4.1. Uso de la tabla de la normal estándar	471
11.4.2. Estandarización de una curva normal	478
11.5. APLICACIONES PRÁCTICAS DE LA CURVA NORMAL	479
Ejercicios de autoevaluación	483
Respuesta ejercicios de autoevaluación	485

Capítulo 12 LA INFERENCIA ESTADÍSTICA

Objetivos	492
Resumen	493
12.1. POBLACIÓN, MUESTRA E INFERENCIA ESTADÍSTICA	493
12.2. VALORES DE LA POBLACIÓN Y ESTIMADORES	496
12.3. LOS ESTIMADORES COMO VARIABLES ALEATORIAS	498

12.4. DISTRIBUCIÓN DE LA MEDIA MUESTRAL (\bar{X}).....	500
12.4.1. Teorema del límite central	501
12.5. EL ERROR ESTÁNDAR DEL PROMEDIO	501
12.6. ESTIMACIÓN DE LA MEDIDA μ	502
12.7. LÍMITES DE CONFIANZA PARA μ	503
12.7.1. Interpretación del intervalo de confianza.....	507
12.8. EL TAMAÑO DE LA MUESTRA.....	509
12.9. EL TAMAÑO DE LA MUESTRA EN POBLACIONES FINITAS	512
12.10. FACTORES DETERMINANTES DEL TAMAÑO DE LA MUESTRA.....	514
12.11. ESTIMACIÓN EN EL CASO DE PROPORCIONES.....	515
12.11.1. Intervalo de confianza para p	517
12.11.2. Tamaño de la muestra para estimar p	517
12.11.3 Error estándar de p	519
Ejercicios de autoevaluación.....	523
Respuesta a los ejercicios de autoevaluación.....	526
Capítulo 13 PRUEBA DE HIPÓTESIS	
<i>Objetivos</i>	532
<i>Resumen</i>	533
13.1. UN PROBLEMA DE DECISIÓN	533
13.2. LA REGLA DE DECISIÓN	535
13.3. DEFINICIÓN DE LA ZONA DE RECHAZO Y DE ACEPTACIÓN.....	536
13.4. TIPOS DE ERROR.....	539
13.5. IMPORTANCIA DE LOS TIPOS DE ERROR.....	541
13.6. RESUMEN DE CONCEPTOS BÁSICOS EN RELACIÓN CON LA PRUEBA DE HIPÓTESIS	543
13.7. PRUEBA DE HIPÓTESIS PARA LA MEDIA POBLACIONAL μ	544
13.7.1. Procedimiento clásico para la prueba de hipótesis.....	547
13.7.2. Prueba de hipótesis usando el cálculo directo de la probabilidad	549
13.8. APLICACIÓN DE LA PRUEBA DE HIPÓTESIS	550
13.9. MUESTRAS PEQUEÑAS: la t de <i>Student</i>	553
13.10. LA DIFERENCIA ENTRE DOS MEDIAS.....	559
13.10.1. Comparación cuando las variancias de las poblaciones son conocidas .	559
13.10.2. Comparación cuando las variancias poblacionales no son conocidas y las muestras son pequeñas.....	560
13.11. MUESTRAS DEPENDIENTES O PAREADAS.....	562
13.12. PRUEBA DE HIPÓTESIS PARA PROPORCIONES.....	565
13.13. DIFERENCIA ENTRE DOS PROPORCIONES	566
Ejercicios de autoevaluación.....	569
Respuesta a los ejercicios de autoevaluación.....	575

Capítulo 14 CORRELACIÓN Y REGRESIÓN

Objetivos.....	584
Resumen.....	585
14.1. NIVELES DE MEDICIÓN.....	585
14.2. ASOCIACIÓN ENTRE VARIABLES: LOS CONCEPTOS DE CORRELACIÓN Y REGRESIÓN.....	588
14.3. ASOCIACIÓN ENTRE VARIABLES CUALITATIVAS.....	593
14.4. CORRELACIÓN LINEAL SIMPLE.....	594
14.5. INTERPRETACIÓN DEL COEFICIENTE DE CORRELACIÓN LINEAL SIMPLE (r).....	598
14.6. CORRELACIÓN Y CAUSALIDAD.....	601
14.7. CORRELACIÓN PARCIAL.....	606
14.8. LA REGRESIÓN LINEAL SIMPLE.....	607
14.9. CONFIABILIDAD DEL MODELO DE REGRESIÓN: EL COEFICIENTE DE DETERMINACIÓN.....	612
14.10. UN EJEMPLO ILUSTRATIVO.....	612
14.11. REGRESIÓN MÚLTIPLE.....	614
14.12. INFERENCIA ACERCA DEL COEFICIENTE DE CORRELACIÓN POBLACIONAL ρ	617
Ejercicios de autoevaluación.....	621
Respuesta a los ejercicios de autoevaluación.....	625
REFERENCIAS BIBLIOGRÁFICAS.....	629

ELEMENTOS DE ESTADÍSTICA DESCRIPTIVA

Capítulo 1

NATURALEZA DE LA ESTADÍSTICA

Capítulo 2

FUENTES DE DATOS Y TÉCNICAS DE RECOLECCIÓN

Capítulo 3

EL CUESTIONARIO

Capítulo 4

LOS NÚMEROS RELATIVOS

Capítulo 5

CONSTRUCCIÓN Y ANÁLISIS DE CUADROS ESTADÍSTICOS

Capítulo 6

CONSTRUCCIÓN Y ANÁLISIS DE GRÁFICOS ESTADÍSTICOS

Capítulo 7

DISTRIBUCIONES DE FRECUENCIAS

Capítulo 8

MEDIDAS DE TENDENCIA CENTRAL

Capítulo 9

MEDIDAS DE VARIABILIDAD O DISPERSION

Capítulo 10

INTRODUCCIÓN A LAS PROBABILIDADES

Capítulo 11

LA CURVA NORMAL

Capítulo 12

LA INFERENCIA ESTADÍSTICA

Capítulo 13

PRUEBA DE HIPÓTESIS

Capítulo 14

CORRELACIÓN Y REGRESIÓN

NATURALEZA DE LA ESTADÍSTICA

Sumario

- 1.1. Introducción
- 1.2. Datos estadísticos o estadísticas
- 1.3. La estadística como disciplina científica
- 1.4. Unidad estadística, característica y observación
- 1.5. El concepto de población
- 1.6. Uso del muestreo
- 1.7. La selección de la muestra
- 1.8. La tabla de números al azar
- 1.9. Atributos y variables
- 1.10. Estadística descriptiva y estadística inferencial

Objetivos específicos

Al finalizar el estudio del capítulo, el estudiante será capaz de:

1. Establecer y explicar las diferencias entre estadísticas y estadística.
2. Identificar los conceptos básicos de unidad estadística, característica y observación.
3. Determinar la diferencia entre atributos y variables.
4. Reconocer y definir el concepto de población.
5. Justificar fundamentalmente la aplicación de los distintos tipos de muestreo.
6. Seleccionar muestras utilizando una tabla de números al azar.
7. Determinar las diferencias entre estadística descriptiva y estadística inferencial.

Resumen

En este capítulo se explica en qué consiste la estadística, se expone su ámbito de aplicación y se presentan los conceptos básicos necesarios para iniciarse en su estudio, tales como: unidad estadística, característica, observación, población, atributos y variables. Se introducen los conceptos de estadística descriptiva y estadística inferencial, se describen los procedimientos básicos para la selección de una muestra y se enuncian las razones que justifican el uso de muestras aleatorias o probabilísticas.

1.1. INTRODUCCIÓN

Es muy posible que todo lector de este libro crea conocer lo que es la *estadística* y está en capacidad de dar una definición razonada de ella; es casi seguro, sin embargo, que esta no coincidirá con la de otras personas. Esto sucede porque la palabra *estadística* es utilizada, en el lenguaje corriente en varios sentidos; también porque aun los propios estadísticos, al definir su disciplina, ponen de relieve unos aspectos o aplicaciones más que otros.

1.2. DATOS ESTADÍSTICOS O ESTADÍSTICAS

Conviene señalar, en primer término, que muchas personas asocian la palabra *estadística* con las *publicaciones censales*¹ o las noticias que recogen cifras de producción, de nacimientos, de admitidos en las universidades, de accidentes de tránsito, entre otros, con los *cuadros o gráficos* que aparecen en las revistas o diarios, con las cifras o porcentajes utilizados por los políticos en sus discursos precedidos de la frase ritual: "Las estadísticas muestran que ...". Este sentido corresponde al plural *estadísticas*, el cual se utiliza para indicar un conjunto de datos numéricos que han sido organizados, resumidos y presentados para mostrar las características o evolución de un cierto fenómeno de interés, y es lo que se tiene en mente cuando se habla, por ejemplo, de estadísticas de población, de producción industrial o de un campeonato de fútbol (véase figura 1.1.).

