APUNTES DE
PROBABILIDAD Y ESTADÍSTICA

Hugo E. Borrás García
Rafael Iriarte Balderrama
Bernardo Fontana de la Cruz

DIVISIÓN DE CIENCIAS BÁSICAS
DEPARTAMENTO DE MATEMÁTICAS APLICADAS

1985

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO
FACULTAD DE INGENIERIA

APUNTES DE PROBABILIDAD Y ESTADISTICA

HUGO E. BORRAS GARCIA
RAFAEL IRIARTE BALDERRAMA
BERNARDO FRONTANA DE LA CRUZ

DIVISION DE CIENCIAS BASICAS
DEPARTAMENTO DE MATEMATICAS APLICADAS

FI/DCB/85-025

APUNTES DE PROBABILIDAD Y ESTADISTICA

PROLOGO

En los últimos años la teoría de la probabilidad y estadística ha llegado a ser una valiosa herramienta dentro del campo de la ingeniería y en general de todas las ciencias, debido a que con ella se pueden entender y manejar fenómenos que no podrían ser estudiados de manera determinística. Asimismo es un importante apoyo en la toma de decisiones, en la planeación de proyectos, en la simulación de sistemas y en muchos otros campos del conocimiento.

En los planes de estudio de todas las carreras que ofrece la Facultad de Ingeniería de la U.N.A.M., se ha incluido un curso de probabilidad y estadística que imparte la División de Ciencias Básicas.

Este curso pretende que el estudiante adquiera los elementos teóricos y prácticos que le permitan analizar fenómenos aleatorios relacionados con la ingeniería.

La presente obra tiene como finalidad apoyar a los estudiantes que cursan la materia y, por ello, se ha tratado de cubrir el contenido temático comprendido en el programa vigente de la asignatura.

En los primeros cinco capítulos se presentan los conceptos fundamentales de la teoría de la probabilidad, en los dos siguientes, los relativos a estadística y, por último, en el capítulo VIII se trata de dar un panorama introductorio a los temas de regresión y correlación, presentando solamente el caso de correlación lineal.

Sin embargo, consideramos que no se presentan en forma exhaustiva todos los temas del curso, por lo cual sugerimos consultar la bibliografía, a fin de que el lector pueda ampliar y profundizar en aquellos tópicos que sean de su interés.

Al final de la obra aparece un apéndice que contiene las tablas de las principales funciones de probabilidad, las cuales son necesarias para la solución de problemas.

El mejoramiento de estos apuntes podrá lograrse con ayuda de las críticas y sugerencias de profesores y alumnos, por lo que agradeceremos las aportaciones que se hagan llegar al Departamento de Matemáticas Aplicadas, con el objeto de incluirlas en futuras ediciones.

Agradecemos la valiosa ayuda que nos brindó la profesora Rossynela Durán Cuevas en la elaboración del primer manuscrito de este material, así como a las licenciadas María Cuairán Ruidíaz e Irma Hinojosa Félix por su participación en el análisis del contenido temático y estructuración didáctica de esta obra.

ING. HUGO E. BORRAS GARCIA
ING. RAFAEL IRIARTE BALDERRAMA
ING. BERNARDO FRONTANA DE LA CRUZ

CONTENIDO

CAPITULO I PROBABILIDAD

INTRODUCCION

El término probabilidad está ligado al grado de incertidumbre que existe sobre la ocurrencia de un acontecimiento. De tal manera que cuando se habla de una probabilidad pequeña se asocia la idea con una gran incertidumbre y cuando se habla de un acontecimiento con probabilidad 1 (ó 100 %), seguramente ocurrirá el acontecimiento y desaparecerá la incertidumbre. Como se puede notar la probabilidad no es más que un número que se asigna al acontecimiento y para hacer la asignación existen varias modalidades.

La teoría de la probabilidad es una rama de las matemáticas y por tal motivo la asignación de probabilidades se encuentra fundamentada en tres axiomas básicos. En este capítulo se presentan además de la definición de algunos conceptos, las principales maneras de asignar probabilidades, los axiomas básicos y algunos teoremas derivados de los axiomas.

I.1 CONCEPTOS GENERALES

La experimentación es una etapa fundamental para llegar al conocimiento científico. En probabilidad la palabra experimento incluye tanto la ocurrencia de un hecho como la observación del resultado. Los experimentos se pueden clasificar, de acuerdo con sus resultados, en dos tipos que son:

EXPERIMENTO DETERMINISTICO

Cuando es posible predecir con certeza su resultado, antes de realizarlo.

EXPERIMENTO ALEATORIO

Cuando no se puede conocer con certeza el resultado antes de ejecutar el experimento o bien, cuando se obtienen resultados diferentes al efectuar el experimento en condiciones aparentemente iguales.

Si un experimento consiste en registrar el número de llamadas telefónicas que llegan a un conmutador en cierto intervalo de tiempo, se trata de un experimento aleatorio. Sin embargo, puede considerarse determinista si el conmutador se encuentra desconectado, y en tal caso se puede asegurar con certeza que no llegará ninguna llamada en el intervalo de tiempo considerado.

Otro experimento aleatorio puede ser, el observar si mañana se descubre la vacuna contra el cáncer. En este ejemplo el experimento se puede repetir, pero solamente hasta el día en que se descubra; después de lo cual no tiene sentido repetir el experimento.

ESPACIO MUESTRAL

Es el conjunto de todos los posibles resultados de un experimento aleatorio y generalmente se representa con las letras S u Ω.

A cada uno de los posibles resultados se le llama punto del espacio muestral o evento elemental.

Ejemplo I.1

Si un experimento aleatorio consiste en lanzar un dado no cargado y su resultado es observar el número de puntos que aparecen en la cara superior del dado, el espacio muestral será el conjunto

$$S = \{1, 2, 3, 4, 5, 6\}$$

y cada uno de estos números es un punto del espacio muestral o evento elemental.

Ejemplo I.2

Si se considera el crecimiento de un árbol como un experimento aleatorio y el resultado es el tamaño que alcanza después de un año de haberse plantado la semilla, entonces el espacio muestral será el conjunto de los números reales positivos, debido a que el árbol no puede presentar un tamaño negativo y no se conoce un valor máximo de crecimiento.

EVENTOS

En general, un evento es una colección o subconjunto de resultados contenidos en el espacio muestral. Cuando el evento está formado por un solo punto del espacio muestral se le llama *evento elemental o simple* y cuando se forma por dos o más puntos se le llama *evento compuesto*. Así en el ejemplo I.1, los eventos simples serán cada uno de los puntos 1, 2, ..., 6. Ahora, si lo que interesa es observar como resultado un número par, se tendrá entonces un evento compuesto:

$$A = \{2, 4, 6\}$$

donde:

$$A \subset S$$

Un evento se denomina *seguro o cierto*, si se produce obligatoriamente como resultado del experimento en cuestión. Se llama *imposible* al evento que no puede suceder como resultado del experimento dado y se denomina *aleatorio* al evento que puede o no ocurrir.

Serán eventos *igualmente posibles*, cuando en la realización de un experimento no existen elementos que favorezcan el resultado de algún o algunos de ellos.

Dados dos eventos A y B cualesquiera, tales que la ocurrencia de uno implica que no puede ocurrir el otro, esto es, $A \cap B = \phi$ entonces A y B son eventos *mutuamente excluyentes*.

Finalmente, a los eventos $A_1, A_2, \ldots, A_n$ se les llama *colectivamente exhaustivos* cuando su unión forma todo el espacio muestral, es decir:

$$S = (A_1 \cup A_2 \cup \ldots \cup A_n)$$

Ejemplo I.3

Un fabricante de acumuladores industriales sabe que está por recibir pedidos de dos clientes C_1 y C_2, dichos pedidos pueden ser de 1 a 5 unidades.

El espacio muestral correspondiente estará dado por el conjunto de eventos S, que se muestra en la siguiente figura.

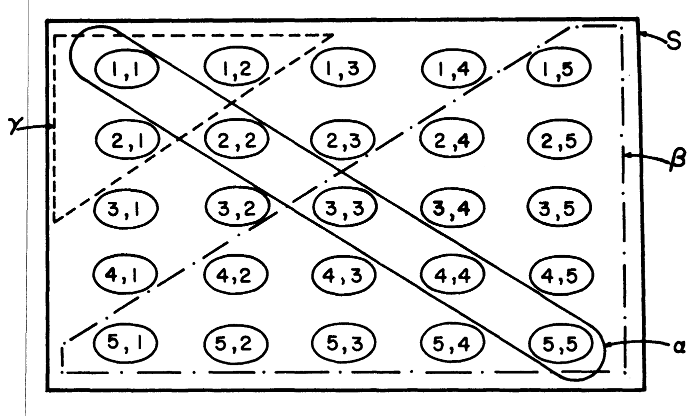


Figura I.1

En donde (i, j) denota al evento simple:

{C_1 demanda i acumuladores y C_2 demanda j acumuladores}

entonces se tendrán 25 eventos simples, de los cuales se pueden determinar eventos compuestos tales como:

α = {que ambos clientes demanden el mismo número de unidades}

este evento corresponde a los puntos en los cuales $i = j$

β = {que la suma de las demandas esté entre 6 y 10 unidades}

es decir que $\{6 \leq i + j \leq 10\}$

γ = {que la suma de las demandas sea a lo más de 3 unidades}

o bien $\{i + j \leq 3\}$

El evento seguro estaría formado por todo el espacio muestral S y un evento imposible sería, por ejemplo, que ambos demanden radiadores, lo cual no puede ocurrir en este experimento.

En este ejemplo, los eventos simples son excluyentes, ya que la ocurrencia de cualquiera de los 25 puntos excluye la posibilidad de ocurrencia de cualquier otro. Sin embargo los eventos compuestos α y β ó α y γ no son excluyentes, pues existen puntos en común.

Los eventos elementales también son colectivamente exhaustivos, debido a que su unión forma todo el espacio muestral.

I.2 INTERPRETACIONES DEL CONCEPTO DE PROBABILIDAD

CLASICA O DE LAPLACE (FINES DEL SIGLO XVII)

Históricamente, la primera aplicación de la teoría de la probabilidad se hizo en los juegos de azar. Los experimentos aleatorios asociados con juegos de azar dieron lugar a espacios muestrales con un número finito de puntos.

De acuerdo con esta interpretación, si se desea asignar probabilidades sobre un espacio muestral de N puntos, la probabilidad de un evento A es igual al cociente del número de puntos muestrales del evento (Na) sobre el número total de puntos muestrales (N), es decir:

$$\boxed{P(A) = \frac{Na}{N}} \qquad \ldots (1)$$

Ejemplo I.4

Considerando que los eventos elementales del ejemplo I.3, tienen la misma probabilidad de ocurrir, calcular la probabilidad de los eventos α, β y γ .

Solución

Los puntos muestrales del evento α son 5 y el total es 25, entonces de la expresión (1):

$$P(\alpha) = \frac{5}{25}$$

de la misma forma:

$$P(\beta) = \frac{15}{25}$$

y

$$P(\gamma) = \frac{3}{25} \quad \text{(ver figura I.1)}$$

LIMITACIONES DE LA INTERPRETACION CLASICA

a) Es necesario que cada resultado del experimento tenga la misma posibilidad de ocurrir.

b) En algunos experimentos el número total de resultados es demasiado grande o muy difícil de determinar, como por ejemplo, cuando se desea calcular la probabilidad de que mañana cierta subestación reciba una descarga eléctrica, durante una tormenta.

FRECUENTISTA O DE VON MISES (1957)

Examínese una sucesión de n experimentos iguales. Supóngase que, como resultado de cada experimento, se registra la llegada de una señal (evento A) o la no llegada (evento A') en intervalos iguales de tiempo.

En esta sucesión, una característica del evento A es la frecuencia de su realización, es decir, la relación entre el número de veces que este evento se produce y el número total de experimentos realizados. Si el evento A ocurre x veces en n experimentos, entonces su probabilidad $P(A)$ se define como el límite de la frecuencia relativa:

$$f^* = \frac{x}{n}$$

cuando el número de experimentos tiende a infinito, es decir:

$$\boxed{P(A) = \lim_{n \to \infty} \frac{x}{n} \text{ o bien } \frac{x}{n} \to P(A)} \qquad \ldots \quad (2)$$

Ejemplo I.5

De 6700 alumnos que se inscribieron al curso de probabilidad y estadística en semestres anteriores, 1500 no lo terminaron, 2000 obtuvieron una calificación de no acreditada y el resto lo aprobaron.

