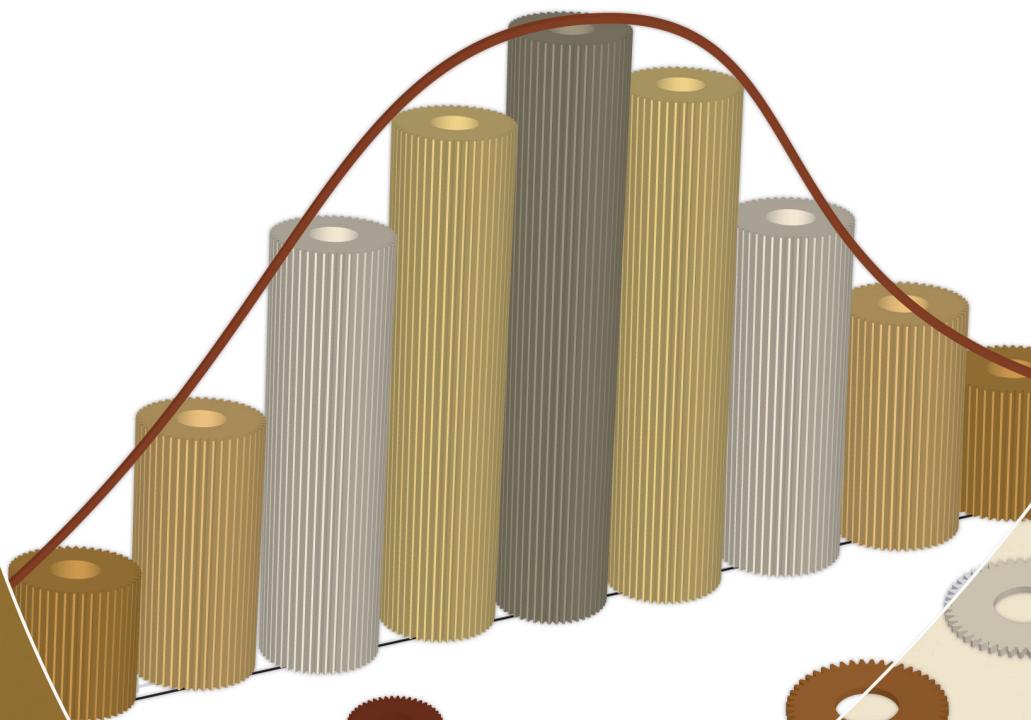




Métodos estadísticos para Ingenieros

Rafael Romero Villafranca
Luisa Zúnica Ramajo



$$V = 2 \times \pi \times R \times \frac{V}{\pi R^2}$$
$$R + \pi R^2$$
$$R = \sqrt[3]{\frac{100}{3,14}} = 3,17$$
$$10a(a+b)x + (4a^3$$
$$S_1 + 10$$

EDITORIAL
UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Rafael Romero Villafranca
Luisa Rosa Zúnica Ramajo

Métodos estadísticos para ingenieros

EDITORIAL
UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Colección Académica

Para referenciar esta publicación utilice la siguiente cita: ROMERO VILAFRANCA, R.; ZÚNICA RAMAJO, L. (2013) *Métodos estadísticos para ingenieros*. Valencia: Universitat Politècnica

Primera edición, 2013

© Rafael Romero Villafranca
Luisa Zúnica Ramajo

© de la presente edición: Editorial Universitat Politècnica de València
Distribución: Telf. 963 877 012/ <http://www.lalibreria.upv.es> / Ref.: 4

Imprime: Byprint Percom, sl

ISBN:978-84-9048-071-7

Impreso en papel Coral Book



Impreso bajo demanda

Queda prohibida la reproducción, la distribución, la comercialización, la transformación y, en general, cualquier otra forma de explotación, por cualquier procedimiento, de la totalidad o de cualquier parte de esta obra sin autorización expresa y por escrito de los autores.