1. Publicaciones en las cuales se recogen los datos proporcionados por el censo.

Campeonato Verano 2009		Estadísticas		Jornada 16				
Clasificación								
Grupo A								
Equipo	PJ	PG	PE	PP	GF	GC	Dif.	Pts
1. Brujas	16	7	5	4	21	14	+7	26
2. Liberia	16	7	5	4	19	16	+3	26
3. Ramonense	16	6	5	5	22	19	+3	23
4. Puntarenas	16	5	6	5	18	21	-3	21
5. Carmelita	16	3	5	8	14	24	-10	14
6. Alajuelense	16	3	4	9	14	22	-8	13
Grupo B								
Equipo	PJ	PG	PE	PP	GFGC	Dif.	Pts	
1. Saprissa	16	10	3	3	26	13	+13	33
2. Herediano	16	9	3	4	22	15	+7	30
3. Pérez Z.	16	6	5	5	19	17	+2	23
4. Cartaginés	16	5	8	3	13	11	+2	23
5. San Carlos	16	4	6	6	18	17	+1	18
6. Universidad	16	3	1	12	11	28	-17	10
Resultados								
Saprissa	0	Cartaginés	0					
San Carlos	3	Herediano	0					
Pérez Zeledón	4	Universidad	0					
Brujas	4	Puntarenas	0					
Alajuelense	0	Ramonense	2					
Liberia Mía	2	Carmelita	0					
Próxima fecha								
Cuartos de final								
Miércoles 6 de mayo								
Ramonense-Herediano, Est. Vargas Roldan								
Pérez Zeledón-Liberia Mía, Est. M. Pérez Z.								
Goleadores								
Mario Camacho (Puntarenas)	10							
Alejandro Sequeira (Ramonense)	9							
Andy Herron (Herediano)	8 (4 p)							
Johnny Woodly (San Carlos)	6 (1p)							
Victor Núñez (Liberia)	5 (1p)							

Figura. 1.1. Ejemplo de unas estadísticas del campeonato de fútbol de Costa Rica del verano del 2009

Fuente: Periódico *La República* del 3 de mayo del 2009.

Así, cuando el Oficial Mayor del Ministerio de Educación Pública pide las estadísticas de deserción escolar, seguro piensa en algún cuadro que presente el número –y el porcentaje– de desertores en el último año académico, clasificados por área geográfica, edad, nivel escolar alcanzado, así como algún gráfico que muestre la evolución del fenómeno en los últimos cinco o diez años. Igualmente, si el gerente de una empresa dice que las estadísticas señalan que se ha tenido un buen año, es muy posible que se refiera a las ventas, a la producción y a las ganancias del año fiscal, quizás detalladas por mes y por regiones del país y comparadas con las del año precedente.

Cuando se usa en este sentido, la palabra *estadística* equivale a información numérica o cuantitativa sobre un tema, para un cierto período y se organiza de tal forma que muestra los aspectos más significativos y de mayor interés. Obviamente, un término más apropiado sería el de *datos estadísticos*, pero la realidad es que el uso de la expresión *estadísticas*, para referirse a este tipo de información, es muy común aun entre los propios estadísticos.

La información, para ser analizada estadísticamente, debe ser cuantitativa o susceptible de ser expresada en esta forma. Sin embargo, no todo este tipo de información constituye un dato estadístico; un ejemplo de datos estadísticos lo constituye un conjunto de números referidos a una misma *característica*, y se recogen de tal modo que se puedan comparar, analizar o interpretar.

Un número aislado, si no se compara o muestra una relación significativa con otros, no es un dato estadístico. La estadística se interesa por los patrones o las regularidades que presenta un conjunto de datos y trata, entonces de determinar las propiedades de esos conjuntos. El detalle individual prácticamente carece de interés.

Debido a lo anterior, por ejemplo, los datos de un paciente que se recogen en un hospital, no obstante lo importante que resultan para el médico (¡y para el paciente!), no son en realidad una estadística; pero sí lo son los de un grupo de pacientes afectados por la misma dolencia, ya que a partir del análisis estadístico de un conjunto de datos se deriva una serie de medidas y de relaciones que permitan tipificar las características de la dolencia, entenderla y eventualmente contribuir a la curación de las personas afectadas. De manera más específica, no es un dato estadístico la cifra aislada que mide la temperatura de un paciente, pero sí la temperatura promedio de un grupo de ellos.

De igual forma, no son datos estadísticos el monto, plazo y tasa de interés de un determinado préstamo concedido por un banco, pero sí la información basada en un conjunto de ellos recogidos en un cierto período. Así, por ejemplo, si se consideran los créditos, estos pueden ser analizados y comparados a fin de determinar cuál es el monto de préstamo más corriente, dentro de qué límites varía la gran mayoría de créditos, entre otros. Además, si fuera de utilidad, podrían relacionarse las cantidades con los plazos y con las tasas de interés, a fin de determinar si existe algún tipo de asociación entre esas características.

1.3. LA ESTADÍSTICA COMO DISCIPLINA CIENTÍFICA

Ahora bien, la estadística no es simplemente un conjunto de datos estadísticos, ni está interesada solamente en recogerlos, resumirlos y presentarlos. Cuando se hable en este libro de "estadística", se tendrá en mente un campo del conocimiento, una disciplina

científica dedicada al desarrollo y aplicación de la teoría y las técnicas apropiadas para la recolección, clasificación, presentación, análisis e interpretación de información cuantitativa obtenida por observación o experimentación.

Como ciencia, la estadística se concibe, a veces como una rama de las Matemáticas Aplicadas, pero también incluye elementos teóricos y técnicas propias. En contraste con otros científicos, cuyos datos vienen de su propia disciplina, los estadísticos frecuentemente analizan datos de diferentes disciplinas.

Donde quiera que se necesite recoger datos por observación, entrevista, experimentación o registro y surja el interés de resumirlos, analizarlos y sacar conclusiones a partir de ellos, aparece la estadística. Por ello, no debe sorprender que se haya convertido en una herramienta esencial de la investigación en casi todos los campos de la actividad humana.

1.4. UNIDAD ESTADÍSTICA, CARACTERÍSTICA Y OBSERVACIÓN

El análisis estadístico se lleva a cabo con base en *observaciones o valores observados* correspondientes a una cierta característica y realizados en lo que se denomina unidades estadísticas elementales o de estudio. Unos ejemplos contribuirán a aclarar estos conceptos.

Ejemplo 1

Una oficina gubernamental desea estimar, para una cierta zona, el salario semanal promedio de los trabajadores industriales. Para ello decide llevar a cabo una *encuesta*. En este caso, la unidad estadística elemental o de estudio la constituye el trabajador industrial; la característica que interesa es el salario semanal en colones. Cada empleado presenta un valor observado u observación, constituido por la suma recibida como pago durante la semana, por ejemplo, 43 500 colones. El conjunto de esos sueldos semanales recogidos en la encuesta, uno por cada unidad de estudio, constituyen las observaciones, o sea el grupo de datos en el cual se basará el análisis estadístico de lo devengado por el trabajador industrial y el cálculo de la remuneración promedio semanal.

Es importante destacar que para llevar a cabo el estudio anterior y llegar a resultados útiles, libres de ambigüedades, es esencial definir claramente qué se entiende por trabajador industrial, a fin de que no haya dudas respecto a quiénes deben ser incluidos en la encuesta. Igualmente, es necesario precisar el salario semanal, estableciendo, con todo detalle si se trata del neto o del bruto, si se consideran las horas extras o no, si se refiere a tiempo completo o no, entre otros. Si esto no se hace, los resultados serán difíciles de interpretar y poco confiables.

Ejemplo 2

El jefe de la sección de créditos personales de un banco decide hacer una investigación para saber el estado de los préstamos, en lo que respecta a la puntualidad en el pago de las amortizaciones mensuales. En este ejemplo, la unidad estadística es el préstamo y la característica que interesa es el cumplimiento en el pago. Las observaciones podrían arrojar resultados de tres tipos: "al día", "atrasado" y "al cobro judicial"; y así, para cada préstamo, existirá una observación que se situaría dentro de una de esas categorías.



Ejemplo 3

El ingeniero jefe de una fábrica desea verificar la calidad de una partida de bombillos producidos el día anterior. Para hacerlo, toma una *muestra* de 50 de ellos y determina, para cada uno, si enciende o no. En este caso, la *unidad de estudio* es el bombillo; la *característica* de interés, su estado; y la *observación*, para cada bombillo en la muestra es, su buen o mal estado, es decir, si enciende o no. Las observaciones las conformarán los 50 resultados.



En los ejemplos anteriores se tenía interés en una sola característica de la unidad de estudio; sin embargo, no debe creerse que esta sea siempre la situación. Hay muchas oportunidades en las cuales, durante una investigación, se llevan a cabo observaciones de varias características diferentes en una misma unidad; luego, estas se analizan en forma separada o simultáneamente, según lo que interese. Considérese el ejemplo 4 para aclarar la idea.

Ejemplo 4

El director de un colegio rural desea analizar los factores asociados al rendimiento escolar de los alumnos del primer año de la institución que dirige; para ello, obtiene de cada uno de los estudiantes, mediante la aplicación de un cuestionario, la siguiente información: nota promedio correspondiente al primer trimestre, sexo, ocupación del padre, madre o encargado, escolaridad del padre y de la madre, ingreso mensual de la familia, número de hermanos y recibo de periódicos en la vivienda. Además, les pasa un *test* de conocimientos generales.



En este caso, la *unidad de estudio* es evidentemente el alumno de primer año y las *características* de interés las ya señaladas: nota promedio del primer semestre, sexo, ocupación del padre, de la madre o del encargado, entre otros. Un ejemplo concreto de observación sería el que aparece en el cuadro 1.1.