Si se selecciona un alumno al azar, entre los que cursan actualmente la materia y con base en la información anterior, cuál es la probabilidad de que dicho alumno:

a) No termine el curso.

b) Acredite la materia.

Solución

a) Si se considera el evento:

$A = \{\text{No termine el curso}\}$

entonces $x = 1500$ y $n = 6700$, con lo cual de la expresión (1):

$$P(A) \approx \frac{1500}{6700} = 0.224$$

b) De igual manera:

$B = \{\text{Acredite la materia}\}$

entonces:

$$x = 6700 - 1500 - 2000 = 3200$$

de donde:

$$P(B) \approx \frac{3200}{6700} = 0.478$$

LIMITACIONES DE LA INTERPRETACION FRECUENTISTA

No se puede aplicar cuando el experimento aleatorio no es repetible o bien cuando es repetible pero cambian las condiciones del experimento.

SUBJETIVISTA (1969)

La escuela subjetivista considera que la probabilidad es una medida del grado de incertidumbre que tiene una persona o un grupo de personas respecto a la verdad de una afirmación o a la ocurrencia de un hecho.

Por ejemplo, cuando un analista requiere determinar la probabilidad de que baje el precio del petróleo, más de dos dólarès por barril en este año, puede utilizar la interpretación subjetivista para estimar la probabilidad de este evento, tomando en cuenta su experiencia y tal vez la de otras personas que conozcan sobre el tema. De otra forma no se podría obtener esta probabilidad debido a que el evento no es repetible.

LIMITACIONES DE LA INTERPRETACION SUBJETIVISTA

Esta interpretación tiene el inconveniente de que la probabilidad asignada cambie de una persona a otra y en ocasiones puede presentar inconsistencias en una misma persona cuando ésta aumente sus conocimientos sobre el fenómeno en cuestión.

I.3 AXIOMAS DE PROBABILIDAD

Para un espacio muestral S, la probabilidad de cada evento $A \subset S$ es un número que se asigna al evento y se representa por $P(A)$, este número debe satisfacer las siguientes condiciones:

Axioma 1

Para cualquier evento A:

$$P(A) \geq 0$$

Axioma 2

Si S es el evento formado por todo el espacio muestral:

$$P(S) = 1$$

Axioma 3

Si A y B son eventos mutuamente excluyentes en S, entonces:

$$P(A \cup B) = P(A) + P(B) \iff A \cap B = \emptyset$$

Ejemplo I.6

En un laboratorio de computación se encuentran 15 calculadoras, de las cuales cinco están descompuestas. Si una persona toma al azar tres de ellas, ¿cuál es la probabilidad de que por lo menos una de las tres calculadoras esté descompuesta?

Solución

La condición de *por lo menos una de las tres calculadoras esté descompuesta* (evento A), se cumplirá si ocurre cualesquiera de los tres eventos mutuamente excluyentes:

1D = {una esté descompuesta y dos no}

2D = {dos estén descompuestas y una no}

3D = {las tres estén descompuestas}

El evento A se puede representar como la unión de estos eventos:

$$A = \{1D \cup 2D \cup 3D\}$$

por el axioma 3:

$$P(A) = P(1D) + P(2D) + P(3D)$$

donde P(1D) se calcula tomando en cuenta que de cinco descompuestas se selecciona una, y de las 10 no descompuestas se seleccionan dos, de manera que hay

$$\binom{10}{2} = \frac{10!}{2!\ 8!} = 45$$

formas de hacerlo; por lo tanto existen 5(45) = 225 maneras en las que se puede presentar 1D.

Por otra parte, el número de casos totales es:

$$\binom{15}{3} = \frac{15!}{3!\ 12!} = 455$$

entonces:

$$P(1D) = \frac{225}{455} = 0.495$$

de la misma forma se obtiene:

$$P(2D) = \frac{\binom{5}{2}\binom{10}{1}}{455} = 0.22$$

$$P(3D) = \frac{\binom{5}{3}}{455} = 0.022$$

por lo tanto:

$$P(A) = 0.495 + 0.22 + 0.022 = 0.737$$

Algunos de los teoremas más importantes que se derivan de los axiomas de probabilidad son:

Teorema I.1

Sea el evento A un evento cualquiera de S y A' el complemento de A, entonces:

$$P(A') = 1 - P(A)$$

Demostración:

Por el axioma 2:

$P(S) = 1$

Como:

$$(A \cup A') = S$$

entonces:

$$P(A \cup A') = 1$$

y por el axioma 3:

$$P(A) + P(A') = 1$$

$$\therefore\ P(A') = 1-P(A)$$

Teorema I.2

Sea ϕ el evento imposible, entonces:

$$P(\phi) = 0$$

Demostración:

Haciendo $S = S \cup \phi$

y tomando en cuenta que:

$$S \cap \phi = \phi$$

por el axioma 3:

$$P(S) = P(S) + P(\phi)$$

además, por el axioma 2:

$$P(S) = 1$$

$$\therefore\ P(\phi) = 0$$

Teorema I.3

Si A y B son dos eventos de S y $A \subset B$

$$P(A) \leq P(B)$$

Demostración:

El evento B de la figura I.2, puede representarse como la unión de los eventos mutuamente excluyentes A y $A' \cap B$, esto es $B = \{A \cup (A' \cap B)\}$.

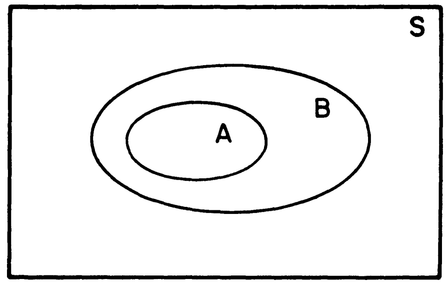


Figura I.2

del axioma 3:

$$P(B) = P(A) + P(A' \cap B)$$

y por el axioma 1:

$$P(A' \cap B) \geq 0$$

$$\therefore\ P(B) \geq P(A)$$

Teorema I.4

Si $A_1, A_2, \ldots, A_n$ es una colección de eventos mutuamente excluyentes y colectivamente exhaustivos, como se muestra en la figura I.3,

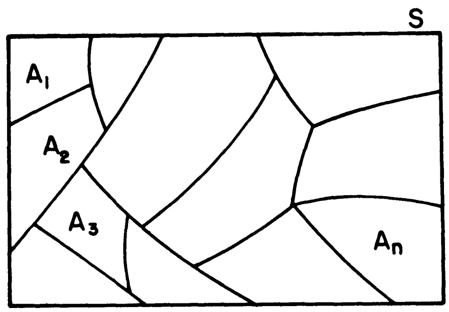


Figura I.3

entonces:

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$$

$$= \sum_{i=1}^{n} P(A_i) \quad \text{con } A_i \cap A_j = \phi \quad \text{para } i \neq j$$

este teorema es sólo una forma general del axioma 3.

Teorema I.5

Para dos eventos cualesquiera A y B:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Demostración:

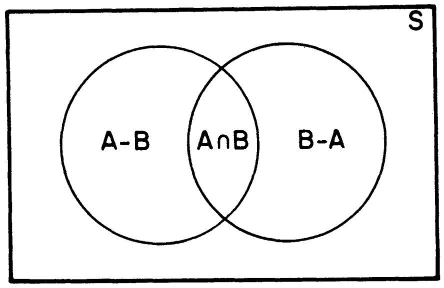


Figura I.4

De la figura I.4:

$$P(A \cup B) = P(A - B) + P(A \cap B) + P(B - A) \quad \ldots \text{(a)}$$

por otra parte:

$$P(A - B) = P(A) - P(A \cap B) \quad \ldots \text{(b)}$$

$$P(B - A) = P(B) - P(A \cap B) \quad \ldots \text{(c)}$$

sustituyendo (b) y (c) en (a):

$$P(A \cup B) = P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B)$$

$$\therefore\ P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Ejemplo I.7

En una compañía distribuidora de equipos electrónicos, se ha registrado el número de aparatos de tipo W que solicitan sus clientes semanalmente. Un resumen de dichos datos se muestra en la siguiente tabla:

N° DE APARATOS	FRECUENCIA (VECES)
0	2
1	5
2	9
3	4
4	1
5	1
	22

Cuál es la probabilidad de que en la próxima semana se soliciten:

a) Más de un aparato.

b) A lo más tres aparatos.

c) Entre 2 y 4, ó más de 2 aparatos.

Solución

a) Se pueden definir los siguientes eventos:

A = {solicitan más de un aparato}

C_i = {solicitan i aparatos}

entonces:

$A = \{C_2 \cup C_3 \cup C_4 \cup C_5\}$

por el teorema 1:

$P(A) = 1 - P(A')$

donde:

$A' = \{C_0 \cup C_1\}$

y por el axioma 3:

$$P(A') = P(C_0) + P(C_1) = \frac{2}{22} + \frac{5}{22} = \frac{7}{22}$$

$$\therefore\ P(A) = 1 - \frac{7}{22} = \frac{15}{22}$$

b) Sea el evento:

B = {solicitan a lo más 3 aparatos}

con lo cual:

$B = \{C_0 \cup C_1 \cup C_2 \cup C_3\}$

entonces:

$$P(B) = P(C_0) + P(C_1) + P(C_2) + P(C_3)$$

$$P(B) = \frac{2}{22} + \frac{5}{22} + \frac{9}{22} + \frac{4}{22} = \frac{20}{22}$$

Obsérvese que $A' \subset B$ y que cumple con el teorema I.3, es decir, $P(A') < P(B)$

c) Sean los eventos:

C = {solicitan entre 2 y 4 aparatos}

y:

D = {solicitan más de 2 aparatos}

de manera que:

$$C = \{C_2 \cup C_3 \cup C_4\} \text{ y } D = \{C_3 \cup C_4 \cup C_5\}$$

por el teorema I.4:

$P(C \cup D) = P(C) + P(D) - P(C \cap D)$

entonces, de acuerdo con los datos:

$P(C) = \frac{14}{22}$; $P(D) = \frac{6}{22}$ y $P(C \cap D) = \frac{5}{22}$

$\therefore \quad P(C \cup D) = \frac{14}{22} + \frac{6}{22} - \frac{5}{22} = \frac{15}{22}$

I.4 PROBABILIDAD CONDICIONAL

La probabilidad de que ocurra un evento A cuando se tiene la certeza que ha ocurrido otro evento B, denotado por P(A/B) (probabilidad de A dado B), se define como el cociente de la probabilidad de la intersección de A y B entre la probabilidad de que ocurra B.

$$\boxed{P(A/B) = \frac{P(A \cap B)}{P(B)}} \quad \ldots (3)$$

Como B es un evento que ya ocurrió, el espacio muestral S se reduce a B, como se muestra en la figura I.5, con lo cual:

$$P(A/B) = \frac{\text{Intersección } (A \cap B)}{\text{nuevo espacio muestral } (B)}$$

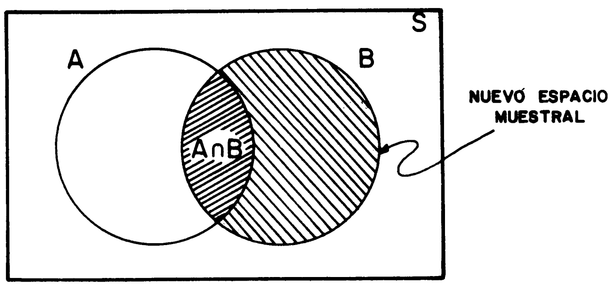


Figura I.5

De modo semejante se determina la probabilidad del evento (B/A).

Ejemplo I.8

Para presentar un examen extraordinario, un alumno considera que la probabilidad de que estudie es de 0.9.

El sinodal del examen ha establecido que si un alumno estudia tiene una probabilidad de 0.8 de aprobarlo; sin embargo si no estudia será sólo de 0.3.

Cuál es la probabilidad de que un alumno:

a) Estudie y apruebe el examen.

b) No estudie y apruebe.