Impreso en España

A nuestros hijos Rafa, Belén y Gerardo

PREFACIO

El presente libro constituye una versión reducida del texto "Métodos Estadísticos en Ingeniería" (al que nos referiremos en adelante en este Preámbulo como MEI), publicado por los autores a principios de los años noventa (SPUPV 93.637). Dicha obra ha venido siendo utilizada desde entonces, como texto de referencia para la docencia de Estadística en varias titulaciones de Ingeniería de la Universidad Politécnica de Valencia.

La necesidad de esta versión reducida es consecuencia de la modificación de los currícula de Estadística en esta Universidad, derivada de implantación de las titulaciones de Grado en el marco de los acuerdos de Bolonia. Mientras que en las titulaciones anteriores la materia de Estadística tenía asignados, en general, 10 o más créditos, en los actuales títulos de Grado esta asignación se ha reducido generalmente a sólo 6 créditos. Con esta carga docente es imposible impartir todos los contenidos que aparecían en MEI.

Este nuevo texto mantiene el espíritu que inspiró la obra anterior, elaborada en un contexto que propugna una enseñanza enfocada hacia la consecución tanto del "saber" como del "saber hacer" por parte del alumno, y potencia del carácter activo, dialéctico, práctico y crítico del proceso de enseñanza/aprendizaje, otorgando al alumno el protagonismo principal en el mismo.

Este libro, que recoge también nuestra experiencia de muchos años enseñando técnicas estadísticas a ingenieros y técnicos en empresas e industrias de sectores muy diversos, mantiene, en nuestra opinión, un conjunto de características distintivas que aparecían ya en MEI:

- Estar planteado para cubrir las materias estadísticas más importantes para un ingeniero (análisis de la varianza y modelos de regresión), partiendo de conocimientos estadísticos previos nulos y en una asignatura con un número reducido de créditos
- Tratar a nivel operativo algunos temas avanzados (efectos de dispersión en Anova, modelos de regresión con variables "dummy", etcétera...) que frecuentemente no se encuentran, incluso en textos de mayor nivel.
- Introducir los conceptos básicos necesarios de probabilidad e inferencia de forma intuitiva, orillando formalizaciones matemáticas imposibles de desarrollar si se pretende cubrir los objetivos anteriores en el tiempo disponible y con la formación actual de nuestros alumnos.

- Dar gran importancia de los ejemplos reales, muchos de ellos fruto de nuestra relación profesional con la industria, que se utilizan en muchos casos para introducir conceptos básicos de probabilidad y de inferencia
- Profusa inclusión de breves apartados de Autoevaluación, pensados no sólo como ejercicios con dicho fin, sino también a veces como cuestiones abiertas para que el lector empiece a plantearse temas que se desarrollan posteriormente.

Las modificaciones más importantes de este texto respecto al libro MEI son las siguientes:

- Los dos primeros capítulos de MEI se han agrupado en uno, prescindiendo de los apartados sobre historia de la Estadística e importancia de esta ciencia en la Ingeniería.
- El tratamiento descriptivo de las variables bidimensionales se ha pospuesto a un breve capítulo previo al de Modelos de Regresión, eliminándose los apartados que trataban la recta de regresión, puesto que ésta se estudia con detalle a nivel inferencial en dicho capítulo posterior.
- Se ha introducido un nuevo capítulo sobre inferencia con datos de frecuencias, tema que no se trataba en MEI.
- Se ha eliminado el capítulo sobre distribuciones en el muestreo. Las ideas básicas al respecto, así como las distribuciones G_i-2 , t de Student y F de Fisher, se introducen a medida que se van necesitando en los capítulos de Inferencia.
- En el capítulo sobre inferencia en poblaciones normales, se han eliminado los apartados que trataban de la comparación de dos poblaciones y del análisis con datos apareados. Estas cuestiones se abordan como casos particulares de los tratados en los temas de Análisis de la Varianza
- Se ha incluido un nuevo capítulo con una introducción general a nivel intuitivo del Análisis de la Varianza, En el capítulo sobre Análisis de la Varianza con uno o varios factores se han eliminado ciertas cuestiones, como el uso de contrastes ortogonales, o la formalización matemática de los conceptos expuestos.
- Se han eliminado los dos capítulos que trataban sobre los diseños 2^k y las Fracciones Factoriales. Algunos ejemplos personales que consideramos especialmente interesantes de diseños 2^k se han conservado como ejercicios en el capítulo sobre Análisis de la Varianza.