Cuadro 1.1

EJEMPLO DE OBSERVACIONES RECOGIDAS EN UN ALUMNO DE PRIMER AÑO

UNIDAD DE ESTUDIO	CARACTERÍSTICA	OBSERVACIÓN
Guillermo Rodríguez V.	Nota promedio primer semestre	8,25
	Sexo	Masculino
	Ocupación del padre	Jornalero
	Ocupación de la madre	Ama de casa
	Escolaridad del padre	6° grado
	Escolaridad de la madre	4° grado
	Ingreso mensual familiar	180 mil colones
	Número de hermanos	4
	Reciben periódicos en la casa	No
	Nota de conocimientos generales	83

Una vez recogida la información, el director puede analizar las notas, la ocupación del principal sostén de la familia, el ingreso mensual del hogar, la educación de la madre, si toma cada una de esas características por separado y considera el conjunto de observaciones correspondientes a todos los alumnos. Así, por ejemplo, tiene la posibilidad de escoger las respuestas relacionadas a la educación de la madre (número de años de estudio) y organizarlas en un cuadro, para llegar posteriormente a algunas conclusiones acerca de dicha característica para el grupo estudiado.

Igualmente, puede tomar las cifras de ingreso mensual de la familia recogidas en la investigación y analizarlas para determinar cuál es el ingreso mensual promedio, dentro de qué límites varía la gran mayoría de las entradas mensuales familiares, qué porcentaje tiene un ingreso mayor de 500 000 colones mensuales, etc. El director, si lo desea, es posible que examine el grado de asociación entre dos características: ingreso mensual familiar y número de hermanos, por ejemplo; para ello, debe tomar, para cada alumno, el valor observado del ingreso mensual familiar y el del número de hermanos, y examinar simultáneamente esas parejas de valores para los estudiantes investigados, a fin de determinar si existe algún tipo de relación entre ellas, así como la intensidad de esa relación. Del mismo modo, puede estudiar la conexión que existe entre la nota promedio y el número de años de estudio de la madre.

1.5. EL CONCEPTO DE POBLACIÓN

Toda investigación tiene como referencia un *conjunto de unidades de estudio* o elementos que pueden ser personas, animales, empresas, organizaciones, objetos, etc. Este conjunto

se denomina técnicamente, en estadística, *población del estudio* o simplemente *población*, definida como *el total o agregado de las unidades de estudio*.

Con la investigación se pretende conocer las características del conjunto y generalizar los resultados o conclusiones que se obtengan de la muestra a la población.

Por ejemplo, en una investigación dirigida a conocer el ingreso mensual de las familias del Área Metropolitana de San José, la unidad de estudio es la familia residente en esa zona geográfica, y la población, evidentemente, la constituyen el conjunto de familias que viven en esta zona, en el momento del estudio.² En forma similar, si un profesor de matemáticas está interesado en analizar el rendimiento de sus alumnos de primer año de secundaria, en la materia que imparte, y tiene dos grupos: uno de 26 estudiantes y otro de 28, la población objeto de su estudio estará formada por los 54 alumnos. Por otra parte, si un agrónomo zootecnista aplica una nueva dieta a una muestra de 20 vacas Holstein para determinar su efecto sobre la producción de leche, la unidad de estudio es la vaca Holstein y la población vendrá dada por todas las reses de esa raza a las que presumiblemente se podría aplicar la dieta.

Una *población* puede ser finita si tiene un número limitado de elementos o infinita si la forma un número ilimitado. Las familias del Área Metropolitana de San José, en una cierta fecha, son un ejemplo de una población grande pero finita, y las personas residentes en la Ciudad de México, en un cierto momento, forman una población muy grande, pero necesariamente finita. Igualmente los 54 alumnos del profesor de matemáticas al que se hizo referencia constituyen un claro ejemplo de una población finita; sin embargo, el número de alumnos actual, pasado y futuro de matemáticas de primer año de secundaria es ilimitado y por ello puede decirse que son una *población infinita*. Igualmente es infinita la compuesta por las vacas Holstein que se tienen en el caso del experimento del agrónomo zootecnista. Es importante señalar que, si se va a estudiar la situación existente en un momento dado, la población quizás sea muy grande, pero necesariamente es finita; en cambio, si se analiza un proceso o experimento, el cual teóricamente puede repetirse indefinidamente bajo las mismas condiciones –caso de la aplicación de la dieta de las vacas Holstein– es infinita. En el caso del ejemplo del agrónomo zootecnista, el total de vacas Holstein que existen en el mundo en un mes o año dado es finito pero desconocido.

El concepto de *población* es muy importante en estadística y es conveniente aclarar algunas confusiones que podrían presentarse al respecto. En primer término, *la estadística realiza sus análisis no con las unidades propiamente dichas, sino con los valores observados para*

2. Se requiere, por supuesto, una definición clara de los distritos o zonas que comprende el Área Metropolitana de San José, al igual que una definición apropiada de familia que pueda ser aplicada sin ambigüedades en la práctica, al realizarse el estudio.

ciertas características de ellas. Por eso, en una encuesta dirigida a conocer el ingreso mensual familiar del Área Metropolitana de San José, la *unidad de estudio* será la familia residente en esa zona geográfica, y la población la formarán el total de las que viven dentro del sector. Para el estadístico, el conjunto de los valores de una característica de la población, por ejemplo el ingreso mensual, constituye una población. El análisis estadístico se basará en el conjunto de ingresos mensuales obtenidos de las familias participantes en el estudio, y de ellos se sacará las conclusiones que luego se extenderán a todas las familias que residan en el Área Metropolitana.

Desde esta perspectiva, la *población* la constituyen no las *unidades estadísticas* propiamente, sino los valores numéricos asociados a ellas. En la práctica, sin embargo, esta diferenciación no es fundamental y la población puede verse tanto como un conjunto de números o de unidades; por ello, en el ejemplo 1 de este capítulo, donde la *unidad de estudio* es el trabajador industrial, la *población* puede ser concebida y descrita como formada por todos los trabajadores industriales de la zona durante la semana anterior a la encuesta.

Otro aspecto importante es que *un mismo conjunto de unidades de estudio puede dar origen a diferentes poblaciones, o sea, a distintos grupos numéricos, según sea la característica que concierna.* Así, por ejemplo, si la *unidad estadística* es el trabajador industrial, esta puede originar diferentes poblaciones, según sean los intereses de la investigación: habrá una población de salarios semanales, una de pesos en kilogramos de los trabajadores, una de estaturas, de número de hijos de esos trabajadores; una de estados civiles, una de cantones de nacimiento, entre otros.

A continuación, se presentan algunos ejemplos de poblaciones y se discute su condición de finitas o infinitas.

- a) Peso de los estudiantes matriculados en la Universidad Estatal a Distancia (UNED) en el presente año.
- b) Número de hijos vivos de cada una las mujeres de 50 años y más residentes en el Área Metropolitana de San José en una cierta fecha (20 de abril del 2009, por ejemplo).
- c) Diámetro de las bolas de tenis producidas por una máquina.
- d) Resultados que se obtendrían al repetir un experimento agrícola bajo las mismas condiciones.
- e) Resultados obtenidos al lanzar un dado al aire 200 veces.
- f) Resultados obtenidos al lanzar una moneda al aire un número infinito de veces.
- g) Respuestas de los empleados de una institución a la pregunta “¿Qué opinión le merece el gerente: buena, regular o mala?”

- h) Sumas depositadas en cada una de las cuentas de ahorro del Banco Nacional, el 30 de setiembre del año pasado.
- i) Número de pacientes nuevos atendidos por una clínica privada en el curso del año pasado.

Los ejemplos c), d) y f) representan poblaciones infinitas. En el caso c), sin embargo, si se toman los diámetros de las bolas de tenis producidas en un cierto lapso, un día por ejemplo, la población es finita. El ejemplo e) constituye una población finita, pero si el proceso no se detiene se estaría en una situación similar a la f), es decir, se tendría una población infinita.

1.6. USO DEL MUESTREO

Los estudios, investigaciones y experimentos se llevan a cabo, normalmente, con el propósito de llegar a conclusiones o resultados aplicables a todos los elementos de la población que constituye el campo de referencia de interés. Así, en el ejemplo antes citado, sobre el ingreso mensual de las familias del Área Metropolitana de San José, los investigadores desean obtener resultados válidos para todas las familias de la zona. Igualmente, el agrónomo zootecnista que realiza su experimento con una nueva dieta desea que sus resultados sean aplicables a todas las vacas Holstein y, si es posible, a otras vacas lecheras. Para el logro de este propósito de generalidad en los resultados de las investigaciones, el sentido común sugiere la necesidad de proceder al análisis completo de la población. Lo cual se hace, precisamente, cuando esta es pequeña, pues el obtener información de todos los elementos que la componen es prácticamente posible y no representa una inversión de tiempo y de recursos desproporcionados o irrazonables. También es la práctica seguida en ciertas operaciones estadísticas especiales, como los *censos* de población y de vivienda.³

En muchas ocasiones, sin embargo, el estudio de todos los elementos que forman la población no solo no es práctico, por los costos y el tiempo que suponen, sino que incluso puede resultar imposible de llevar a cabo porque la población es muy grande o infinita. Por ello se toma una parte, *una muestra*, estudiarla y luego generalizar los resultados observados a toda la población de la cual fue seleccionada. Lo mismo se aplica cuando las unidades se destruyen o transforman al ser estudiadas, como sería el caso de un cargamento de semillas que, para conocer su estado, deben ponerse a germinar.