Solución

Las diferentes alternativas del problema se pueden representar mediante un diagrama de árbol, como el que se muestra en la figura I.6, donde:

$E = \{\text{estudie}\}$

$A = \{\text{apruebe el examen}\}$

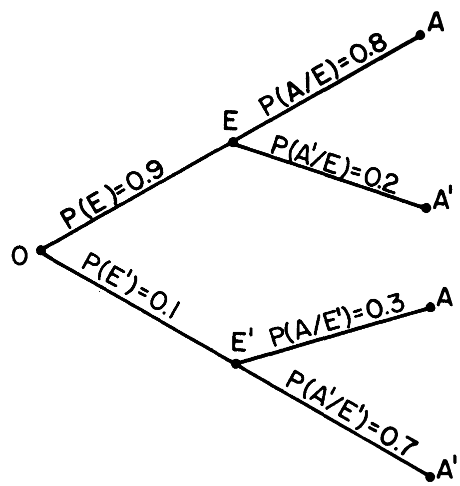


Figura I.6

De la definición de probabilidad condicional:

a) $P(E \cap A) = P(A/E)\ P(E)$

$= 0.8(0.9) = 0.72$

b) $P(E' \cap A) = P(A/E')\ P(E')$

$= 0.3(0.1) = 0.03$

EVENTOS INDEPENDIENTES

Dos eventos A y B son mutuamente independientes, cuando la ocurrencia de uno de ellos no afecta la probabilidad de ocurrencia del otro, esto es:

$$P(A/B) = P(A) \quad \text{ó} \quad P(B/A) = P(B)$$

Considerando los eventos:

$A = \{$sube el precio del petróleo en este año$\}$

$B = \{$se funden dos focos en un salón de clase$\}$

se puede observar que son independientes, ya que la probabilidad de que suba el precio del petróleo no tiene ninguna relación con lo que ocurre en el salón de clase y viceversa.

PROBABILIDAD DE LA INTERSECCION DE n EVENTOS

Frecuentemente se desea encontrar la probabilidad de que ocurran conjuntamente una serie de eventos $A_1, A_2, \ldots, A_n$; ya sea simultáneamente o bien, uno a continuación del otro.

En el caso de dos eventos, de la expresión (3):

$$P(A_1 \cap A_2) = P(A_1/A_2)\, P(A_2) \qquad \ldots (4)$$

para tres eventos se puede definir otro evento:

$$B = (A_2 \cap A_3)$$

y de la expresión (4):

$$\begin{aligned} P(A_1 \cap B) &= P(A_1 \cap A_2 \cap A_3) \\ &= P[A_1/(A_2 \cap A_3)]\, P(A_2 \cap A_3) \\ &= P(A_1/(A_2 \cap A_3))\, P(A_2/A_3)\, P(A_3) \end{aligned}$$

siguiendo un procedimiento similar para n eventos:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1/(A_2 \cap A_3 \cap \ldots \cap A_{n-1}))\, P(A_2/A_3 \cap \ldots \cap A_n) \ldots P(A_n) \qquad \ldots (5)$$

Cuando todos los eventos son independientes entre sí, $P(A_1/A_2) = P(A_1)$, entonces la expresión (4) se transforma en:

$$P(A_1 \cap A_2) = P(A_1)\, P(A_2)$$

y se puede demostrar por inducción matemática que:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)\, P(A_2) \ldots P(A_n) \qquad \ldots (6)$$

A esta última expresión se le conoce como la *regla de la multiplicación*.

Ejemplo I.9

Calcular la probabilidad de que se encuentren todas las piezas defectuosas en un lote de cinco unidades obtenidas de la producción de una máquina, considerando que es independiente la probabilidad de defecto en cada una de ellas y que en promedio la proporción de piezas defectuosas en esa máquina es del 10%.

Solución

Sea:

$$D_i = \{\text{la unidad } i \text{ resulta defectuosa}\}$$

de la expresión (6):

$$\begin{aligned} P(D_1 \cap D_2 \cap D_3 \cap D_4 \cap D_5) &= P(D_1)\ P(D_2)\ P(D_3)\ P(D_4)\ P(D_5) \\ &= (0.10)(0.10)(0.10)(0.10)(0.10) \\ &= (0.10)^5 \end{aligned}$$

TEOREMA DE LA PROBABILIDAD TOTAL

Sea un conjunto de eventos B_1, B_2, ..., B_n mutuamente excluyentes y colectivamente exhaustivos. Véase la siguiente figura.

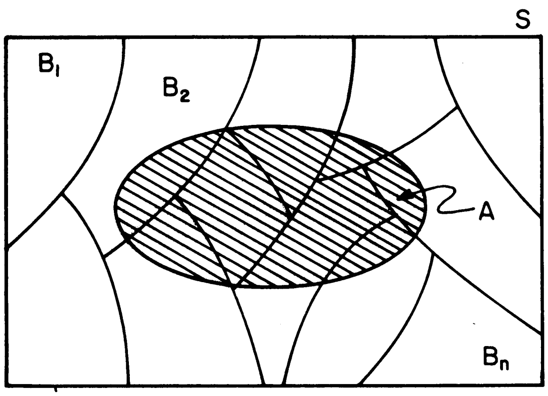


Figura I.7

Entonces para cualquier otro evento $A \subset S$; se tiene:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_n)$$

cuya probabilidad es:

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_n) = \sum_{i=1}^{n} P(A \cap B_i)$$

considerando que:

$$P(A \cap B_i) = P(A/B_i)\ P(B_i)$$

entonces:

$$P(A) = \sum_{i=1}^{n} P(A/B_i)\ P(B_i) \qquad \ldots (7)$$

Ejemplo I.10

Tres líneas aéreas (1, 2 y 3) hacen vuelos nocturnos de México a Acapulco. La experiencia ha demostrado que salen con retraso el 40% de los vuelos nocturnos de la línea 1, el 50% de la línea 2 y el 70% de la línea 3. Pasado mañana Carlos desea ir a Acapulco en un vuelo nocturno y elige aleatoriamente una línea aérea.

Cuál es la probabilidad de que Carlos:

a) Seleccione la línea 1 y que salga con retraso.

b) Salga con retraso.

Solución

El primer punto del problema involucra la selección de una de las 3 líneas por las cuales puede viajar. Si A_1, A_2, A_3 son los eventos que representan la elección de la línea 1, 2 y 3 respectivamente entonces $P(A_1) = P(A_2) = P(A_3) = 1/3$ (eventos igualmente posibles).

Una vez seleccionada la línea, se tendrán los eventos:

$$R = \{\text{el avión sale con retraso}\}$$

y:

$$R' = \{\text{sale a tiempo}\}$$

La información que se puede obtener del enunciado, determina las probabilidades:

$$P(R/A_1) = 0.4,\ P(R/A_2) = 0.5,\ P(R/A_3) = 0.70$$

El diagrama de árbol de la figura I.8 representa los eventos de este experimento.

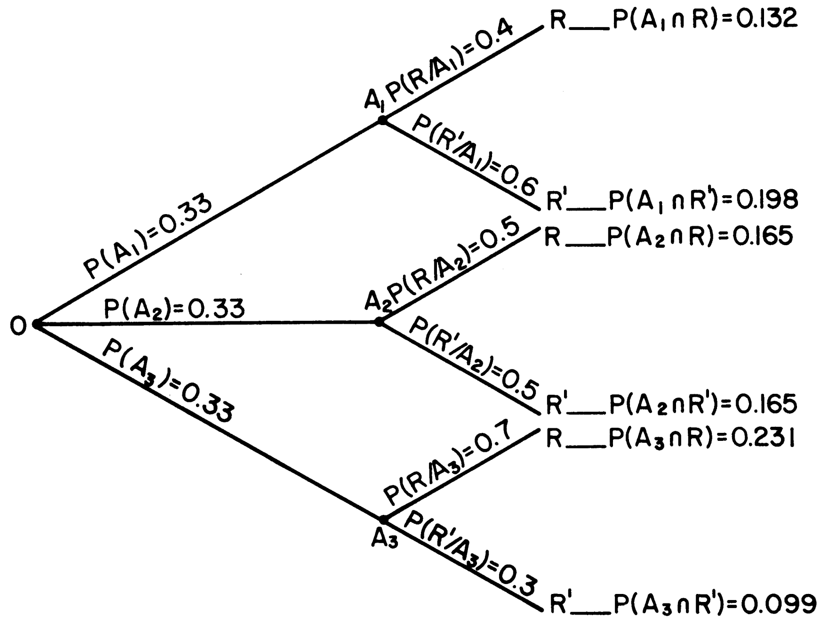


Figura I.8

a) De la expresión (4) se obtiene la probabilidad conjunta

$$P(A_1 \cap R) = P(R/A_1)\ P(A_1) = 0.4(0.33) = 0.132$$

b) El evento R se puede presentar con cada una de las tres líneas aéreas, por lo cual de la expresión (7):

$$P(R)=P(R/A_1)P(A_1)+P(R/A_2)P(A_2)+P(R/A_3)P(A_3)=P(A_1\cap R)+P(A_2\cap R)+P(A_3\cap R)$$

$$=0.132+0.165+0.231=0.528$$

TEOREMA DE BAYES

Este teorema es solamente una aplicación de la definición de probabilidad condicional. Sin embargo en ella se fundamenta la teoría bayesiana de la estimación, la cual permite modificar las probabilidades a partir de nueva información.

Sean $B_1, B_2, \ldots, B_k, \ldots, B_n$ eventos mutuamente excluyentes y colectivamente exhaustivos y sea A otro evento del mismo espacio muestral S, como se muestra en la figura I.7; entonces:

$$\boxed{P(B_k/A) = \frac{P(A/B_k)\, P(B_k)}{\sum_{k=1}^{n} P(A/B_k)\, P(B_k)}} \qquad \ldots (8)$$

Demostración:

Por la definición de probabilidad condicional:

$$P(B_k/A) = \frac{P(B_k \cap A)}{P(A)} \qquad \ldots (a)$$

de la expresión (4):

$$P(B_k \cap A) = P(A/B_k)\, P(B_k) \qquad \ldots (b)$$

y del teorema de la probabilidad total:

$$P(A) = \sum_{k=1}^{n} P(A/B_k)\, P(B_k) \qquad \ldots (c)$$

sustituyendo (b) y (c) en (a):

$$P(B_k/A) = \frac{P(A/B_k)\, P(B_k)}{\sum_{k=1}^{n} P(A/B_k)\, P(B_k)}$$

Ejemplo I.11

Considérese el enunciado del ejemplo I.10.

Si después de efectuado el viaje, Carlos comenta que el avión salió a tiempo, cuál es la probabilidad de que haya utilizado:

a) La línea 2.

b) La línea 3.

Solución

a) Utilizando la misma notación y de la expresión (8):

$$P(A_2/R') = \frac{P(R'/A_2)\ P(A_2)}{P(R'/A_1)P(A_1)+P(R'/A_2)P(A_2)+P(R'/A_3)P(A_3)}$$

$$= \frac{P(R' \cap A_2)}{P(R' \cap A_1)+P(R' \cap A_2)+P(R' \cap A_3)}$$

sustituyendo los valores de la figura I.8:

$$P(A_2/R') = \frac{0.165}{0.198+0.165+0.099} = \frac{0.165}{0.462} = 0.357$$

b) Considerando que el denominador representa la probabilidad de que salga a tiempo P(R'), de la expresión (7):

$$P(A_3/R') = \frac{P(R' \cap A_3)}{P(R')} = \frac{0.099}{0.462} = 0.214$$

Es importante observar que las probabilidades condicionales que se encuentran en las ramas de la figura I.8, se refieren a eventos que ocurrirán después del análisis y por ese motivo se les denomina probabilidades *a priori*.

Por otra parte, las probabilidades que se calculan con el teorema de Bayes se refieren al posible origen de un evento que ocurrió en el pasado y, en consecuencia, a esas probabilidades condicionales se les llama *a posteriori*.

CAPITULO II VARIABLES ALEATORIAS

INTRODUCCION

En este capítulo se define el concepto de variable aleatoria que representa en forma cuantitativa el resultado de un experimento aleatorio, así como diferentes funciones e indicadores que facilitan el manejo matemático de los posibles resultados de un experimento aleatorio.

La ventaja de manejar variables aleatorias en lugar de los eventos de un espacio muestral, se hace evidente cuando el resultado del experimento es cualitativo.

II.1 CONCEPTO DE VARIABLE ALEATORIA

Se define como una función real que asocia un elemento numérico con cada uno de los eventos elementales del espacio muestral. Al término variable aleatoria también se le conoce como variable posible, cantidad incierta o variable estocástica.