Cada capítulo se completa con un apartado anejo con las respuestas a muchas de las Autoevaluaciones planteadas en el mismo (aquéllas de mayor dificultad o interés), así como con un conjunto de ejercicios resueltos y ejercicios propuestos adicionales.

Los autores deseamos manifestar nuestro agradecimiento a todos los profesores y alumnos que a lo largo de estos años han venido utilizando nuestro anterior texto para impartir o estudiar sus asignaturas de estadística. Sus sugerencias y críticas han sido un input muy importante para la elaboración de esta nueva versión, en la que esperamos haber satisfecho al menos algunas de ellas. De forma muy especial, los autores al elaborar esta nueva versión hemos tenido permanentemente presente el recuerdo de nuestro querido hijo Rafael Romero Zúnica (q.e.p.d.) que tanto nos ayudó en la preparación del texto anterior

Valencia, junio de 2013.

ÍNDICE

CAPÍTULO 1. INTRODUCCIÓN

1.1. OBJETO DE LA ESTADÍSTICA	17
1.2. POBLACIONES	18
1.3. VARIABLES ALEATORIAS	19
1.3.1. Concepto	19
1.3.2. Variables discretas y variables continuas.....	20
1.3.3. Variables k-dimensionales	20
1.4. MUESTRAS. DATOS ESTADÍSTICOS	21
1.5. ESTADÍSTICA DESCRIPTIVA E INFERENCIA ESTADÍSTICA	22
1.6. ENCUESTA.....	24

CAPÍTULO 2. ESTADÍSTICA DESCRIPTIVA UNIDIMENSIONAL

2.1. INTRODUCCIÓN.....	27
2.2. TABLAS DE FRECUENCIAS	27
2.3. DIAGRAMAS DE BARRAS Y DE TARTA	30
2.4. HISTOGRAMAS	31
2.5. PARAMETROS DE POSICIÓN.....	32
2.5.1. Media.....	33
2.5.2. Mediana.....	34
2.5.3. Cuartiles	34
2.6. PARÁMETROS DE DISPERSIÓN	35
2.6.1. Recorrido.....	36
2.6.2. Varianza. Desviación típica	36
2.6.3. Intervalo intercuartílico	38
2.7. PARÁMETROS DE ASIMETRÍA Y DE CURTOSIS	38
2.7.1. Coeficiente de Asimetría	39
2.7.2. Coeficiente de Curtosis	40
2.8. DIAGRAMAS BOX-WHISKER	42
2.A. AUTOEVALUACIONES RESUELTAS Y EJERCICIOS	43
2.A.1. Respuesta a algunas Autoevaluaciones	43
2.A.2. Ejercicios adicionales.....	49

CAPÍTULO 3. CONCEPTOS BÁSICOS DE CÁLCULO DE PROBABILIDADES

3.1. INTRODUCCIÓN.....	51
3.2. SUCESOS. OPERACIONES CON SUCESOS	52
3.3. PROBABILIDAD	55
3.3.1. Concepto de Probabilidad	55
3.3.2. Propiedades de la Probabilidad	55
3.3.3. Probabilidad de la suma de sucesos.....	56
3.4. ESPACIOS DE PROBABILIDAD SIMÉTRICOS	57
3.5. PROBABILIDAD CONDICIONAL	58
3.6. TEOREMA DE LA PROBABILIDAD TOTAL	60
3.7. INDEPENDENCIA DE SUCESOS	60

3.8. TEOREMA DE BAYES	61
3.A. AUTOEVALUACIONES RESUELTAS Y EJERCICIOS.....	63
3.A.1. Respuesta a algunas Autoevaluaciones	63
3.A.2. Ejercicios resueltos.....	65
3.A.3. Ejercicios adicionales	68