3. Recuento exhaustivo de la población y viviendas de un país que se hace cada cierto tiempo –usualmente diez años– y en el que se recogen datos demográficos, económicos y sociales referentes a todos los habitantes en un momento determinado, así como sobre las características de las viviendas donde ellos residen.

El uso del *muestreo* es un medio más rápido y barato y, en ciertos casos, el único posible. Sin embargo, por tratarse de un *procedimiento inductivo*, de una *inferencia* que se hace de la parte hacia el todo, se corre el riesgo de que la parte seleccionada no sea *representativa*⁴ del todo y, por lo tanto, que la generalización no sea precisa y tenga cierto margen de error, el cual, asociado a las inferencias realizadas a partir de muestras, es el precio por los beneficios que en costo y rapidez brinda el *muestreo*. En el caso del estudio sobre ingreso familiar en el Área Metropolitana de San José, por ejemplo, si la muestra de familias no es representativa y quedan en ella muchas de nivel socioeconómico medio o alto de las que corresponden proporcionalmente, el ingreso mensual promedio arrojado por la muestra será superior al real y, en consecuencia, se concluirá que el ingreso familiar promedio de la zona es mayor de lo que es en realidad. Entonces, al realizar inferencias, un punto crucial es lograr que la muestra utilizada sea representativa de la población de la cual se toma. En esto influyen dos factores básicos: el tamaño de la muestra y su forma de selección.

En general, cuanto más grande sea el tamaño de la muestra, mayor confianza se tiene que es representativa de la población. Conviene señalar, sin embargo, que el tamaño de la muestra por utilizar en un estudio depende de la homogeneidad mostrada por los elementos de la población, del nivel de precisión deseado para las estimaciones que se derivarán de la muestra y del grado de confianza requerido en la inferencia. Así, cuanto más homogénea sea una población, menor es la muestra para obtener conclusiones de un cierto grado de confianza.

Para un examen de sangre, por ejemplo, se utiliza una cantidad muy pequeña de la que existe en el cuerpo humano, pues elementos que entran en su composición están bien mezclados y, por lo tanto, bien distribuidos dentro del conjunto. Con ese nivel de homogeneidad, las conclusiones obtenidas a partir de unas pocas gotas son válidas para toda la sangre del cuerpo.

Si se quisiera estudiar el ingreso familiar en el Área Metropolitana de San José, evidentemente no sería aconsejable una muestra tan reducida en relación con el total, como la mencionada para el examen de sangre, ya que el ingreso es mucho más heterogéneo –varía más– entre familias y, en consecuencia se requiere una muestra mayor.

Ahora bien, que la muestra sea de un tamaño razonable o aun grande no garantiza que sea representativa de la población. Hay muchos ejemplos de estudios basados en muestras muy grandes, los cuales han producido resultados errados porque, a pesar de su

4. Se dice que una muestra es representativa cuando su composición, de acuerdo con las variables pertinentes, es muy similar –para fines prácticos– a la composición de la población de la cual fue seleccionada. El ideal se logra cuando la muestra es, en realidad, una miniatura de la población. Obviamente, cuanto más representativa de la población es la muestra, más confiables serán las inferencias que se realicen a partir de ella.

tamaño, tenían fallas de selección –sesgos– que las hacían no representativas. Esto lleva a la necesidad de discutir con cierto detalle el tema de los métodos de selección la razón por la que los estadísticos recomiendan y defienden la selección aleatoria.

CUÁNDO USAR MUESTRAS

- La población es infinita o muy grande y es imposible física o económicamente cubrir todos los elementos que la componen.
- La población es finita, pero lo suficientemente grande para que el estudio de todos los elementos no sea aconsejable, porque resultaría demasiado costoso y tomaría tanto tiempo que los datos serían obsoletos o inútiles cuando estén disponibles.
- La unidad de estudio se transforma o se destruye al ser examinada.
- Los resultados que arrojaría una muestra bien seleccionada, de tamaño razonable, serían suficientemente precisos para los fines prácticos perseguidos con los datos.

1.7. SELECCIÓN DE MUESTRA

En la práctica, pueden distinguirse las siguientes tres formas de selección de muestras:

- a) *Aleatoria o al azar*, es decir, dándole a cada uno de los elementos de la población una probabilidad conocida y no nula de incluirse en la muestra. Un caso particular es en el que a todos los elementos se les da la misma probabilidad; este procedimiento recibe el nombre de muestreo simple al azar o muestreo igualmente probable.
- b) *Intencional*, o sea utilizar el juicio de una persona con experiencia y conocimiento con respecto a la población que se estudia para la selección de los elementos en la muestra.
- c) *Por conveniencia*, es decir, escoger las unidades o elementos que están disponibles o son más fáciles de conseguir.

En el caso c), evidentemente se corre el riesgo de que la muestra no sea representativa, ya que las unidades estadísticas disponibles o fáciles de conseguir podrían ser de naturaleza diferente a las demás.

Considere el siguiente ejemplo: un empresario residente en Los Pinares⁵ quiere estimar la proporción de viviendas del Área Metropolitana de San José que disponen de

5. La urbanización Los Pinares es un barrio del Área Metropolitana de San José, ubicado en el distrito Sánchez de Curridabat, donde predominan las familias con ingresos altos.

conexión a internet en la vivienda. Para ello, decide visitar una muestra de 400 casas de ese barrio y determinar si tienen o no ese servicio instalado, y luego generalizar el resultado a todas las viviendas del Área Metropolitana.

Tan solo con una idea general de las características socioeconómicas de los barrios del Área Metropolitana de San José, se concluye que una muestra de casas del barrio Los Pinares no será representativa del Área Metropolitana de San José y concluirán a una sobrestimación importante de la proporción de viviendas que disponen de conexión a internet. Cualquier persona razonablemente informada objetaría esta muestra.

Si se utiliza el juicio de un experto, es decir, una persona con mucho conocimiento de las características de los barrios del Área Metropolitana, para seleccionar las viviendas de la muestra de acuerdo con su conocimiento de la población, es probable que la muestra resultante *será representativa*, pues el experto empleará todos sus conocimientos tratando de que así sea.

Sin embargo, la experiencia indica que las personas tienen tendencias, muchas de ellas inconscientes; por lo tanto, cuando seleccionan una muestra, le dan preferencia a ciertos elementos en perjuicio de otros, por lo que no resulta representativa. En otras palabras, el juicio personal puede tender sistemáticamente a dar más oportunidad de incluir ciertos elementos de la población.

En el ejemplo comentando, la selección hecha por una persona conocedora de la situación económica de las familias que viven en los diferentes sectores del Área Metropolitana es muy superior a una selección de viviendas hecha completamente al azar de esa zona, pero esto no garantiza que represente con exactitud la situación de la población, las tendencias de parte del experto permiten que ciertos sectores queden representados en la muestra con mayor importancia de la que realmente tienen. Por ejemplo, si el especialista tiene la tendencia a escoger viviendas situadas en las esquinas en mayor proporción que las ubicadas en otras partes de la cuadra y, precisamente por esto, aquellas sean más caras, estén en lotes más grandes y correspondan a familias de un mayor nivel de ingreso. Si la tenencia de conexión a internet está asociada con el nivel económico, esta tendencia del especialista puede resultar en una sobrestimación de tenencia de la proporción de hogares que cuentan con el servicio de internet.

Supóngase, finalmente, que se utiliza un procedimiento aleatorio para seleccionar la muestra. Para ese fin, se toma un mapa del Área Metropolitana de San José y se divide en segmentos de alrededor de 50 viviendas cada uno, estos se numeran consecutivamente de 1 en adelante. Luego con algún sistema aleatorio, como el utilizado en la lotería, se escogen 40 de ellos; después dentro de cada uno de los segmentos seleccionados se eligen, también al azar, una de cada 5 casas y se visitan para hacer el estudio.

¿Será esta muestra representativa de las viviendas del Área Metropolitana de San José?

No necesariamente. Por azar, la muestra podría contener un porcentaje de residencias con conexión a internet mayor del que realmente hay en la población, o que ese porcentaje sea menor. Es decir, la muestra al azar puede no ser representativa, sub o sobrestimando la presencia en la población de la circunstancia que interesa.

Ahora bien, si la selección aleatoria puede llevar y lleva a muestras no representativas, ¿por qué razón, entonces, los estadísticos utilizan muestras al azar en lugar de intencionales o por juicio?

La respuesta a esta pregunta es clara si se considera la diferencia entre los tipos de errores a los que lleva cada una de las formas de selección: al azar produce discrepancias entre el resultado obtenido de la muestra y el valor de la población, cuyas magnitudes y sentido no pueden predecirse para una específica porque se trata de errores aleatorios o de muestreo, las cuales se originan exclusivamente en el hecho de que se trabaja con una muestra de la población y esta es escogida al azar. La selección intencional o de juicio, en cambio, produce básicamente discrepancias o errores en un solo sentido, es decir los denominados *sesgos de selección*.

Esta distinción es de fundamental importancia y hay que detenerse en ella con cierto detalle.