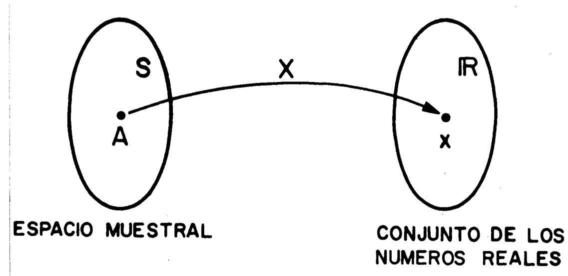


Figura II.1

En la figura anterior se observa la transformación $X:S \rightarrow \mathbb{R}$, donde al punto muestral A se le asigna el número x. La naturaleza del experimento no representa ningún problema para asignar valores numéricos a los elementos del espacio muestral; por ejemplo:

Si un experimento consiste en observar en un día determinado, si llueve o no, se puede hacer la asignación:

$$S = \{\text{no llueve, llueve}\} \longrightarrow X\ \{0,\ 1\}$$

Mediante la variable aleatoria X se asigna el número uno al evento llueve y cero cuando no llueve; sin embargo, ésta no es la única forma de hacer la asignación y se podrían asignar también los números $\{-1,1\}$ respectivamente.

Cuando el resultado del experimento aleatorio es un número, generalmente se le asigna ese mismo valor. Por ejemplo, en un experimento que consiste en observar la temperatura de una máquina de combustión interna, la variable aleatoria X puede tomar los mismos valores (x) que se observan en el termómetro, es decir que $X = x$.

II.2 VARIABLES ALEATORIAS DISCRETAS

Se dice que una variable aleatoria X es discreta, cuando el número de valores que puede tomar es finito o infinito numerable, es decir, cuando los resultados del experimento aleatorio al que está asociada la variable sólo se definen para algunos valores de X.

Por ejemplo:

- El número de automóviles que pasa por cierta avenida en un intervalo de tiempo.
- La demanda mensual de tractores en una distribuidora de maquinaria.
- El número de personas que visita anualmente una localidad.

En estos ejemplos no tiene sentido hablar de resultados fraccionarios y por consiguiente sólo se definen para valores enteros de X.

FUNCION MASA DE PROBABILIDAD

El comportamiento de una variable aleatoria discreta X puede estudiarse a través de su función masa de probabilidad (F.M.P.), la cual relaciona los resultados de un experimento aleatorio con su probabilidad de ocurrencia. Esta función se denota como $P_X(x)$, $P(X=x)$ ó $P(x)$ y se interpreta como la probabilidad de que la variable aleatoria X tome el valor particular x.

PROPIEDADES DE LA F.M.P.

$$P_X(x) \geq 0 \qquad \text{para toda } x \qquad \ldots (1)$$

$$\sum_{\forall x} P_X(x) = 1 \qquad \ldots (2)$$

Ejemplo II.1

Cierta fábrica de chocolates está realizando una promoción publicitaria que consiste en incluir en la envoltura de algunas barras de chocolate un cupón que da derecho al consumidor a recibir otra. La proporción de chocolates premiados es de 20%.

Si la variable aleatoria X representa el número de cupones que hay en cuatro barras seleccionadas al azar y en forma independiente, determinar la F.M.P. de X y a partir de ella la probabilidad de encontrar cuando menos dos barras con cupón.

Solución

Considerando el evento

$$B_i = \{\text{la barra i tiene cupón}\}$$

Para x = 0, ninguna barra tiene cupón, entonces:

$$P_X(0) = P(B_1' \cap B_2' \cap B_3' \cap B_4')$$

donde B_i' es el complemento de B_i.

Como la proporción de chocolates premiados es de 20%

$$P(B_i) = 0.2 \text{ y } P(B_i') = 1 - 0.2 = 0.8$$

por lo tanto:

$$P_X(0) = P(B_1') \, P(B_2') \; P(B_3') \; P(B_4')$$

$$= (0.8)\,(0.8)\,(0.8)\,(0.8)$$

$$= 0.4096$$

Para x = 1

existen C(4,1) = 4 maneras de obtener una barra premiada y tres sin premio. Cada una de las alternativas tiene una probabilidad.

P (una premiada y tres sin premio) $= 0.2\,(0.8)^3 = 0.1024$

con lo cual:

$$P_X(1) = 4(0.1024) = 0.4096$$

De manera similar se obtiene:

$$P_X(2) = 0.1536$$

$$P_X(3) = 0.0256$$

$$P_X(4) = 0.0016$$

Gráficamente se puede observar la F.M.P. en la siguiente figura.

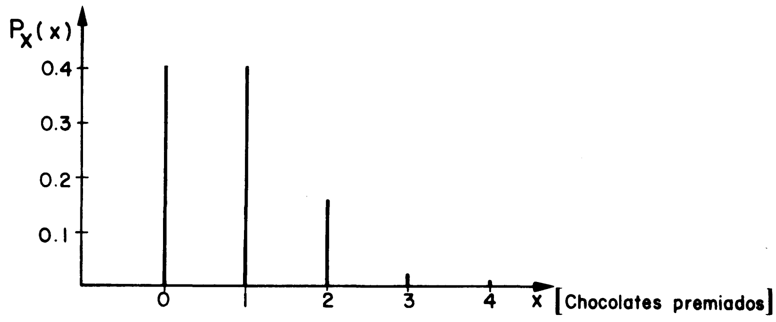


Figura II.2

Encontrar cuando menos dos barras equivale a encontrar dos, tres o cuatro, ya que la probabilidad de encontrar más de cuatro en este experimento es cero, con lo cual:

$$
\begin{aligned}
P(X \geq 2) &= P_X(2) + P_X(3) + P_X(4) \\
&= 0.154 + 0.025 + 0.002 \\
&= 0.181
\end{aligned}
$$

FUNCION DE DISTRIBUCION ACUMULADA PARA VARIABLES ALEATORIAS DISCRETAS

Otra forma de caracterizar el comportamiento de una variable aleatoria, es mediante la función de distribución acumulada (F.D.A.), la cual se denota como $F_X(x_0)$ y se define:

$$\boxed{F_X(x_0) = \sum_{\forall x < x_0} P_X(x)} \qquad \ldots (3)$$

En esta relación se puede observar que en cada valor $F_X(x_0)$ se acumulan las probabilidades anteriores o iguales a $P_X(x_0)$ y por eso la función lleva ese nombre.

PROPIEDADES DE LA F.D.A.

$$F_X(x_0) \geq 0 \quad \forall x_0 \qquad \ldots (4)$$

$$F_X(-\infty) = 0 \qquad \ldots (5)$$

$$F_X(\infty) = 1 \qquad \ldots (6)$$

$$F_X(x_0+\varepsilon) \geq F_X(x_0), \quad \text{para toda constante } \varepsilon > 0 \;. (7)$$

$$P(x_1 \leq X \leq x_2) = F_X(x_2) - F_X(x_1) \qquad \ldots (8)$$

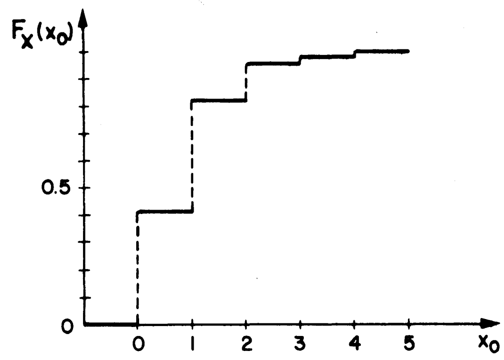
Ejemplo II.2

Determinar la F.D.A. del ejemplo II.1 en forma tabular y gráfica.

Solución

De la expresión (3):

X	$P_X(x)$	$F_X(x_0)$
0	0.4096	0.4096
1	0.4096	0.8192
2	0.1536	0.9728
3	0.0256	0.9984
4	0.0016	1.0000



II.3 VARIABLES ALEATORIAS CONTINUAS

Una variable aleatoria es continua si su dominio de definición es el conjunto de los números reales; es decir, cuando la variable puede tomar cualquier valor dentro de cierto intervalo.

Este tipo de variables se presenta con más frecuencia en los sistemas físicos que las variables discretas, debido a que la masa, el calor, el volumen, el tiempo y otras características, pueden tomar cualquier nivel, dentro de una escala continua.

FUNCION DENSIDAD DE PROBABILIDAD

Para estudiar el comportamiento de una variable aleatoria continua, se puede definir la función densidad de probabilidad (F.D.P.) de manera análoga a la F.M.P.; sin embargo, no es posible relacionar cada valor de la variable X con su probabilidad de ocurrencia, como sucede con la F.M.P. Una variable aleatoria continua puede tomar una infinidad de valores en un pequeño intervalo; en consecuencia, la probabilidad de que X tome un valor particular es prácticamente cero y para estudiar el comportamiento de la variable se debe hablar de subintervalos, dentro de los cuales puede encontrarse el valor de X.

La *densidad de probabilidad* es un concepto auxiliar que indica cuáles son los subintervalos con mayor o menor probabilidad, sin proporcionarla directamente.

La función densidad de probabilidad denotada por $f_X(x)$, se define de tal manera que la probabilidad del evento $(x_1 \leq X \leq x_2)$ es igual a la integral definida de $f_X(x)$ en el intervalo $[x_1, x_2]$, es decir:

$$\boxed{P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f_X(x)\, dx}$$

Por la interpretación geométrica de la integral, también se puede relacionar la probabilidad de que X se encuentre en el intervalo $[x_1, x_2]$ con el área bajo la F.D.P., como se muestra en la siguiente figura.

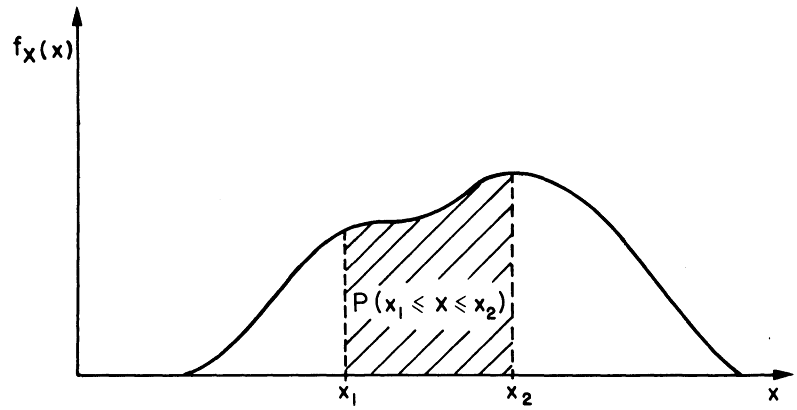


Figura II.3

Debe observarse que la probabilidad de que X tome un valor específico x, sería el área bajo un punto de la F.D.P., la cual es prácticamente cero. Sin embargo, esto no implica que sea imposible ya que el evento $(X=x)$ puede llegar a ocurrir.

PROPIEDADES DE LA F.D.P.

$$f_X(x) \geq 0 \quad \forall\, x \qquad \ldots (9)$$

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1 \qquad \ldots (10)$$

Ejemplo II.3

Se ha observado el tránsito en la intersección de dos avenidas de mucha afluencia, y se determinó la F.D.P. del tiempo que transcurre para que un automóvil logre pasar por el crucero; la función está dada por:

$$f_X(x) = \begin{cases} 0.1e^{-0.1x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

a) Demostrar que la función cumple con la propiedad (10).

b) Calcular la probabilidad de que un automóvil tarde en pasar cuando menos cinco minutos.

Solución

a) $$\int_{-\infty}^{\infty} f_X(x)\,dx = \int_{-\infty}^{0} f_X(x)\,dx + \int_{0}^{\infty} f_X(x)\,dx$$

Como la función vale cero en el intervalo $(-\infty, 0)$

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \int_{0}^{\infty} 0.1e^{-0.1x}\,dx$$

$$= 0.1\left[\frac{e^{-0.1x}}{-0.1}\right]_0^{\infty} = \frac{0.1}{-0.1}\,[0 - 1] = 1$$

b) $$P(X \geq 5) = 0.1\int_{5}^{\infty} e^{-0.1x}\,dx = \frac{0.1}{-0.1}\left[0 - e^{-0.1(5)}\right] = 0.6065$$

En la siguiente figura se muestra el área que representa la probabilidad buscada.