CAPÍTULO 4. DISTRIBUCIONES DE PROBABILIDAD

4.1. INTRODUCCIÓN.....	71
4.2. FUNCIÓN DE DISTRIBUCIÓN.....	72
4.3. DISTRIBUCIONES DISCRETAS.....	74
4.4. DISTRIBUCIONES CONTINUAS. FUNCIÓN DE DENSIDAD	75
4.5. VARIABLES BIDIMENSIONALES. INDEPENDENCIA	78
4.6. ESPERANZA MATEMÁTICA	79
4.6.1. Variables discretas	79
4.6.2. Variables continuas	80
4.7. MOMENTOS CENTRALES	81
4.7.1. Varianza. Desviación Típica.....	81
4.7.2. Coeficientes de asimetría y de Curtosis	82
4.A. AUTOEVALUACIONES RESUELTAS Y EJERCICIOS.....	82
4.A.1. Respuesta a algunas Autoevaluaciones	82
4.A.2. Ejercicios resueltos.....	83
4.A.3. Ejercicios adicionales	87

CAPÍTULO 5. PRINCIPALES DISTRIBUCIONES DISCRETAS

5.1. INTRODUCCIÓN.....	89
5.2. DISTRIBUCIÓN DE DOS PUNTOS	89
5.3. DISTRIBUCIÓN BINOMIAL.....	90
5.4. DISTRIBUCIÓN DE POISSON.....	91
5.5. PLANES DE INSPECCIÓN	95
5.A. AUTOEVALUACIONES RESUELTAS Y EJERCICIOS.....	97
5.A.1. Respuesta a algunas Autoevaluaciones	97
5.A.2. Ejercicios adicionales	100

CAPÍTULO 6. LA DISTRIBUCIÓN NORMAL

6.1. INTRODUCCIÓN.....	103
6.2. DISTRIBUCIÓN NORMAL: CONCEPTO Y PROPIEDADES	103
6.3. TABLA DE LA DISTRIBUCIÓN NORMAL.....	105
6.4. PAPEL PROBABILÍSTICO	108
6.5. APROXIMACIONES NORMALES.....	110
6.6. LA DISTRIBUCIÓN LOGNORMAL.....	112
6.A. AUTOEVALUACIONES RESUELTAS Y EJERCICIOS.....	114
6.A.1. Respuesta a algunas Autoevaluaciones	114
6.A.2. Ejercicios resueltos.....	116
6.A.3. Ejercicios adicionales	117

CAPÍTULO 7. INFERENCIA ESTADÍSTICA CON DATOS DE FRECUENCIAS	
7.1. INTRODUCCIÓN.....	119
7.2. EJEMPLOS DE TESTS DE CONTRASTE DE PROPORCIONES	120
7.3. HIPÓTESIS NULA.....	120
7.4. RIESGOS DE 1ª Y DE 2ª ESPECIE.....	121
7.5. CONTRASTE DE PROPORCIONES	122
7.6. LA DISTRIBUCIÓN GI-DOS.....	124
7.7. ANÁLISIS DESCRIPTIVO DE TABLAS DE FRECUENCIAS.....	127
7.8. INFERENCIA EN TABLAS DE FRECUENCIAS	128
7.8.1. Test de independencia.....	128
7.8.2. Test de homogeneidad.....	130
7.8.3. Generalizaciones.....	132
7.A. EJERCICIOS	134
7.A.1. Ejercicio resuelto	133
7.A.2. Ejercicios adicionales.....	134
CAPÍTULO 8. INFERENCIA SOBRE UNA POBLACIÓN NORMAL	
8.1. INTRODUCCIÓN.....	137
8.2. UN EJEMPLO.....	137
8.3. NORMALIDAD DE LOS DATOS	138
8.4. ANÁLISIS DESCRIPTIVO DE LA MUESTRA	140
8.5. ESTUDIO DE LA HIPÓTESIS $m = 2000$. ENFOQUE DEL PROBLEMA ..	141
8.6. DISTRIBUCIONES EN EL MUESTREO	142
8.6.1. Introducción.....	142
8.6.2. Distribución de \bar{x}	143
8.6.3. La distribución t de Student.....	144
8.7. CONTRASTE DE LA HIPÓTESIS $m = 2000$	147
8.8. INTERVALO DE CONFIANZA PARA m	149
8.9. INTERVALO DE CONFIANZA PARA σ	150
8.10. ANÁLISIS MEDIANTE STATGRAPHICS	151
8.11. INTERVALO DE CONFIANZA PARA UNA PROPORCIÓN	152
8.A. AUTOEVALUACIONES RESUELTAS Y EJERCICIOS	153
8.A.1. Respuesta a algunas Autoevaluaciones	153
8.A.2. Ejercicios adicionales.....	155
CAPÍTULO 9. INTRODUCCIÓN AL ANÁLISIS DE LA VARIANZA	
9.1. COMPARACIÓN DE 2 POBLACIONES.....	157
9.1.1. Ejemplos.....	157
9.1.2. Terminología	158
9.2. COMPARACIÓN DE K POBLACIONES	159
9.3. ESTUDIOS MULTIFACTORIALES	160
9.3.1. Estudio simultáneo de varios factores.....	160
9.3.2. Poblaciones estudiadas	161
9.3.3. Concepto de efecto simple de un factor	162
9.3.4. Concepto de interacción entre dos factores.....	163