En general, los *sesgos* son errores sistemáticos, o sea, en un solo sentido, y pueden ser de selección o de medición. Los de selección se presentan únicamente cuando se utilizan muestras, mientras que los de medición son comunes a censos y a muestras. Un ejemplo de este último tipo de *sesgo* se da en las cifras sobre ingresos, las cuales, como es sabido, usualmente no indican el ingreso real, debido a la tendencia de las personas a declarar entradas menores a las que tienen, así como por la propensión a excluir cierto tipo de ingresos provenientes de alquileres o intereses, o a declarar solo el neto en vez del bruto.

Si se conoce su sentido y magnitud, el *sesgo* puede corregirse fácilmente: tal sería el caso de una persona que mida con un metro al cual se le han quitado 2 centímetros. Al medir siempre estaría exagerando, pero en cuanto se descubra esta falla es posible corregir las mediciones ya realizadas. En la práctica, sin embargo, lo corriente es que las tendencias de las personas para reportar el ingreso y, en general, los *sesgos*, se desconozcan por ello resulta imposible evaluarlos y corregirlos.

Los errores de muestreo, por su naturaleza aleatoria, no son predecibles individualmente, pero sí pueden estudiarse y evaluarse, por esa misma razón, en muestras sucesivas, utilizando los modelos matemáticos que brinda la teoría de las probabilidades.

De vuelta al ejemplo del empresario interesado en la proporción de viviendas del Área Metropolitana que disponen de conexión a internet, se señalan las características de ambos tipos de procedimiento para la selección de la muestra:

- a) Si un experto escoge un número grande de muestras de casas y su tendencia es a seleccionar las ubicadas en las esquinas, todas sus muestras sobrestimarán el porcentaje de viviendas con internet en el Área Metropolitana.
- b) Si un número grande de muestras se selecciona aleatoriamente, una parte de ellas sobrestimará la proporción de familias con internet, otra la subestimará y otro grupo estará muy cerca del verdadero valor de la proporción en la población. La falta de representatividad debida al azar, es decir, *los errores del muestreo*, se alternarán en muestras sucesivas, por lo que en ocasiones son positivos y, a veces, negativos y tienden a compensarse en un número grande de ellas. Debido a este comportamiento característico, *los errores aleatorios pueden ser descritos con un modelo de probabilidades y utilizados para evaluar la confianza que merece el resultado obtenido en una sola muestra.*

En resumen, los *sesgos* de selección no pueden ser evaluados o medidos con un modelo; los *errores de muestreo*, por su naturaleza aleatoria, sí. Cuando se toma una muestra intencional o de juicio, no se sabe el efecto del *sesgo*; en cambio, si se toma al azar, el posible efecto del error de muestreo sí es medible tomando en cuenta el comportamiento que se esperaría en muestras sucesivas de este tipo.

Las ventajas de usar muestreo aleatorio surgen claramente después de la discusión anterior y permiten responder a la pregunta de por qué se prefiere el muestreo aleatorio o probabilístico al de juicio; la predilección se debe, fundamentalmente, a que el aleatorio elimina los *sesgos de selección* y produce *errores aleatorios*, los cuales, por su naturaleza, se pueden medir utilizando modelos probabilísticos.

De esta manera, la *validez* o *precisión* de una inferencia realizada a partir de una muestra aleatoria es medible, mientras que no lo es cuando está basada en una intencional o de juicio. Adicionalmente, el uso del muestreo aleatorio permite, al aumentar el tamaño de la muestra, reducir el error probable de la inferencia a una magnitud tan pequeña como se desee.

Es importante señalar, por último, que cuando se utiliza una muestra muy pequeña, la selección intencional resulta corrientemente más apropiada que la aleatoria. Así, por ejemplo, se desea hacer un estudio antropológico cuidadoso del estudiante universitario costarricense. Debido a la naturaleza de la investigación, que requiere un largo período de observación del alumno, conocer sus hábitos, vivir en su casa durante un tiempo, entre otros, el antropólogo a cargo de este trabajo solo podrá analizar eficazmente un número muy restringido de casos (alrededor de cinco). Por ello, no sería apropiado utilizar un procedimiento aleatorio para escoger los sujetos; la muestra, por el

contrario, debe ser intencional, y realiza la selección de manera que esos cinco alumnos reflejen adecuadamente las características típicas del estudiante costarricense. Escoger cinco discentes al azar implicaría correr el gran riesgo de que el grupo seleccionado no reúna las características típicas o del promedio.

LAS MUESTRAS ALEATORIAS SE PREFIEREN PORQUE

- Eliminan los sesgos de selección.
- Producen errores aleatorios que son medibles utilizando modelos probabilísticos.
- El error de muestreo puede hacerse tan pequeño como se quiera con el aumento del tamaño de la muestra.

1.8. LA TABLA DE NÚMEROS AL AZAR

Ahora, conviene prestar atención a la forma de llevar a cabo la selección aleatoria de una muestra. En realidad, se busca un método que garantice una probabilidad determinada, a cada elemento de la población, de ser incluido en esta. Considere, para facilitar las explicaciones, algunos de estos.

Ejemplo 5

Seis amigos se reúnen a comer. Una vez finalizada la comida, deciden echar a suerte quién debe pagar la cuenta. La rifa podría realizarse en varias formas, se señalan dos:

- a) Se toman seis pedazos de papel de igual tamaño y color, en cada uno de ellos se escribe el nombre de uno de los amigos y luego se enrollan en forma de cigarrillos. Estos "papeliitos" se revuelven bien dentro de la canastilla del pan y luego se extrae uno, el nombre que aparece en él indicará quién debe pagar la cuenta.
- b) Se hace una lista de los amigos y luego se numeran los nombres de 1 a 6. Se tira un dado corriente y el número que quede hacia arriba indicará el que debe cubrir la cuenta.⁶



6. Una opción muy frecuente consiste en asignar un número a cada persona y llamar luego al mesero y pedirle que diga un número entre 1 y 6. Aunque es una forma de resolver el problema, se trata de un procedimiento sesgado, pues las personas rara vez escogen los valores extremos 1 o 6 y tienden, más bien, a mencionar con mayor frecuencia valores centrales.

Ejemplo 6

Se desea escoger, de un grupo de 60, a 5 alumnos para que los representen en un acto social. El profesor quiere que todos participen en la elección. Para ello, se puede proceder en varias formas:

- a) Se utiliza el método de los "cigarrillos" ya mencionado, y se seleccionan 5 de la caja o recipiente donde se mezclaron.
- b) Asignar los números entre 01 y 60, uno a cada alumno; se consiguen 60 fichas de lotería numeradas de 01 a 60, se colocan en una bolsa, se revuelven bien y luego 5 de ellas son escogidas al azar.



En los ejemplos que se presentaron, los procedimientos de selección de la muestra mencionados ("cigarrillos", fichas de lotería, dados, entre otros) son cómodos y rápidos y, en términos generales, garantizan un suficiente grado de aleatoriedad. Sin embargo, cuando la población es de mayor tamaño, estas maneras pueden resultar laboriosas y complicadas; además, no existe completa seguridad de que sean realmente aleatorias. Por ejemplo, el conocido método de los papelitos o "cigarrillos" puede fallar, ya sea porque no son todos de igual tamaño o no quedan adecuadamente mezclados.

Debido a esto, los estadísticos prefieren utilizar, para la selección de las muestras, procedimientos que garanticen una completa aleatoriedad y sean simples de aplicar: uno de ellos, con larga tradición en la estadística, es la *tabla de número al azar* y el otro son los métodos computarizados que permiten generar muestras de números aleatorios.

En una *tabla de números al azar* aparece un gran número de dígitos, los cuales han sido generados por un proceso mecánico o electrónico, usando una fórmula o algoritmo matemático, y garantiza que cada dígito es el resultado de un procedimiento aleatorio, es decir, independiente de los dígitos previamente seleccionados. Desplazarse por una columna, una hilera, una diagonal o en alguna forma sistemática por la tabla de números al azar equivale a elegir dígitos por un proceso aleatorio.

En una *tabla de números al azar* la proporción de los que tienen una cifra: 0, 1, 2 ... 9, es muy similar, por ello la probabilidad de que el escogido, sea por ejemplo 3, es de $1/10$, y también es de un décimo la de obtener cualquier otro dígito. En igual forma la proporción de números de dos cifras: 00, 01, ... 99, es muy similar y la probabilidad de seleccionar cualquiera número de dos cifras es de $1/100$.

Tradicionalmente se han preparado extensas tablas de números al azar que incluyen hasta un millón de números, como la que se observa en la figura 1.2.