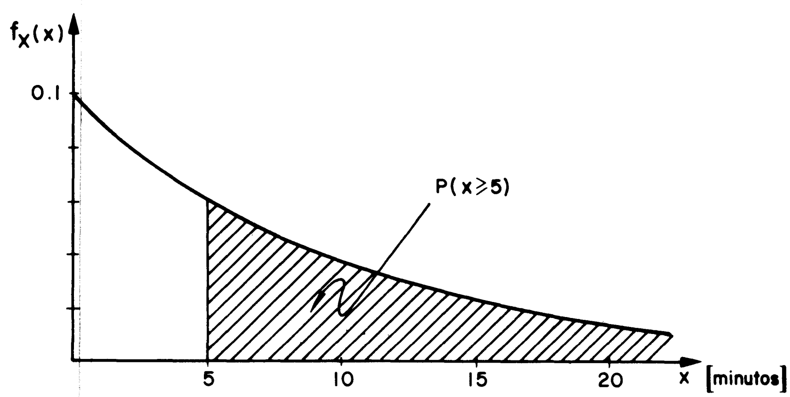


Figura II.4

FUNCION DE DISTRIBUCION ACUMULADA PARA VARIABLES ALEATORIAS CONTINUAS

La definición y las propiedades de la función de distribución acumulada (F.D.A.) no dependen del tipo de variable aleatoria que se estudia. Sin embargo, la manera de obtenerla es diferente en los casos discreto y continuo.

Para las variables continuas:

$$\boxed{F_X(x_0) = P(x \leq x_0) = \int_{-\infty}^{x_0} f_X(x)\, dx} \qquad \ldots (11)$$

Ejemplo II.4

Determinar la F.D.A. del ejemplo II.3

Solución

De la expresión (11)

$$F_X(x_0) = \int_{-\infty}^{x_0} 0.1e^{-0.1x}\, dx = \frac{0.1}{-0.1}\left[e^{-0.1x}\right]_{-0}^{x_0}$$

$$= (-1)\left[e^{-0.1x_0} - 1\right] = 1 - e^{-0.1x_0}$$

En la siguiente figura se puede observar que la F.D.A. cumple con las propiedades (4), (5), (6), (7) y (8).

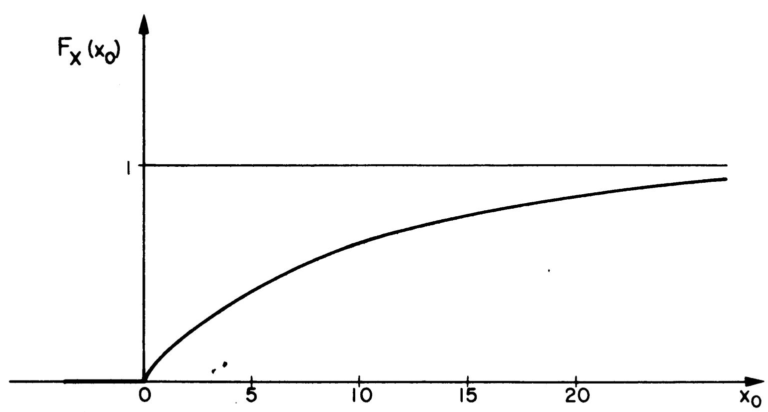


Figura II.5

II.4 VARIABLES ALEATORIAS MIXTAS

Una variable aleatoria X es mixta cuando puede tomar cualquier valor real dentro de cierta región del dominio de definición, y fuera de él sólo se define para un número finito o infinito numerable de puntos.

Su distribución de probabilidad está dada por una F.M.P. y una F.D.P. En la figura II.6 se encuentra graficada la distribución de probabilidad de una variable aleatoria mixta.

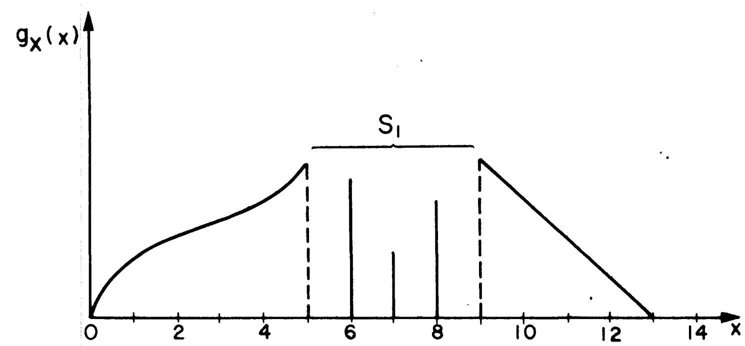


Figura II.6

Obsérvese que en la región:

$$S_1 = \{x | 5 < x < 9\}$$

la variable es discreta, mientras que en el resto del dominio es continua.

Existen pocos ejemplos de variables aleatorias mixtas y normalmente se puede discretizar la parte continua, dividiéndola en pequeños subintervalos iguales, con el objeto de simplificar el estudio del problema.

Sin embargo, es sencillo obtener las propiedades de estas variables a partir de las expresiones presentadas en los casos anteriores.

II.5 ESPERANZA MATEMATICA

Sea X una variable aleatoria discreta con función masa de probabilidad $P_X(x)$ y sea $h(x)$ una función cualquiera de X, entonces se define la esperanza de $h(x)$ como:

$$\boxed{E\{h(x)\} = \sum_{\forall x} h(x) \quad P_X(x)} \qquad \ldots (12)$$

Si la variable aleatoria X es de tipo continuo con función densidad de probabilidad $f_X(x)$, la esperanza de una función $h(x)$ se define como:

$$\boxed{E\{h(x)\} = \int_{-\infty}^{\infty} h(x) \quad f_X(x) \; dx} \qquad \ldots (13)$$

Para darle una interpretación al concepto de esperanza matemática, se puede observar que si $h(x) = x$ y se considera la siguiente F.M.P.

x	3	4	5	6
$P_X(x)$	0.2	0.4	0.3	0.1

El valor esperado de X (o esperanza de X) es:

$$E[X] = \sum_{x=3}^{6} x \, P_X(x)$$

$$= 3(0.2) + 4(0.4) + 5(0.3) + 6(0.1)$$

$$= 4.3$$

que representa el valor medio de la variable X, calculado como un promedio ponderado, en donde la probabilidad de que X tome el valor x, es precisamente la ponderación.

En general, la esperanza se puede interpretar como un *promedio ponderado* de $h(x)$ y como se verá posteriormente es de gran utilidad para definir algunos parámetros que caracterizan el comportamiento de la variable aleatoria.

PROPIEDADES DE LA ESPERANZA

Considerando la operación $E[h(x)]$ como un operador algebraico que al aplicarse sobre una función $h(x)$ produce un número real, se tienen las siguientes propiedades del operador:

a) $E[K] = K$ para cualquier constante K ... (14)

b) $E[K\,h(x)] = K\,E[h(x)]$ ... (15)

c) $E[h_1(x) + h_2(x)] = E[h_1(x)] + E[h_2(x)]$ ... (16)

Demostración:

Para el caso continuo

a) De la expresión (13)

$$E\left[K\right] = \int_{-\infty}^{\infty} K \; f_X(x)\; dx \;=\; K \int_{-\infty}^{\infty} f_X(x)\; dx$$

como $\int_{-\infty}^{\infty} f_X(x)\; dx = 1$

$$E\left[K\right] = K(1) = K$$

b) $$E\left[K\, h\,(x)\right] = \int_{-\infty}^{\infty} K\, h(x)\; f_X(x)\; dx$$

$$= K \int_{-\infty}^{\infty} h(x)\; f_X(x)\; dx$$

$$= K\, E\left[h(x)\right]$$

c) $$E\left[h_1(x) + h_2(x)\right] = \int_{-\infty}^{\infty} \left[h_1(x) + h_2(x)\right]\; f_X(x)\; dx$$

$$= \int_{-\infty}^{\infty} h_1(x)\; f_X(x)\; dx + \int_{-\infty}^{\infty} h_2(x)\; f_X(x)\, dx$$

$$= E\left[h_1(x)\right] + E\left[h_2(x)\right]$$

La demostración para el caso discreto se puede hacer de manera semejante.

Ejemplo II.5

Considerando los datos del ejemplo II.3, determinar:

a) $E\left[X\right]$

b) $E\left[X + 2\right]$

Solución

a) De la expresión (13):

$$E[X] = \int_{-\infty}^{\infty} x\,[0.1\ e^{-0.1x}]\,dx$$

$$= \int_{0}^{\infty} 0.1\ x\, e^{-0.1x}\,dx$$

resolviendo la integral por partes:

$$u = 0.1x \qquad du = 0.1dx$$

$$dv = e^{-0.1x}\,dx \qquad v = -10e^{-0.1x}$$

con lo cual:

$$E[X] = [-\ xe^{-0.1x}]_0^{\infty} + \int_0^{\infty} e^{-0.1x}\,dx$$

$$= \lim_{x \to \infty} \left[-\ \frac{x}{e^{0.1x}}\right] - 0 + \left[\frac{1}{-0.1}\ e^{-0.1x}\right]_0^{\infty}$$

obteniendo el límite mediante la regla de L'hopital:

$$E[X] = 0 + [0 - (1/-0.1)] = 10$$

b) Por las propiedades de linealidad:

$$E[X + 2] = E[X] + E[2]$$

de la expresión (14):

$$E[2] = 2$$

con lo cual:

$$E[X + 2] = 10 + 2 = 12$$

La esperanza matemática es de gran utilidad para calcular algunos parámetros que describen ciertos aspectos de las distribuciones de probabilidad y, en ocasiones, son suficientes para tener una idea precisa del comportamiento de la variable aleatoria.

Entre los parámetros más importantes se encuentran la *media y la variancia*. Para estudiarlos, se presentarán dos casos particulares de la esperanza matemática que son: los momentos con respecto al origen y los momentos con respecto a la media.

II.6 MOMENTOS CON RESPECTO AL ORIGEN

Para una distribución de probabilidad, se define el momento de orden n con respecto al origen (donde n es un número entero positivo), como la esperanza de la función $h(x) = X^n$; es decir:

$$\boxed{E\left[X^n\right] = \sum_{\forall x} x^n \, P_X(x) \qquad \text{(caso discreto)}} \quad \ldots (17)$$

$$\boxed{E\left[X^n\right] = \int_{-\infty}^{\infty} x^n \, f_X(x) \, dx \quad \text{(caso continuo)}} \quad \ldots (18)$$

Los momentos con respecto al origen que más se utilizan son el primero y el segundo, sin embargo se puede obtener el momento para cualquier valor de n.

Al primer momento con respecto al origen se le llama *media de la distribución* y, generalmente, se presenta por μ_X. La media es uno de los parámetros más representativos de la distribución y como se vio en los dos últimos ejemplos, conociendo la media o valor esperado de la distribución se tiene una idea del *valor central* de la distribución. Sin embargo, como se verá más adelante, no es el único.

Haciendo una analogía con el concepto físico de primer momento, la media es la abscisa del centroide en la F.M.P. o la F.D.P.

En la siguiente figura se muestra la media de la F.D.P. de los últimos dos ejemplos y en ella se observa que para el caso discreto, la variable no puede tomar el valor μ_X, por ser un número fraccionario.

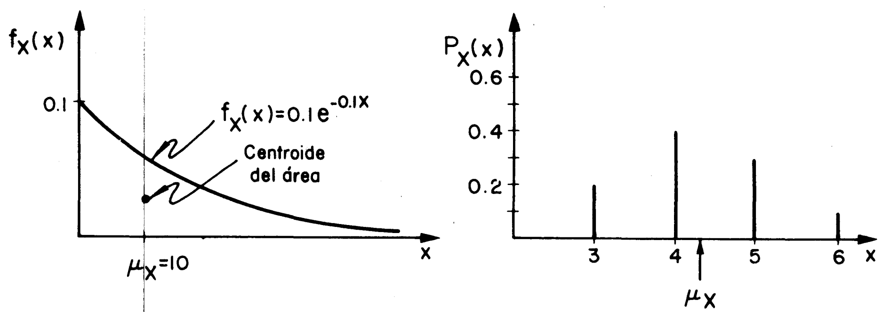


Figura II.7

El segundo momento con respecto al origen no tiene la misma importancia de la media; sin embargo, se utiliza para simplificar el cálculo de otros parámetros de la distribución que se presentan posteriormente.