9.4. INTRODUCCIÓN AL ANÁLISIS DE LA VARIANZA	164
9.4.1. Introducción	164
9.4.2. Idea intuitiva del ANOVA	164
9.4.3. Cuadro resumen del Análisis de la Varianza	167
9.5. LA DISTRIBUCIÓN F	171

CAPÍTULO 10. ANÁLISIS DE LA VARIANZA UNIFACTORIAL Y MULTIFACTORIAL

10.1. ANOVA CON UN SOLO FACTOR	175
10.1.1. Un ejemplo	175
10.1.2. Cuadro Resumen del ANOVA	176
10.1.3. Análisis de los residuos	176
10.1.4. Comparación de medias. Intervalos LSD	178
10.1.5. Estudio de efectos sobre varianzas	179
10.1.6. Realización práctica de los cálculos	181
10.1.7. Análisis mediante Stagraphics	182
10.2. ANOVA EN PLANES MULTIFACTORIALES	182
10.2.1. Concepto de Plan Factorial Equilibrado	182
10.2.2. Ejemplo de un Plan Factorial Equilibrado con 2 Factores	183
10.2.3. Predicciones en Análisis de la Varianza	187
10.2.4. Ejemplo: mejora en un proceso de aditivación en PEAD	187
10.2.5. Ejemplo: Mejora de un proceso de adhesivado	193
10.A. EJERCICIOS ADICIONALES	199

CAPÍTULO 11. ESTADÍSTICA DESCRIPTIVA BIDIMENSIONAL

11.1. DISTRIBUCIONES MARGINALES Y CONDICIONALES	201
11.2. DIAGRAMAS DE DISPERSIÓN	203
11.3. COVARIANZA. COEFICIENTE DE CORRELACIÓN	204
11.4. INTERPRETACION DE RELACIONES	206
11.5. MODELOS DE REGRESIÓN	207

CAPÍTULO 12. MODELOS DE REGRESIÓN

12.1. INTRODUCCIÓN	209
12.2. MODELOS DE REGRESIÓN: IDEAS GENERALES	210
12.3. FASES DE UN ESTUDIO DE REGRESIÓN	211
12.4. MODELO DE REGRESIÓN LINEAL SIMPLE	215
12.5. GENERALIZACIONES DEL MODELO	217
12.5.1. Modelo de regresión lineal múltiple	217
12.5.2. Consideración de relaciones no lineales	218
12.5.3. Inclusión de variables cualitativas	220
12.5.4. Consideración de interacciones con variables cualitativas	222
12.6. ESTIMACIÓN DEL MODELO	224
12.6.1. Objetivos	224
12.6.2. Fundamento del proceso de estimación	224
12.7. COEFICIENTE R ² . ANOVA DEL MODELO	225
12.8. TESTS DE HIPÓTESIS SOBRE LAS β_i	226