Tabla de números al azar										
	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
00	54463	22662	62305	70639	79365	67382	29085	69831	42058	08185
01	15389	85205	18850	39226	42294	90669	96325	23248	60933	26927
02	85941	40756	82414	02015	13858	78030	16269	65978	01285	15345
03	61149	69440	11286	88218	58925	03638	52862	62733	33451	77455
04	05219	81619	10651	57079	92511	59888	84502	72095	83463	75577
05	41417	98326	87719	92294	46614	50948	64886	20002	97365	30975
06	28357	94070	20652	35774	15249	79019	21145	05217	42286	76305
07	17783	00015	10076	83091	91530	36466	39981	52481	49117	75229
08	40950	84820	29881	85966	62800	20326	84740	62660	77379	90229
09	82995	64157	66164	41180	10089	41252	78258	96488	88629	12231
10	96754	12676	55669	44105	42361	34833	86629	23930	53249	27083
11	34357	88040	53364	21726	45690	56314	80332	22554	90600	61113
12	06318	37403	49927	57715	50423	67372	63306	48888	21505	80182
13	62111	52820	07243	29931	89292	84267	85693	23947	22278	10551
14	47534	09243	57869	00544	23410	02740	02540	54440	32949	13490
15	98514	75993	84460	62846	59844	14522	48730	23443	48157	34770
16	24856	03548	44898	09351	98795	18644	39265	21058	90368	44104
17	96887	12479	80621	66223	86085	28285	02432	53342	42846	94771
18	90801	21472	42815	77408	37390	76786	57615	32141	30268	18105
19	55165	77310	83666	36028	28420	20219	81369	41943	42066	41067
20	25884	12932	84318	95108	72305	64620	91308	89032	45325	85436
21	16777	37116	58550	42958	21460	43910	01125	82894	81389	10620
22	46230	43877	80207	88877	89380	32992	91380	03164	98656	58337
23	42902	66892	46134	01432	94210	23474	20423	60137	60609	10119
24	81007	00333	39693	28039	10154	95425	39220	19774	31782	49032
25	68089	01122	51311	72370	06002	74373	96199	92013	41273	21546
26	20411	67081	89950	16944	93054	87687	96593	87235	77054	33846
27	58212	13160	06458	15718	82623	76989	05999	58680	96739	63700
28	70577	42866	24969	61210	76046	67699	42054	12690	93758	03283
29	94522	74358	71659	62038	79643	79169	44741	05437	39038	13163
30	42626	86819	85651	88678	17401	03252	99547	32404	17918	62880
31	16051	33763	57194	16752	54450	19031	58680	47629	54132	60631
32	08244	27647	33851	44705	94211	06716	11738	55784	95374	72555
33	59497	04097	09419	89964	51211	04894	22882	17805	21896	83864
34	97155	13428	40293	09985	58434	01412	69124	82171	59058	82859
35	98409	66162	95260	47420	20792	61527	20441	09435	11859	41567
36	45476	84882	62109	95597	25930	66290	69706	61203	53634	22557
37	89300	69700	50741	30329	11658	70166	05400	66669	48708	03887
38	50051	95137	91631	66315	91428	12275	24816	58081	71710	33258
39	31253	85178	31310	89642	98364	02306	24617	09609	83947	22716
40	79152	53829	77250	20190	56535	18760	69942	77448	33718	48805
41	44560	38750	83635	56540	64900	42912	11953	79049	19710	68618
42	68328	83378	63369	71383	39564	05618	42491	64559	97501	65747
43	46439	38689	58625	08342	30459	85863	20281	09284	26333	91777
44	83544	86141	15707	96256	23068	13782	08467	89469	93842	55349
45	91621	00881	04900	54224	46127	55109	17852	27491	89415	23466
46	91896	57126	04151	03795	53077	11848	12630	98325	52068	60142
47	55751	62515	21308	80830	02261	29303	37204	96926	30506	08808
48	85156	87689	95493	88847	00564	55017	95534	17371	69448	82530
49	07521	96898	12236	60273	39102	62315	12239	07105	11844	01117

Figura 1.2. Tabla de números al azar

Para ilustrar el manejo de la tabla de números al azar, considere el ejemplo 2, en el cual se quería tomar una muestra de cinco estudiantes de un grupo de sesenta.

En primer término, debe tomarse la lista de los alumnos (esta se usará como marco de la muestra) y asignarle a cada uno de ellos un número de orden entre 01 y 60. Luego se elige una tabla de números al azar si el libro contiene varias páginas con tablas de números, se selecciona una página cualquiera y se pone, sin mirar, el lápiz sobre ella. Suponga que la página con la tabla seleccionada es la que se muestra en la figura 1.2. y el lápiz cayó sobre el dígito 6, situado en la línea 27 y la columna 22, que forma parte del número 82 623 (ver figura 1.2.). Para hacer la selección, se anotan todos los números de dos cifras a partir del 62, hacia abajo, o se

62
04
64
40
45
21
21
43

El proceso se detiene cuando han aparecido cinco números diferentes entre 01 y 60. Debe notarse que solo interesan los de dos cifras entre 01 y 60; por lo tanto, cualquiera que surja entre 61 y 99, así como 00, no debe considerarse. Además, si uno se repite, como el caso del 21 en el presente ejemplo, solo se considera una vez.

Para este caso los estudiantes escogidos serán los que se señalan en la lista con los números de orden 04, 40, 45, 21 y 43.

Cuando sea necesario hacer otra selección aleatoria, puede procederse de nuevo a escoger tabla de números al azar y un punto de arranque, o partir de donde se detuvo la elección anterior.⁷

En la actualidad, es frecuente recurrir directamente a opciones ofrecidas por los programas de generación de números aleatorios de las computadoras. Así, en el caso de la selección de los 5 estudiantes, la escogencia puede hacerse, por ejemplo, creando los cinco números entre 01 y 60 mediante la opción disponible (ALEATORIO (), en la versión en Español o RAND (), en la versión en Inglés) en la hoja de cálculo Excel.

7. En las tablas de números al azar se incluyen explicaciones sobre métodos más refinados para la selección de la página y del punto de partida.

1.9. ATRIBUTOS Y VARIABLES

Los datos con los cuales se realiza el análisis estadístico provienen de las operaciones de medir o de contar. Así, cuando interesa el peso de estudiantes, lo que se hace es medir esta característica con una balanza y se reporta usando cierto tipo de unidades. Cada estudiante que se pese arrojará un número de kilogramo, 61, 57, 73. Si estudian parejas y la característica que interesa es la cantidad de hijos, simplemente se cuenta el total de hijos y, se anota el número correspondiente a cada una: 0, 1, 2, 3, etcétera.

Ahora bien, hay otras situaciones en las cuales la dimensión cuantitativa de la característica no es evidente, como sí lo es en el caso del peso y el número de hijos, y la operación de medir –aun en un sentido más amplio– no parece tan simple. Suponga que tenemos un grupo de mujeres entre 20 y 49 años cumplidos y nos interesa la característica estado conyugal (o marital), ¿se puede medir el estado conyugal? ¿se puede medir una cualidad? En la forma como se hace con características como el peso y el número de hijos, claro que no; se adopta un procedimiento un poco diferente. Se definen las categorías que comprende el estado conyugal: soltero, casado, viudo, divorciado, separado y en unión libre; luego, se determina para cada *unidad de estudio* a cuál de las categorías pertenece cada mujer. Para distinguirlas se asigna a cada una, un símbolo, letra o número, por ejemplo: puede acordarse que S representa soltera; C, casada; V, viuda; D, divorciada; SE, separada y UL, unión libre. Bajo este acuerdo a una viuda se le asignará una V, a una soltera una S, entre otras. Hecha esa identificación resulta fácil contar cuántas solteras hay, cuántas casadas, etcétera.

En la práctica, sin embargo, para facilitar el manejo electrónico de los datos, lo usual es asignar números a las categorías: por ejemplo, que unión libre sea 1; casado, 2; separado, 3; divorciado, 4; viudo, 5 y soltero, 6. Conviene señalar que esta asignación de números tiene únicamente fines de clasificación y es totalmente arbitraria, puede emplearse, para el mismo fin, cualquier juego de seis enteros positivos. Por ejemplo, la siguiente asignación sería totalmente válida: unión libre, 9; casado, 8; separado, 5; divorciado, 4; viudo, 3 y soltero, 1. Por esta razón, no es lícito ejecutar operaciones aritméticas o realizar cálculos con estos números como si lo es en el caso de las variables anteriores: peso y número de hijos.⁸

Desde esta perspectiva, cuando se mide una característica se obtiene un número para cada *unidad de estudio observada*; como estos números usualmente varían de una *unidad de estudio* a otra, se habla de *variables*. Así, el peso es una variable, y lo son también la cantidad de hijos y el estado civil, similar a cualquiera que varía de un elemento a otro de la población del estudio.

8. En las variables peso y número de hijos, el cálculo del promedio es lícito y tiene una interpretación válida. Pero el promedio de los valores de la variable estado conyugal, aunque es posible realizarlo, carece de sentido.

Las *variables* antes citadas tienen ciertas diferencias de comportamiento, de posibilidades de variación. Atendiendo a estas discrepancias, se habla de *variables cuantitativas* (peso y número de hijos, por ejemplo) y de *variables cualitativas o atributos* (como el estado civil). Dentro de las cuantitativas se distingue entre *continuas*, aquellas que toman cualquier valor dentro de un intervalo de números reales: el peso y la estatura, y *discretas*, cuyos valores son numerables, como el total de hijos o de goles en un partido de fútbol. La figura 1.3 resume los tipos de características discutidos.