II.7 MOMENTOS CON RESPECTO A LA MEDIA

El momento de orden n con respecto a la media se define como la esperanza de la función $h(x) = (X - \mu_X)^n$, donde n es un número entero, positivo y μ_X es la media de la distribución, es decir:

$$\boxed{\begin{aligned} E\left[(X - \mu_X)^n\right] &= \sum_{\forall x} (x - \mu_X)^n \, p_X(x) \\ E\left[(X - \mu_X)^n\right] &= \int_{-\infty}^{\infty} (x - \mu_X)^n \, f_X(x)\, dx \end{aligned}}$$

El primer momento con respecto a la media es cero para cualquier distribución de probabilidad debido a que:

$$E\left[(X - \mu_X)\right] = E\left[X\right] - \mu_X$$

y como $E\left[X\right] = \mu_X$

$$E\left[(X - \mu_X)\right] = 0$$

El segundo momento con respecto a la media llamado *variancia o varianza* es, junto con la media, uno de los más importantes parámetros descriptivos, ya que es un indicador de la dispersión de los posibles valores de la variable aleatoria con respecto a su media μ_X.

La variancia se representa generalmente como σ_X^2 ó $VAR\left[X\right]$ y de acuerdo con lo anterior, se define como:

$$\boxed{\sigma_X^2 = E\left[(X - \mu_X)^2\right]} \qquad \ldots (19)$$

Si se considera a la esperanza como un promedio y a la diferencia $(X - \mu_X)$ como la distancia entre un valor particular de X y su media, entonces se interpreta a la variancia como el promedio de los cuadrados de las diferencias entre el valor medio (μ_X) y los posibles valores de la variable aleatoria.

Debe observarse que si el exponente n toma cualquier otro valor par, se tendrá también un indicador de la dispersión de los datos y éste será siempre positivo, mientras que para una potencia impar, el momento puede ser positivo, negativo o cero.

En la siguiente figura se observan dos distribuciones de probabilidad $f_{X_1}(x)$ y $f_{X_2}(x)$, con variancias $\sigma^2_{X_1}$ y $\sigma^2_{X_2}$; comparando las funciones se establece que $\sigma^2_{X_1} > \sigma^2_{X_2}$

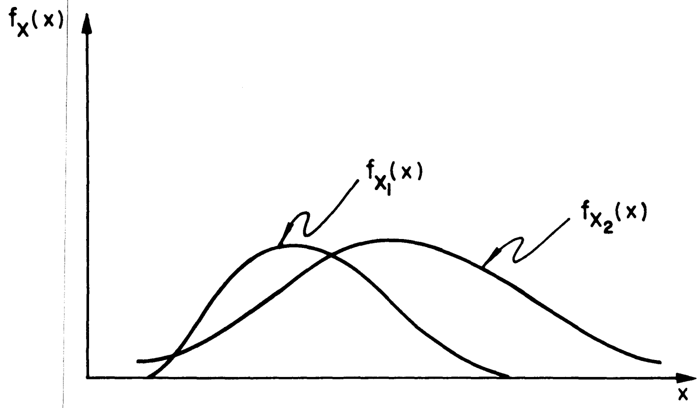


Figura II.8

PROPIEDADES DE LA VARIANCIA

Si a es una constante cualquiera y X una variable aleatoria con media μ_X y variancia σ^2_X, entonces:

$$VAR[X + a] = \sigma^2_X \qquad \ldots (20)$$

Demostración:

De la definición de variancia:

$$VAR[X + a] = E\left[(X + a) - E[X + a]\right]^2$$

$$= E[X + a - E[X] - a]^2$$

como $E[X] = \mu_X$

$$VAR[X + a] = E[(X - \mu_X)^2] = \sigma^2_X$$

Esta propiedad indica que al sumar una constante a la variable aleatoria, solamente se recorre la distribución sobre el eje X, manteniendo su forma, como se indica en la figura II.9, por lo que la dispersión de la función se mantiene igual.

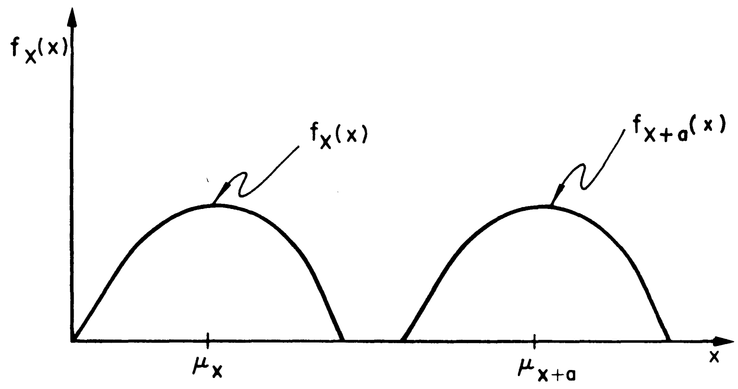


Figura II.9

De manera semejante se puede demostrar que:

$$VAR(aX) = a^2 \sigma_X^2 \qquad \ldots (21)$$

En esta expresión lo importante es observar que al multiplicar la variable por una constante, la variancia se modifica con el cuadrado de la constante.

Otra propiedad que en ocasiones simplifica el cálculo de la variancia es:

$$\sigma_X^2 = E\left[X^2\right] - \mu_X^2 \qquad \ldots (22)$$

La demostración es similar a la que se hizo para la expresión (20).

Como una medida de dispersión, la variancia tiene el inconveniente de proporcionar valores con unidades que son el cuadrado de las que tiene la variable aleatoria. Otra medida de dispersión es la *desviación estándar*, denotada por σ ó σ_X y definida como la raíz cuadrada de la variancia.

Es decir:

$$\boxed{\sigma_X = \sqrt{\sigma_X^2} = \sqrt{E\left[(X - \mu_X)^2\right]}} \qquad \ldots (23)$$

Si se desea comparar la variancia o la desviación estándar de varias variables con diferentes unidades entre sí, la comparación resultará injusta, debido a que las distribuciones de variables aleatorias, con unidades pequeñas, tendrán aparentemente menos variación.

Para hacer este tipo de comparaciones se define otro indicador, denominado *coeficiente de variación*, como el cociente de la desviación estándar entre la media, es decir:

$$C.V. = \frac{\sigma_X}{\mu_X} \qquad \ldots (24)$$

Como la media y la desviación estándar tienen las mismas unidades, el coeficiente de variación es adimensional.

Ejemplo II.6

Sea X una variable aleatoria discreta con la siguiente distribución de probabilidad:

x	-2	3	5	7
$P_X(x)$	0.2	0.4	0.3	0.1

Determinar:

a) La media.

b) La desviación estándar.

c) El coeficiente de variación.

Solución

a) $E[X] = \sum_{\forall x} x\, P_X(x) = (-2)(0.2) + 3(0.4) + 5(0.3) + 7(0.1)$

$E[X] = 3$

b) Por las expresiones (22) y (23):

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{E\{X^2\} - \mu_X^2}$$

en donde el segundo momento con respecto al origen es:

$$E\{X^2\} = (-2)^2(0.2) + 3^2(0.4) + 5^2(0.3) + 7^2(0.1) = 16.8$$

por lo tanto:

$$\sigma_X = \sqrt{16.8 - 3^2} = 2.79$$

c) Con los resultados anteriores y de la expresión (24):

$$C.V. = \frac{2.79}{3} = 0.93$$

II.8 TRANSFORMADAS

Para calcular los momentos de una distribución con respecto al origen, se pueden presentar dificultades al tratar de aplicar directamente las definiciones.

Una forma alternativa de hacerlo es mediante transformadas. Entre las más importantes se encuentran las siguientes:

TRANSFORMADA GEOMETRICA

Sea X una variable aleatoria discreta, cuya F.M.P. es $P_X(x)$, entonces se define la transformada geométrica de X como:

$$\boxed{T_X(Z) = E\left[Z^x\right] = \sum_{x=0}^{\infty} Z^x P_X(x)} \quad \ldots (25)$$

Propiedades de la transformada geométrica:

$$T_X(1) = 1 \quad \ldots (26)$$

$$\left.\frac{d\ T_X(Z)}{dZ}\right|_{Z=1} = E\left[X\right] = \mu_X \quad \ldots (27)$$

$$\left.\frac{d^2 T_X(Z)}{dZ^2}\right|_{Z=1} = E\left[X^2\right] - \mu_X \quad \ldots (28)$$

debe notarse que de la expresión (22) y (28):

$$\sigma_X^2 = E\left[X^2\right] - \mu_X^2 = \left.\frac{d^2\ T_X(Z)}{dZ^2}\right|_{Z=1} + \mu_X - \mu_X^2 \quad \ldots (29)$$

Demostración de las propiedades

Para la expresión (26), por definición de transformada geométrica:

$$T_X(1) = \sum_{x=0}^{\infty} (1)^x P_X(x)$$

$$= \sum_{x=0}^{\infty} P_X(x) = 1$$

para la expresión (27):

$$\left.\frac{d\ T_X(Z)}{dZ}\right|_{Z=1} = \left.\frac{d}{dZ}\left[\sum_{x=0}^{\infty} Z^x\ P_X(x)\right]\right|_{Z=1}$$

$$= \left.\sum_{x=0}^{\infty} x\ Z^{x-1}\ P_X(x)\right|_{Z=1} = \sum_{x=0}^{\infty} x(1)^{x-1}\ P_X(x)$$

$$= \sum_{x=0}^{\infty} x\ P_X(x) = \sum_{x=0}^{\infty} x\ P_X(x)$$

que es definición de $E[X]$

Para la expresión (28):

$$\left.\frac{d^2T_X(Z)}{dZ^2}\right|_{Z=1} = \sum_{x=0}^{\infty} x(x-1)\ Z^{x-2}\ P_X(x)\Big|_{Z=1}$$

$$= \sum_{x=0}^{\infty} (x^2 - x)\ P_X(x)$$

$$= \sum_{x=0}^{\infty} x^2\ P_X(x) - \sum_{x=0}^{\infty} x\ P_X(x)$$

simplificando:

$$= E\left[X^2\right] - \mu_X$$

Ejemplo II.7

A partir de la F.M.P.

$$P_X(x) = \begin{cases} \dfrac{2}{3^x} & \text{para } x = 1, 2, \ldots \\ 0 & \text{para cualquier otro caso} \end{cases}$$

y mediante la transformada geométrica, calcular:

a) La media.

b) La variancia.

Solución

a) Por definición:

$$T_X(Z) = \sum_{x=0}^{\infty} Z^x\ P_X(x)$$

Como $P_X(x)$ toma valores diferentes de cero $(x = 1, 2, \ldots)$ y de la expresión (27):

$$\left.\frac{d\,T_X(Z)}{dZ}\right|_{Z=1} = \frac{d}{dZ}\left[\sum_{x=1}^{\infty} Z^x \frac{2}{3^x}\right]_{Z=1}$$

$$= 2\,\frac{d}{dZ}\left[\sum_{x=1}^{\infty}\left(\frac{Z}{3}\right)^x\right]_{Z=1}$$

donde la sumatoria es una serie geométrica de la forma:

$$\sum_{x=1}^{\infty} a^x = \frac{a}{1-a}$$

entonces:

$$\left.\frac{d\,T_X(Z)}{dZ}\right|_{Z=1} = 2\,\frac{d}{dZ}\left[\frac{\frac{Z}{3}}{1-\frac{Z}{3}}\right]_{Z=1}$$

$$= 2\left[\frac{\left(1-\frac{Z}{3}\right)\frac{1}{3}+\frac{Z}{9}}{\left(1-\frac{Z}{3}\right)^2}\right]_{Z=1}$$

$$= 2\left[\frac{\frac{1}{3}}{\left(1-\frac{Z}{3}\right)^2}\right]_{Z=1} = \left[\frac{\frac{2}{3}}{\left(1-\frac{Z}{3}\right)^2}\right]_{Z=1}$$

$$= \frac{2}{3\left(1-\frac{1}{3}\right)^2} = 1.5$$

b) De la expresión (29):

$$\sigma_X^2 = \left.\frac{d^2T_X(Z)}{dZ^2}\right|_{Z=1} + \mu_X - \mu_X^2$$

donde:

$$\left.\frac{d^2T_X(Z)}{dZ^2}\right|_{Z=1} = \frac{d}{dZ}\Big[T'(Z)\Big]_{Z=1} = \frac{d}{dZ}\left[\frac{\frac{2}{3}}{\left(1-\frac{Z}{3}\right)^2}\right]_{Z=1}$$

$$= \left[\frac{-\frac{2}{3}\,(2)\,\left(1-\frac{Z}{3}\right)\left(-\frac{1}{3}\right)}{\left(1-\frac{Z}{3}\right)^4}\right]_{Z=1} = \left[\frac{\frac{4}{9}\left(1-\frac{Z}{3}\right)}{\left(1-\frac{Z}{3}\right)^4}\right]_{Z=1}$$

$$= \left[\frac{\frac{4}{9}}{\left(1-\frac{Z}{3}\right)^3}\right]_{Z=1} = \frac{4}{9\left(1-\frac{Z}{3}\right)^3} = 1.5$$

con lo cual:

$$\sigma_X^2 = 1.5 + 1.5 - (1.5)^2 = 0.75$$

TRANSFORMADA EXPONENCIAL O DE LAPLACE

Sea X una variable aleatoria continua cuya F.D.P. es $f_X(x)$; la transformada exponencial de X se define como:

$$\boxed{L_X(S) = \int_{-\infty}^{\infty} e^{-sx}\, f_X(x)\, dx = E\left[e^{-sx}\right]}$$