12.9. PREDICCIONES EN MODELOS DE REGRESIÓN	227
12.10. VALIDACIÓN DEL MODELO. ANÁLISIS DE RESIDUOS	228
12.11. EJEMPLO DE SÍNTESIS: UN MODELO PARA EL CONTROL DEL CONSUMO DE ENERGÍA	229
12.11.1. Objetivo del estudio	229
12.11.2. Modelo inicial	230
12.11.3. Recogida de datos	231
12.11.4. Estimación del modelo inicial	232
12.11.5. Relación funcional entre E(Consumo) y Temperatura	232
12.11.6. Consideración del efecto del día de la semana	236
12.11.7. Efectos sobre la varianza del Consumo	238
12.11.8. Explotación del modelo	238
12.A. AUTOEVALUACIONES RESUELTAS Y EJERCICIOS	239
12.A.1. Respuesta a algunas Autoevaluaciones	239
12.A.2. Ejercicios resueltos	242
12.A.3. Ejercicios adicionales	243

CAPÍTULO 1

INTRODUCCIÓN

1.1. OBJETO DE LA ESTADÍSTICA

La Estadística constituye una disciplina científica extremadamente amplia y que puede ser conceptualizada desde enfoques diferentes e incluso contrapuestos. No es raro, por tanto, que se hayan propugnado para ella distintas definiciones que, en el fondo, implican diferentes visiones sobre lo que constituye la característica esencial de esta ciencia.

Desde nuestro punto de vista, en un texto cuyo objetivo es la enseñanza de la Estadística a futuros ingenieros, una definición adecuada, que sintetiza las propuestas entre otros por Gnanadesikan y Bisgaard, podría ser la siguiente:

"La Estadística es la ciencia cuyo objeto es la obtención y el análisis de datos mediante el recurso a modelos matemáticos y a herramientas informáticas"

En nuestra opinión es precisamente la palabra "**datos**" la clave en la definición de la Ciencia Estadística. Como afirma Joiner *"el foco de nuestra ciencia son los **datos**, no la variación aleatoria ni la probabilidad"*.

La Estadística no es una rama de las Matemáticas, aunque el recurso al lenguaje de las Matemáticas sea fundamental en ella, de la misma forma que la Mecánica o la Termodinámica tampoco son una rama de las Matemáticas pese al extensivo uso que en las mismas se hace de modelos matemáticos. Es un error considerar que la importancia real de las técnicas estadísticas es proporcional a su complejidad matemática. De hecho métodos gráficos y otras herramientas de análisis descriptivo son extremadamente útiles para la interpretación de datos reales, pese a la sencillez de su aparato matemático.

La definición que hemos avanzado también resalta el papel crucial que en la Estadística aplicada moderna desempeñan las herramientas informáticas. De hecho sin la existencia del software adecuado la mayor parte de la metodología estadística que utilizan los ingenieros (modelos de regresión múltiple, análisis multivariante, series temporales, análisis de fiabilidad, etc...) sería inaplicable. El recurso a paquetes estadísticos es imprescindible tanto para la estimación como para la validación de la mayoría de los modelos estadísticos avanzados. No es concebible, en nuestra opinión, que pueda explicarse una Estadística útil a futuros ingenieros sin recurrir a estas herramientas informáticas.

1.2. POBLACIONES

En la terminología estadística se denomina población al conjunto de todos los individuos o entes que constituyen el objeto de un determinado estudio y sobre los que se desea obtener ciertas conclusiones.

Ejemplo 1: en un estudio sobre la intención de voto de los ciudadanos españoles, la población la constituirá el conjunto de los aproximadamente 35 millones de españoles con derecho a voto.

Ejemplo 2: en un estudio sobre el desarrollo de la tristeza de los cítricos en la Comunidad Valenciana, la población estará formada por la totalidad de árboles de cítricos existentes en esta Comunidad.

Ejemplo 3: al realizar en una industria el control de calidad en recepción de una partida de piezas, la población estará constituida por la totalidad de las piezas que componen la partida.

Los ejemplos anteriores tratan en todos los casos de poblaciones con una existencia física real, constituidas por un número finito, aunque posiblemente muy elevado, de individuos.

Aunque pueda parecer sorprendente no es ésta la situación más frecuente en la práctica, sino que en general las poblaciones a estudiar son de carácter abstracto, fruto del necesario proceso de conceptualización que debe preceder al estudio científico de cualquier problema real.