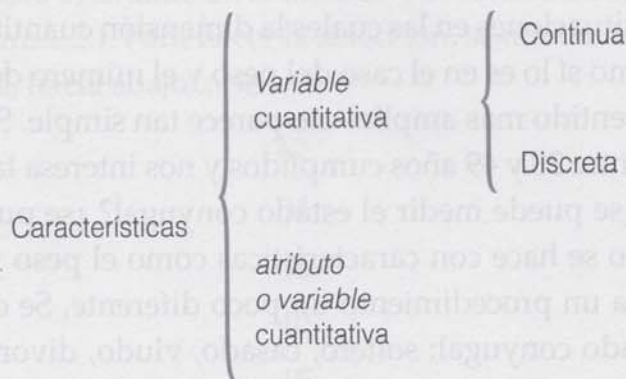


Figura 1.3. Tipos de características que se pueden recoger en un estudio

Una *variable* es continua en un intervalo si teóricamente puede tomar cualquier valor en él, ejemplos: peso, estatura, edad, velocidad. No importa cuán pequeña sea la diferencia entre las estaturas de dos personas, siempre es posible encontrar otra con una estatura entre esas dos. El instrumento de medida puede impedir esa distinción, pero aparentemente existe esa medida intermedia. En el caso de la edad, también es clara la naturaleza continua de la variable: para pasar de una edad a otra, la persona debe ir teniendo todas las edades intermedias. Cualquier edad existe, cualquiera es real, aunque por costumbre y comodidad se da en años y, en ocasiones, en años y meses cumplidos.

Si una variable solo puede tomar un número determinado de valores dentro de un intervalo, y no todos los posibles, se dice que es discontinua o *discreta*. Por lo tanto, una *variable discreta* cuantitativa tiene la particularidad de que no puede asumir todos los posibles valores dentro de un intervalo. Un ejemplo muy claro de una variable discreta es el total de goles de un partido de fútbol, el cual puede ser 0, 1, 2, 3,... Su naturaleza le impide asumir un valor fraccionario, así jamás podría decirse que las anotaciones de un partido fueron 1,37 o 2,1857; tampoco tendría sentido decir que Brasil venció a Argentina 2,71 a 1,58.

La situación es otra para las *variables continuas*. El peso, por ejemplo, puede darse en un número exacto de kilogramos, pero si se desea mayor precisión, se obtendría en gramos, miligramos, entre otros. El valor 2,4853 kg es completamente real.

Conviene aclarar que las *variables discretas* no cambian solo en unidades enteras. Así, si la mantequilla se vende solo en octavos de kilogramo, entonces las ventas medidas en kilogramos serían discretas y variaría en la siguiente forma: 0,125 kg; 0,25; 0,375, entre otras. También podría ser expresada en octavos de kilogramo y las cantidades vendidas asumirían los valores: un octavo, dos octavos, tres octavos, entre otros. A continuación se presentan, con fines ilustrativos, algunos ejemplos de variables y atributos.

Cuadro 1.2
EJEMPLOS DE VARIABLES

VARIABLES			
Unidad elemental	Característica	Unidades medida	Ejemplo de observación
Un estudiante	Peso	Kilogramos	64,5
Una casa	Valor	Colones	35 000 000
Una casa	Número de dormitorios	Dormitorios	3
Un bombillo	Duración	Horas	480
Una máquina	Producción diaria	Unidades producidas	525
Una venta	Monto	Colones	4450
Un maestro	Experiencia	Años	5

Cuadro 1.3
EJEMPLOS DE ATRIBUTOS

ATRIBUTOS		
Unidad elemental	Característica	Posibles observaciones
Un estudiante	Clase de alumno	Regular-especial-oyente
Una casa	Condición de alquiler	Amueblada-sin muebles
Un bombillo	Condición	Defectuoso-no defectuoso
Una venta	Tipo de artículo	Jabón, jugo naranja, betún, etc.
Un empleado	Opinión del jefe	Muy bueno, bueno, regular, malo
Un maestro	Sexo	Masculino, femenino

En ciertas situaciones, por razones prácticas, puede tratarse de una *variable discreta* como si fuera *continua*, tal es el caso de las cifras que expresan montos de pagos, de ventas y de ingresos en colones, las cuales, en la Costa Rica actual, usualmente solo asumen valores

en colones enteros, pero pueden considerarse continuas sin que sea importante el error que significa esta aproximación.⁹

También, cuando se tienen *variables cualitativas* que solo asumen dos valores, como en los casos antes mencionados del sexo y de condición de un bombillo (defectuoso, no defectuoso), se puede asignar 0 y 1 a esas dos categorías y obtener las denominadas *variables dicotómicas*¹⁰ o “cero-uno”, las cuales pueden manejarse como si fueran *discretas*.

1.10. ESTADÍSTICA DESCRIPTIVA Y ESTADÍSTICA INFERENCIAL

El campo de acción de la estadística, en la actualidad, es muy amplio y variado, pero algo común a todas sus aplicaciones es que en ellas predomina generalmente uno de estos dos intereses: a) la descripción del conjunto de datos observados o b) la obtención de conclusiones o *inferencias* para un grupo mayor que el observado, o sea, la población de interés. Así, algunos estudian y usan la estadística porque permite recoger y resumir una colección de datos, de tal manera que las características y relaciones entre los elementos sean fácilmente comprensibles. En estas aplicaciones, el interés se limita exclusivamente a los datos que se analizan y la preocupación básica es lograr una descripción o caracterización apropiada de ellos. Por eso se le denomina *estadística descriptiva* y también se utiliza este nombre para hacer referencia a aquellas técnicas o instrumentos empleados cuando el interés primordial es describir un conglomerado de datos, sin importar la profundidad y detalle con que se haga, pues no se pretende hacer generalizaciones o inferencias para un conjunto mayor. La confección de *cuadros* y *gráficos*, el cálculo de *promedios*, *porcentajes* y *medidas de variabilidad*, son algunas de las técnicas ordinariamente empleadas en la *estadística descriptiva*.

Por otra parte, la estadística puede ser aplicada con el propósito de extraer, de los datos disponibles, conclusiones que se generalizan a un conjunto de unidades mucho mayores y de las cuales fueron extraídas las unidades que proporcionaron los datos analizados. Así, un profesor que prueba un nuevo método de enseñanza de la química en unos pocos grupos de estudiantes de secundaria está interesado en generalizar esos resultados a todos los otros alumnos de secundaria del país o, al menos, del Área Metropolitana de San José. Su propósito básico no es describir el comportamiento en los que se prueba el método, sino inferir o extender sus conclusiones a todo el sistema de educación secundaria –la población– a partir de lo observado en los grupos de estudiantes estudiados en la muestra.

9. Por excepción, en algunos casos, como cuando se paga con cheque, se pide escribir la suma con detalle en céntimos, pero aún así se tiene una variable discreta.
10. Son aquellas que por su propia naturaleza solo pueden manifestarse según dos modalidades: por ejemplo, vivo-muerto; hombre-mujer, encendido-apagado, etcétera.

Este tipo de aplicaciones corresponde a la llamada *estadística inferencial* o *inductiva*, este nombre se le aplica también a aquellas técnicas o procedimientos empleados cuando el propósito perseguido es no solo describir los datos sino generalizar lo observado en ellos para un conjunto o universo mayor del cual fueron tomados y analizados. La *inferencia estadística* es, por lo tanto, un proceso esencialmente inductivo: se toma una parte –muestra–, se examina y los resultados se generalizan al conjunto o universo –población– del cual fue seleccionada la parte que se analiza.¹¹

El proceso de inferencia se esquematiza en la figura 1.4, como puede notarse, se desea conocer las características de una población a partir de una muestra. Para que esta sea representativa, deben resolverse previamente dos problemas básicos, a saber: el *tamaño* que debe tener la muestra y el *método* que se seguirá para seleccionarla. Una vez analizada la información recolectada, los resultados se generalizan a la población. Este proceso implica haber resuelto dos dificultades: ¿qué método usar para la generalización?, y ¿cómo medir el error asociado a esa generalización o inferencia? Si la muestra es seleccionada aleatoriamente, se tendrá un modelo probabilístico para describir el proceso de muestreo y será posible realizar una medición válida del error asociado a la inferencia.¹² En este caso, cuando la muestra ha sido escogida aleatoriamente y es posible medir el error de muestreo, se tiene una *inferencia estadística*. Si la selección es de juicio o de conveniencia, no se dispone de un procedimiento objetivo para medir el error asociado a la inferencia.

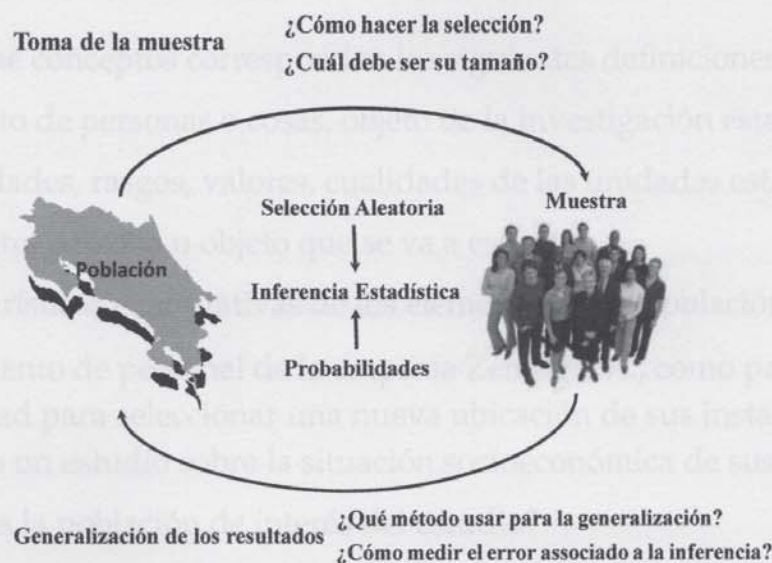


Figura. 1.4. Esquema de inferencia estadística

11. El que el proceso de inferencia incluya elementos descriptivos y se utilicen procedimientos que también se emplean en la estadística descriptiva, no debe oscurecer esta diferencia.
12. Técnicamente, se diría que la muestra es “medible”, o sea, reúne las características y contiene la información requerida para calcular el error de muestreo usando los conocimientos de la teoría estadística.