Propiedades de la transformada exponencial:

$$L_X(0) = 1 \qquad \ldots (30)$$

$$\left.\frac{d^n L_X(S)}{dS^n}\right|_{S=0} = (-1)^n\, E\left[x^n\, e^{-Sx}\right]_{S=0} = (-1)^n\, E\left[x^n\right] \qquad \ldots (31)$$

Demostración:

$$L_X(0) = \int_{-\infty}^{\infty} e^{(0)x}\, f_X(x)\, dx = \int_{-\infty}^{\infty} f_X(x)\; dx$$

$$L_X(0) = 1$$

Para $n = 1$:

$$\left.\frac{d\; L_X(S)}{dS}\right|_{S=0} = \frac{d\; E\left[e^{-Sx}\right]}{dS}$$

Como la esperanza es un operador lineal:

$$\left.\frac{d\; L_X(S)}{dS}\right|_{S=0} = E\left[\frac{de^{-Sx}}{dS}\right]_{S=0}$$

$$= E\left[-xe^{-Sx}\right]_{S=0} = -E\left[x\right]$$

Para $n = k - 1$

$$\left.\frac{d^{k-1}L_X(S)}{dS^{k-1}}\right|_{S=0} = (-1)^{k-1}\,E\left[x^{k-1}\,e^{-Sx}\right]_{S=0} = (-1)^{k-1}\,E\left[x^{k-1}\right]$$

finalmente para $n = k$

$$\left.\frac{d^k L_X(S)}{dS^k}\right|_{S=0} = \frac{d}{dS}\left[\frac{d^{k-1}L_X(S)}{dS^{k-1}}\right]_{S=0}$$

$$\frac{d^k L_X(S)}{dS^k} = \frac{d}{dS}\,E\left[(-1)^{k-1}\,x^{k-1}\,e^{-Sx}\right]_{S=0}$$

$$= E\left[(-1)^k\,x^k\,e^{-Sx}\right]_{S=0}$$

$$\frac{d^k L_X(S)}{dS^k} = (-1)^k\,E\left[x^k\right]$$

Ejemplo II.8

Determinar la media y la variancia de la siguiente F.D.P. a través de la transformada exponencial:

$$f_X(x) = \begin{cases} \dfrac{e^{-x/4}}{4} & ; \quad x \geq 0 \\ 0 & ; \quad x < 0 \end{cases}$$

Solución

De la definición:

$$L_X(S) = \int_0^\infty e^{-Sx}\,\frac{1}{4}\,e^{-x/4}\,dx$$

$$= \frac{1}{4}\int_0^{\infty} e^{-x(S + 1/4)}\ dx$$

$$= \frac{1}{4}\left[\frac{1}{-(S+\frac{1}{4})}\ e^{-x(S + 1/4)}\right]_0^{\infty} = \frac{1}{4}\left[0 + \frac{1}{(S+\frac{1}{4})}\right]$$

$$= \frac{1}{4(S + \frac{1}{4})}$$

De la expresión (31):

$$\mu_X = \left.\frac{-d\ L_X(S)}{dS}\right|_{S=0} = \frac{-d}{dS}\left[\frac{1}{4(S+\frac{1}{4})}\right]_{S=0}$$

$$= \left. - \frac{1}{4(S+\frac{1}{4})^2}\right|_{S=0} = \frac{1}{4(\frac{1}{16})} = 4$$

Para obtener la variancia, de la expresión (22):

$$\sigma_X^2 = E\ [X^2] - \mu_X^2$$

donde:

$$E\ [X^2] = \frac{d}{dS}\left[\frac{d\ T_X(S)}{dS}\right]_{S=0}$$

$$= \left.\frac{2}{4(S+\frac{1}{4})^3}\right|_{S=0} = \frac{2}{\frac{1}{16}} = 32$$

por lo tanto:

$$\sigma_X^2 = 32 - (4)^2 = 16$$

FUNCION GENERATRIZ DE MOMENTOS

Para una variable aleatoria X (discreta o continua), se define la función generatriz de momentos como:

$$\boxed{M_X(S) = E\left[e^{Sx}\right] = \begin{cases} \sum_{x=0}^{\infty} e^{Sx}\ p_X(x) & \text{(caso discreto)} \\ \int_0^{\infty} e^{Sx}\ f_X(x)\ dx & \text{(caso continuo)} \end{cases}}$$

Debe notarse que, en el caso continuo, esta transformada es casi igual a la transformada de Laplace y únicamente cambia por el signo de la variable s.

Propiedades de la función generatriz de momentos:

$$M_X(0) = 1 \qquad \ldots (32)$$

$$\left.\frac{d^n M_X(S)}{dS^n}\right|_{S=0} = E\left[x^n\right] \quad \text{para } n = 0, 1, 2, \ldots \qquad \ldots (33)$$

A esta última propiedad le debe su nombre, debido a que se puede generar cualquier momento con respecto al origen. La demostración de cada una de las propiedades es en todo similar a las presentadas en la transformada exponencial.

La función generatriz de momentos puede sustituir a la transformada geométrica y a la transformada de Laplace si se consideran las siguientes relaciones:

$$M_X(S) = T_X(e^S)$$

$$M_X(S) = L_X(-S)$$

es decir, que se sustituye el argumento z por e^S en el primer caso y S por -s en el segundo caso.

Otra característica importante de la función generatriz de momentos es que cuando existe, es única e identifica plenamente a la distribución de probabilidad correspondiente; es decir, que a cada distribución de probabilidad le corresponde una y sólo una función generatriz de momentos.

FUNCION CARACTERISTICA

Para una variable aleatoria X (discreta o continua) se define la función característica de X como:

$$\boxed{\phi_X(t) = E\left[e^{it\,x}\right] = \begin{cases} \displaystyle\sum_{x=0}^{\infty} e^{it\,x} \quad P_X(x)\,dx & \text{caso discreto} \\ \displaystyle\int_0^{\infty} e^{it\,x} \quad f_X(x)\,dx & \text{caso continuo} \end{cases}}$$

donde $i = \sqrt{-1}$

Las propiedades de esta función son:

1. $\phi_X(0) = 1 \qquad \ldots (34)$

2. $\dfrac{d^n \phi_X(t)}{dt^n} = (-i)^n\, E\left[x^n\right] \qquad \ldots (35)$

La demostración de estas propiedades es también similar a la efectuada en la transformada exponencial y la ventaja que tiene con respecto a las anteriores es que para cualquier distribución de probabilidad existe su función característica, lo cual no se puede asegurar para las transformadas anteriores.

La función característica se relaciona con la transformada de Laplace y con la función generatriz de momentos en la siguiente forma:

$$M_X(S) = \phi_X(-is)$$

$$L_X(S) = \phi_X(is)$$

CAPITULO III MODELOS PROBABILISTICOS

INTRODUCCION

Cuando se estudia un experimento aleatorio mediante la teoría de la probabilidad y la estadística, es necesario definir el comportamiento de las variables involucradas. Una forma de hacerlo es mediante su función masa o densidad de probabilidad; sin embargo, encontrar para cada variable su verdadera distribución de probabilidad es una tarea muy difícil y, por tal motivo, se han desarrollado una gran cantidad de modelos probabilísticos que explican satisfactoriamente el comportamiento de muchas variables aleatorias.

En este capítulo se presentan algunos de los modelos teóricos, tanto para variables aleatorias discretas como continuas, incluyendo su forma, sus parámetros y otras características particulares.

III.1 DISTRIBUCIONES TEORICAS DE VARIABLES ALEATORIAS DISCRETAS

Algunos de los modelos que se presentan en este capítulo se han desarrollado a partir de un experimento denominado *ensayo de Bernoulli*, en el cual sólo se pueden presentar dos posibles resultados, denominados *éxito* y *fracaso*. La probabilidad de que ocurra un éxito se representa con p y la probabilidad de que ocurra un fracaso con q ó $1-p$.

Como ejemplos de un ensayo de Bernoulli se puede considerar el lanzamiento de una moneda, cuyos posibles resultados son águila o sol, o bien tomar un producto de una línea de producción para verificar si cumple con las normas de calidad preestablecidas, en cuyo caso los posibles resultados son bueno o defectuoso.

Llamarle éxito a uno de los posibles resultados no significa forzosamente que sea el más favorable, generalmente se le llama así al evento que tiene menos probabilidad de ocurrencia; por ejemplo, se puede llamar éxito al posible evento de que un vehículo se descomponga durante un intervalo de tiempo y fracaso, si no sufre ningún daño.

PROCESO O SECUENCIA DE BERNOULLI

Consiste en una serie de experimentos o ensayos de Bernoulli independientes, donde la probabilidad de éxito p permanece constante durante todo el proceso. Por las características anteriores, se dice que el proceso es dicotómico, estacionario e independiente.

Ejemplo III.1

En una fábrica de automóviles se ha observado que el 10% de las unidades salen defectuosas de la línea de producción. Si se selecciona un lote de cinco para ser inspeccionado, ¿cuál es la probabilidad de que el primero y el cuarto salgan defectuosos?

Solución

Si se define el éxito y el fracaso respectivamente como:

E = {encontrar un automóvil defectuoso} y

F = {encontrar un automóvil sin defecto}

se tiene entonces, por la regla de la multiplicación, que:

$$P(E \cap F \cap F \cap E \cap F) = P(E)\ P(F)\ P(F)\ P(E)\ P(F)$$

como $P(E) = 0.1$ y $P(F) = 1 - 0.1 = 0.9$

$$P(E \cap F \cap F \cap E \cap F) = (0.1)^2\ (0.9)^3 = 0.007$$

De este resultado se puede observar que si se cambia el orden de la secuencia, la probabilidad es la misma. A esta propiedad se le llama intercambiabilidad del proceso de Bernoulli.

III.1.1 DISTRIBUCION DE BERNOULLI

Si X *es el número de éxitos en un ensayo de Bernoulli* con probabilidad p de éxito y 1-p ó q de fracaso, entonces se dice que X tiene distribución binaria o de Bernoulli con parámetro p.

Como X solamente puede tomar el valor de cero o uno, su función masa de probabilidad está definida por:

$$\boxed{p_X(x) = \begin{cases} 1-p & \text{para } x = 0 \\ p & \text{para } x = 1 \\ 0 & \text{para cualquier otro caso} \end{cases}}$$

que gráficamente se muestra en la siguiente figura.