Ejemplo 4: Un ejemplo trivial sacado de los juegos de azar sirve para ilustrar la idea anterior. Se desea estudiar si un dado es correcto o está trucado. ¿Qué quiere decir la afirmación de que el dado es correcto? En la práctica, que si se tira un número muy elevado de veces los seis resultados posibles saldrán aproximadamente con la misma frecuencia. Al abordar este problema nos referiremos a la población abstracta constituida por infinitos lanzamientos del dado en cuestión, población sobre la que deseamos estudiar si las frecuencias relativas con las que se presentan los seis resultados posibles son idénticas.

Ejemplo 5: En una estudio sobre el contenido de agua en los tanques de acrilonitrilo fabricados por una empresa, la población sobre la que interesa obtener resultados podrían constituirlos todos los tanques que pueda fabricar en el futuro la empresa.

Ejemplo 6: En un estudio sobre la eficiencia de diversos algoritmos de encaminamiento de mensajes entre nudos en una red de procesadores, la población a investigar la constituirían todos los mensajes que puedan llegar a generarse en la red.

Como se desprende de los ejemplos anteriores los "individuos" que forman una población pueden corresponder a entes de naturaleza muy diversa (personas, árboles, piezas, lanzamientos de dados, parcelas, mensajes, etcétera...). En los casos de los tres primeros ejemplos dichos individuos tienen una existencia real, previa a la realización del estudio. En casos como los de los ejemplos 4 5, y 6 los individuos que constituyen la población pueden irse generando mediante la realización de un determinado proceso (lanzar un dado, fabrica un tanque de acrilonitrilo, emitir un mensaje desde un nudo,...). A estos procesos, que en sucesivas realizaciones pueden ir generando los diferentes individuos de la población les denominamos experimentos aleatorios

1.3. VARIABLES ALEATORIAS

1.3.1. Concepto

¡En toda población real existe VARIABILIDAD! Unos españoles votan a ciertos partidos y otros a otros; unos naranjos tienen tristeza y otros no; una determinada dimensión varía algo de una pieza a otra; el número que sale al lanzar el dado varía de unas tiradas a otras; el rendimiento obtenido varía de unas parcelas a otras; unos mensajes tienen retardos más elevados que otros,...

A cualquier característica que puede constatarse en cada individuo de una población se le denomina característica aleatoria. Así el partido a que piensan votar los individuos (Ejemplo 1) la ausencia o presencia de tristeza en los árboles (Ejemplo 2), el contenido de agua en los tanques (Ejemplo 5) o el retardo de un mensaje (Ejemplo 6) son características aleatorias. Muchas características aleatorias se expresan numéricamente (como el número de puntos obtenidos al lanzar un dado, el contenido de agua en un tanque o el retardo de un mensaje); a este tipo de características aleatorias se las denomina variables aleatorias.

Cuando una característica aleatoria es de tipo cualitativo (como por ejemplo el partido político a votar) nada impide codificar numéricamente sus diferentes alternativas y tratarla como una variable aleatoria. Hay que tener, sin embargo, cuidado en estos casos, porque operaciones perfectamente legítimas con características intrínsecamente numéricas (como, por ejemplo, sumar y promediar los rendimientos de diferentes parcelas) carecerían de sentido en este caso.

Autoevaluación: ¿Qué sentido práctico tendría el resultado de sumar y promediar los códigos de los problemas considerados más importantes por un conjunto de individuos?

1.3.2. Variables discretas y variables continuas

Cuando el conjunto de los valores que podría tomar una determinada variable aleatoria es discreto (es decir, finito o infinito numerable) se dice que dicha variable es de tipo discreto (a veces a las variables de este tipo se les denomina también atributos), por oposición a aquellos casos en que dicho conjunto es un infinito continuo en los que la variable se denomina continua.

Ejemplos de variables discretas serían el número de puntos al lanzar un dado, el número de picadas de ceratitis en cada naranja de un huerto, el número de errores en un programa de ordenador y también cualquier variable que se origine al codificar las diferentes alternativas de una característica cualitativa (sexo, partido votado, etcétera..).

Ejemplos de variables continuas serían todas las características que se miden sobre una escala de naturaleza básicamente continua (estaturas, pesos, rendimientos, tiempos, resistencias, etcétera...)