EJERCICIOS DE AUTOEVALUACIÓN

- Subraye en cada columna los ejemplos correspondientes con el término que encabeza la columna.

ESTADÍSTICAS

Número de personas en el padrón electoral por distritos
 Anuario estadístico
 Unidades estadísticas
 Publicaciones censales
 Balanza de pagos de Costa Rica
 Datos estadísticos
 Datos numéricos
 Número turistas recibidos el año pasado
 Diagrama

ESTADÍSTICA

Observación de la población
 Población real
 Tablas numéricas
 Es una ciencia
 Estudio de fenómenos de masa
 Descripción de un fenómeno
 Vehículos registrados en el 2008
 Mortalidad infantil
 Estudio del fenómeno turístico
 Producto interno bruto

- Una vez efectuado el ejercicio anterior, defina los términos “estadísticas” y “estadística”
- Indique a qué conceptos corresponden las siguientes definiciones:
 - Conjunto de personas o cosas, objeto de la investigación estadística.
 - Propiedades, rasgos, valores, cualidades de las unidades estadísticas.
 - Elemento, persona u objeto que se va a estudiar.
 - Características cuantitativas de los elementos de la población.
- El departamento de personal de la empresa Zemog S.A., como parte de un estudio de factibilidad para seleccionar una nueva ubicación de sus instalaciones, requiere llevar a cabo un estudio sobre la situación socioeconómica de sus empleados.
 - ¿Cuál es la población de interés del estudio?
 - ¿Liste 4 o 5 características que podrían ser recogidas en el estudio?
 - ¿Cuál es la unidad estadística del estudio?
 - Escriba tres ejemplos distintos de poblaciones, desde el punto de vista estadístico, que pudieran derivarse de este estudio.
 - Para cada una de las características indicadas en (b), señale ejemplos de observaciones posibles.

5. Exponga la diferencia entre unidad de estudio y población.
6. Cuando se refiere a la población, puede hacerlo desde dos puntos de vista. Señale e indique cuál es el enfoque estadístico.
7. A continuación se dan ejemplos de poblaciones, especifique si se tratan de poblaciones finitas o infinitas, situando dentro del cuadro correspondiente la letra que las precede.

	FINITAS	INFINITAS
POBLACIÓN		

- a) Salarios de los trabajadores de una fábrica determinada, durante el año anterior.
 - b) Personas a las que se les puede aplicar una vacuna.
 - c) Estudiantes que se matricularán en la carrera de Biología en los próximos años.
 - d) Población económicamente activa de Costa Rica, por rama de actividad, según el último censo de población.
 - e) Resultados obtenidos al lanzar, sin límite, un dado.
 - f) Producción anual de energía eléctrica en los últimos 10 años.
 - g) Salarios de los obreros.
8. Determine si las afirmaciones que a continuación se exponen son verdaderas o falsas, anotando una V o una F, según corresponda.
 - a) Cuando se selecciona una muestra utilizando el juicio de un experto, hay que tener muy en cuenta los errores aleatorios cometidos.
 - b) La representatividad de la muestra está en razón directa a su tamaño y a la homogeneidad que presenten los elementos de la población.
 - c) En un análisis estadístico, cuanto mayor es la muestra, menor es el error aleatorio.
 - d) Los sesgos de selección son realmente errores aleatorios o de muestreo.
 9. Señale tres características por las que se prefiere el uso de las muestras aleatorias a las de juicio o de conveniencia.
 10. En la lista de características que a continuación se le presenta, para cada una indique, si se trata de un atributo (variable cualitativa) o de una variable (variable cuantitativa) y, en este caso, si es discreta o continua.
 - a) Peso de los estudiantes
 - b) Estado civil de los profesores

- c) Ingresos familiares
 - d) Número de hijos
 - e) Edad de los estudiantes
 - f) Clases de insectos
 - g) Duración de las carreras
 - h) Categorías profesionales
 - i) Producción de la editorial
 - j) Tipos de papel
 - k) Número de habitantes
 - l) Consumo familiar de electricidad
 - m) Humedad relativa
 - n) Especies botánicas
 - o) Calidad de un producto
 - p) Calificaciones de un curso
11. Determine si las afirmaciones que a continuación se exponen son verdaderas o falsas, anotando una V o una F, según corresponda.
- a) La selección aleatoria garantiza que la muestra refleja exactamente lo que sucede en la población.
 - b) La estadística inferencial trata de generalizar los datos obtenidos en la muestra a la población.
 - c) El cálculo de promedios, porcentajes y medidas de variabilidad es el objeto exclusivo de la estadística inferencial.
 - d) La tabla de números al azar se utiliza para garantizar que los elementos escogidos no se repetirán al seleccionar otras muestras.
 - e) La estadística descriptiva trata de describir el conjunto de datos, sin pretender generalización alguna.
12. Defina en sus propias palabras los términos “estadística descriptiva” y “estadística inferencial”. Utilice para ello sus respuestas a los ejercicios anteriores.

RESPUESTA A LOS EJERCICIOS DE AUTOEVALUACIÓN

1.

ESTADÍSTICAS

Número de personas en el padrón electoral por distrito
 Anuario estadístico
 Publicaciones censales
 Balanza de pagos de Costa Rica
 Datos estadísticos
 Número de turistas recibidos el año pasado
 Datos numéricos

ESTADÍSTICA

Es una ciencia
 Estudio de fenómenos de masa
 Descripción de un fenómeno
 Estudio del fenómeno turístico

2. Las estadísticas son un conjunto de datos numéricos que han sido organizados, resumidos y presentados para mostrar las características o evolución de un fenómeno de interés. Esto es lo que se tiene en mente cuando se habla de estadísticas de población, estadísticas de educación, estadísticas de los Juegos Deportivos Nacionales, etc. El término estadística se refiere a una disciplina científica dedicada al desarrollo de la teoría y técnicas apropiadas para la recolección, clasificación, presentación, análisis e interpretación de información cuantitativa obtenida a través de la observación o experimentación.

3. a) Población

b) Características

c) Unidad estadística

d) Variables

4. a) Todos los empleados actuales de la empresa Zemog S.A.

b) Ingreso familiar, nivel educativo del empleado, tamaño de la familia, número de años en la empresa, tenencia de automóvil; casa propia o alquilada.

c) El empleado de la empresa Zemog S.A.

d) Población de:

- Salarios percibidos por los empleados de la empresa en un cierto período.
- Tamaño de las familias de los empleados de la empresa.
- Número de años trabajados por cada empleado de la empresa.

- e) Ejemplos de observaciones
- Ingreso familiar: 415 000 colones mensuales
 - Nivel educativo empleado: tercer año de secundaria
 - Tamaño de la familia: 5 personas
 - Número de años en la empresa: 11 años
 - Tiene automóvil: no
 - Forma de tenencia de la vivienda: propia pagándola
5. La diferencia entre unidad de estudio y población estriba en que sobre la primera recae la observación, mientras que la población incluye todas las posibles observaciones de una misma característica.
6. Cuando se hace referencia a la población, puede ser desde dos puntos de vista:
- a) Como un conjunto de elementos con una característica común, objeto del estudio.
 - b) Como el conjunto de todas las observaciones de la misma característica para un conjunto de elementos. Este es el enfoque estadístico.

7.

	FINITAS	INFINITAS
POBLACIÓN	a, d, f	b, c, e, g

8. a) F
b) F
c) V
d) F
9. a) Eliminan los sesgos de selección.
b) Permiten medir la confiabilidad de las inferencias realizadas con la muestra.
c) El error de muestreo puede hacerse tan pequeño como se quiera aumentando el tamaño de la muestra.

10. a) Variable continua
 b) Atributo
 c) Variable continua
 d) Variable discreta
 e) Variable continua
 f) Atributo
 g) Variable continua
 h) Atributo
 i) Variable discreta
 j) Atributo
 k) Variable discreta
 l) Variable continua
 m) Variable continua
 n) Atributo
 o) Atributo
 p) Variable continua
11. a) F
 b) V
 c) F
 d) F
 e) V
12. La estadística descriptiva comprende las técnicas e instrumentos que se emplean cuando únicamente se desea describir y analizar un conjunto de datos, sin pretender ninguna generalización a un grupo mayor.

La estadística inferencial es aquella a la cual le interesa generalizar lo observado en una muestra de datos a un conjunto mayor, del cual fueron seleccionados; comprende la teoría y las técnicas para ese fin.

- Salarios percibidos por los empleados de la empresa en un cierto periodo.
- Tamaño de las familias de los empleados de la empresa.
- Número de años trabajados por cada empleado de la empresa.