1.3.3. Variables k-dimensionales

Cuando sobre cada individuo de la población se estudian K características diferentes (todas ellas expresables numéricamente) se tiene una variable aleatoria K -dimensional. Por ejemplo si en la población constituida por los estudiantes de la UPV se estudia el sexo (codificado como 1 ó 2), la edad, la estatura y el peso, estaremos ante una variable aleatoria de dimensión 4.

En estos casos es frecuente utilizar los valores de aquellas componentes cuya naturaleza intrínseca es cualitativa (por ejemplo, el sexo) para dividir la población inicial en subpoblaciones (en nuestro caso: chicos y chicas) entre las cuales interesa estudiar las diferencias en las pautas de variabilidad existentes en las otras componentes de la variable aleatoria (ejemplo: ¿cómo difieren las pautas de variabilidad del peso o la estatura entre chicos y chicas en la UPV?).

Es importante darse cuenta de la diferencia entre una variable aleatoria K -dimensional, en la que las K variables se miden sobre los individuos de una única población, y un conjunto de K variables aleatorias unidimensionales, definidas sobre K poblaciones distintas.

Autoevaluación: El contenido en zumo y el calibre de las naranjas de un huerto ¿constituyen una variable aleatoria bidimensional? ¿Y el número de líneas de código y el número de errores en los programas preparados por una empresa de software? ¿Y el contenido de leucocitos en la sangre de individuos alcohólicos y no alcohólicos? ¿Y las estaturas del marido y de la mujer en los matrimonios jóvenes de un país? (Ver respuesta en el Anejo al final del Tema)

Autoevaluación: La definición clara de la población sobre la que se desea obtener conclusiones es el primer paso de cualquier estudio. El alumno deberá plantearse 3 problemas que le interesen de su vida cotidiana y definir en cada caso, con la mayor precisión posible, la población implicada y la(s) variable(s) aleatoria(s) implicada(s), analizando su naturaleza discreta o continua.

Autoevaluación: En el estudio de insecticidas se define la LD50 (Dosis Letal 50) de un producto como aquella dosis mínima que administrada a ratas provoca la muerte al 50% de las mismas. Al estudiar la LD50 de un determinado producto: ¿Cuál es la población implicada, y cual la variable aleatoria considerada? (Ver respuesta en el Anejo al final del Tema)

Autoevaluación: En una factoría interesa cuantificar, con el fin de controlar el consumo de energía (utilizada en su mayor parte en la climatización de las naves), la relación existente entre el consumo diario de electricidad y la temperatura media del día correspondiente. ¿Cuál es en el contexto anterior la población implicada y la variable aleatoria considerada? (Ver respuesta en el Anejo al final del Tema)

1.4. MUESTRAS. DATOS ESTADÍSTICOS

En general no resulta posible estudiar la totalidad de los individuos de una población para obtener información sobre ésta. Incluso cuando esta posibilidad existe técnicamente, como es el caso al tratar de poblaciones reales finitas, dicho procedimiento suele ser impracticable por consideraciones económicas.

En consecuencia para obtener información sobre una población hay que limitarse a analizar sólo un subconjunto de individuos de la misma. A éste subconjunto se le denomina muestra.

La forma de seleccionar los individuos que han de constituir la muestra tiene, como es lógico, una importancia capital para garantizar que ésta permita obtener conclusiones que puedan extrapolarse validamente a la población de la que la muestra procede. No hay que olvidar nunca que el objeto final del estudio es siempre la población y que la muestra es sólo un medio para obtener información sobre ésta.

Con el fin de permitir inferir conclusiones válidas sobre una población la muestra debe ser "representativa" de ésta. En teoría la única forma de garantizar la representatividad de una muestra es seleccionando al azar los individuos que la van a componer, de forma que todos los individuos de la población tengan "a priori" una probabilidad idéntica de pertenecer a la muestra. Aunque ésta forma de proceder rara vez sea aplicable de forma estricta en la práctica, siempre hay que extremar las precauciones para que la forma real de obtener la muestra sea lo más parecida posible a la ideal.

Para seguir leyendo haga click aquí