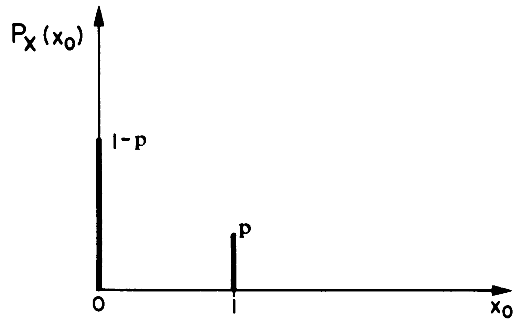


Figura III.1

La función generatriz de momentos de la distribución de Bernoulli es:

$$M_X(s) = \sum_{\forall x} e^{sx} P_X(x) = e^{s(0)} (1 - p) + e^{s(1)} p = 1 - p + p e^{s} \quad ..(1)$$

La media o valor esperado de esta distribución se puede obtener de acuerdo con las propiedades de la función generatriz de momentos como:

$$E[X] = \left.\frac{d\, M_X(s)}{ds}\right|_{s=0} = p\, e^{(0)} = p$$

de la expresión (22) del capítulo II, la variancia se obtiene como:

$$\sigma_X^2 = E[X^2] - E[X]^2$$

donde:

$$E[X^2] = \left.\frac{d^2 M_X(s)}{d s^2}\right|_{s=0} = p \quad \ldots (2)$$

con lo cual:

$$\sigma_X^2 = p - p^2 = p(1 - p) = pq \quad \ldots (3)$$

Aun cuando esta distribución no se aplica por sí sola en el estudio de fenómenos de la realidad, es la parte fundamental de los modelos basados en procesos bernoullianos.

Conviene destacar el hecho de que ésta y las demás distribuciones que se presentan en este capítulo son en realidad *familias de funciones*, debido a que para cada valor del parámetro (o parámetros) se obtienen diferentes funciones que pertenecen a la misma familia.

III.1.2 DISTRIBUCION BINOMIAL

Sea K una variable aleatoria que representa *el número de éxitos en una secuencia de* n *ensayos de Bernoulli* con probabilidad p de éxito y $1-p$ de fracaso, entonces K tiene distribución binomial con parámetros n y p.

Para determinar la función masa de probabilidad de la distribución binomial, supóngase que se desea determinar la probabilidad de encontrar K piezas defectuosas de un lote de tres unidades, tomadas de una de las líneas de producción de una planta industrial. Considérese que la probabilidad de que una pieza resulte defectuosa es p, esto es $p(\text{éxito}) = p$ y $p(\text{fracaso}) = 1 - p$.

Tener cero éxitos ($k = 0$) equivale a no encontrar piezas defectuosas en el lote, de manera que:

$$P(k = 0) = P_K(0) = P(F \cap F \cap F)$$

como los eventos son independientes:

$$P(k = 0) = P(F)\, P(F)\, P(F)$$
$$= (1 - p)^3$$

Para $k = 1$ existen $C(3, 1) = \frac{3!}{1!\ 2!} = 3$ maneras de obtener un éxito, que son:

E F F
F E F
F F E

con lo cual:

$$P_K(1) = P\,[(E \cap F \cap F) \cup (F \cap E \cap F) \cup (F \cap F \cap E)]$$

como las tres alternativas tienen un éxito y dos fracasos:

$$P_K(1) = 3P(1 \text{ éxito y } 2 \text{ fracasos})$$
$$= 3P(1 - p)^2$$

Para $k = 2$ existen $C(3, 2) = \frac{3!}{2!\ 1!} = 3$ maneras de obtener dos éxitos, que son:

E E F
F E E
E F E

y de manera semejante al caso anterior:

$$P_K(2) = 3\, p^2(1 - p)$$

Por último, para $k = 3$ sólo pueden ocurrir tres éxitos cuando las tres piezas sean defectuosas, por lo cual:

$$P_K(3) = P(E \cap E \cap E)$$
$$= p^3$$

Resumiendo los resultados anteriores, la F.M.P. de K está dada por:

$$P_K(0) = (1 - P)^3 = q^3$$

$$P_K(1) = 3\,P(1 - P)^2 = 3\,pq^2$$

$$P_K(2) = 3\,P^2(1 - P) = 3\,p^2q$$

$$P_K(3) = p^3$$

Como puede observarse, los resultados anteriores corresponden a los términos que se obtienen al desarrollar el binomio $(p + q)^3$, por este motivo se le llama distribución binomial.

El número de maneras en las cuales se pueden presentar k éxitos en una serie de n ensayos, es el número combinatorio:

$$C(n, k) = \frac{n!}{k!(n - k)!}$$

y la probabilidad de cada uno de los posibles resultados es $P^k(1 - P)^{n-k}$, por lo cual la F.M.P. de la distribución binomial se representa en forma general como:

$$\boxed{P_K(k) = \begin{cases} C(n, k)\; p^k\; q^{n-k} & \text{para } k \geq 0 \\ 0 & k < 0 \end{cases}} \quad \ldots (4)$$

donde k y n son enteros positivos y las constantes n y p son los parámetros de esta distribución.

MOMENTOS DE LA DISTRIBUCION BINOMIAL

Para obtener la media y la variancia de la distribución binomial, nuevamente conviene obtener su función generatriz de momentos y aplicar las propiedades vistas en el capítulo anterior.

La función generatriz de momentos está dada por:

$$M_K(s) = \sum_{k=0}^{n} e^{sk}\, P_K(k)$$

en este caso:

$$M_K(s) = \sum_{k=0}^{n} e^{sk}\, C(n, k)\; p^k\; q^{n-k}$$

por las propiedades de los exponentes:

$$M_K(s) = \sum_{k=0}^{n} C(n, k) \left[p\; e^s\right]^k q^{n-k}$$

que es equivalente al desarrollo de un binomio elevado a la n-ésima potencia de la forma:

$$M_K(s) = (p\; e^s + q)^n \quad \ldots (5)$$

para obtener la media:

$$\mu_K = \frac{d\, M_K(s)}{d\, s}\Bigg|_{s=0} = n(p\, e^s + q)^{n-1}\, p\, e^s\Bigg|_{s=0} = n(p + q)^{n-1}\, p$$

como $p + q = 1$

$$\mu_K = n\, p \qquad \ldots (6)$$

Para obtener la variancia, de la expresión (22) del capítulo II :

$$\sigma_K^2 = E\left[K^2\right] - \mu_K^2$$

donde:

$$E\left[X^2\right] = \frac{d^2 M_K(s)}{ds}\Bigg|_{s=0} = n(p\, e^s + q)^{n-1}\, p\, e^s + p\, e^s\left[n(n - 1)(p\, e^s + q)^{n-2}\, p\, e^s\right]_{s=0}$$

$$= n(p + q)^{n-1}\, p + p^2\left[n(n - 1)(p + q)^{n-2}\right]$$

$$= n\, p + p^2\left[n(n - 1)\right]$$

$$= n\, p + p^2 n^2 - p^2 n$$

con lo cual:

$$\sigma_K^2 = n\, p + p^2\, n^2 - p^2\, n - (n\, p)^2$$

$$= n\, p - p^2\, n$$

factorizando $n\, p$:

$$\sigma_K^2 = n\, p\, (1 - p) = n\, p\, q \qquad \ldots (7)$$

Ejemplo III.2

Considerando la línea de producción del ejemplo III.1, cuál es:

a) La probabilidad de no encontrar unidades defectuosas en un lote de cuatro automóviles.

b) El número esperado de unidades defectuosas en el mismo lote.

c) La variancia y la desviación estándar de la distribución.

d) La gráfica de la función masa de probabilidad.

Solución

a) Como en este caso la secuencia de observaciones se puede considerar un proceso de Bernoulli, la variable K que representa el número de unidades defectuosas que hay en un lote, tiene distribución binomial con parámetros $n = 4$ y $p = 0.1$, con lo cual:

$$p_K(0) = C(4, 0)(0.1)^0(0.9)^4$$
$$= (0.9)^4 = 0.656$$

b) De la expresión (6):

$$\mu_K = 4(0.1) = 0.4$$

c) De la expresión (7):

$$\sigma_K^2 = 4(0.1)(0.9) = 0.36 \quad y \quad \sigma_K = \sqrt{0.36} = 0.6$$

d) De la expresión (4) se obtiene el valor de la función para los demás puntos lo cual se muestra en la siguiente figura.

k	$p_K(k)$
0	0.6561
1	0.2916
2	0.0486
3	0.0036
4	0.0001

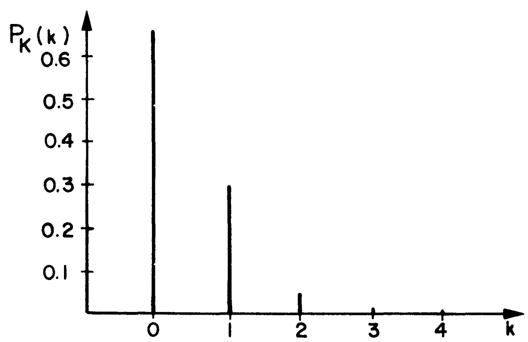


Figura III.2

III.1.3 DISTRIBUCION BINOMIAL NEGATIVA O PASCAL

Sea N *una variable aleatoria que representa el número de ensayos necesarios para encontrar* k *éxitos en una secuencia de Bernoulli* con probabilidad de éxito p, entonces N tiene distribución binomial negativa con parámetros k y p.

Comparando esta definición con la correspondiente a la binomial, se debe notar que el número de éxitos K es una variable en la distribución binomial, mientras que en esta distribución es solamente un parámetro. También se debe observar que en esta distribución, el último ensayo de la secuencia tiene que ser un éxito para completar los k éxitos que aparecen en la definición.

Tomando en cuenta esta última observación, se puede obtener la F.M.P. haciendo:

$$P_N(n) = P\left[(\text{obtener } k - 1 \text{ éxitos en los primeros } N - 1 \text{ ensayos}) \cap (\text{un éxito en el último ensayo})\right]$$

donde el primer evento corresponde a una variable con distribución binomial y la probabilidad del segundo es p; por lo tanto, de la expresión (4):

$$P_N(n) = \left[C(n - 1, k - 1)p^{k-1} q^{n-k}\right] p$$

$$= C(n - 1, k - 1)p^k q^{n-k}$$

Como no tiene sentido que haya menos de K ensayos para encontrar k éxitos, entonces:

$$\boxed{P_N(n) = \begin{cases} C(n - 1, k - 1)\, p^k q^{n-k} & \text{para } n \geq k \\ 0 & n < k \end{cases}} \quad \ldots (8)$$

MOMENTOS DE LA DISTRIBUCION BINOMIAL NEGATIVA

Siguiendo un procedimiento semejante al practicado en la distribución binomial, se obtiene la función generatriz de momentos:

$$M_N(s) = \frac{p\, e^s}{(1 - q\, e^s)^k} \quad \ldots (9)$$

donde $q = 1 - p$

Aplicando las propiedades de esta función, se obtiene la media y la variancia como:

$$\mu_N = \frac{k}{p} \quad \ldots (10)$$

$$\sigma_N^2 = \frac{k\, q}{p^2} \quad \ldots (11)$$

Ejemplo III.3

Considerando nuevamente el enunciado del ejemplo III.1, cuál es:

a) La probabilidad de que la quinta unidad observada sea la segunda defectuosa.

b) El número promedio de unidades que se debe observar para encontrar cinco defectuosas.

Solución

a) En este caso $k = 2$, $n = 5$ y $p = 0.1$ con lo cual de la expresión (8):

$$P_N(5) = C(5-1,\ 2-1)(0.1)^2(1-0.1)^{5-2}$$

$$= \frac{4!}{1!\ 3!}(0.1)^2(0.9)^3 = 0.02916$$

b) Considerando ahora $k = 5$ y de la expresión (10):

$$\mu_N = \frac{5}{0.1} = 50 \text{ unidades}$$

III.1.4 DISTRIBUCION GEOMETRICA

Esta distribución es un caso especial de la binomial negativa, en donde la variable aleatoria representa *el número de ensayos necesarios para encontrar el primer éxito en la secuencia de Bernoulli.*

La función masa de probabilidad se obtiene de la expresión (8), haciendo $k = 1$, con lo cual:

$$\boxed{P_N(n) = P(N = n \mid k = 1,\ p) = \begin{cases} p\ q^{n-1} & \text{para} \quad n \geq 1 \\ 0 & \qquad\quad n < 1 \end{cases}} \quad \ldots (12)$$

De las expresiones (9), (10) y (11) se obtienen también la función generatriz de momentos, la media y la variancia de la distribución geométrica (haciendo $k = 1$).

Ejemplo III.4

El propietario de diez terrenos campestres en un centro turístico, está tratando de venderlos por medio de entrevistas personales con los posibles compradores. Considera que al entrevistarse con un posible comprador, existe la misma probabilidad de vender o no vender y que el resultado de una entrevista es independiente de lo que ocurre en las demás. Cuál es:

a) La probabilidad de que la cuarta persona entrevistada sea la primera que compre.

b) La media de la variable N que representa el número de clientes que se tienen que entrevistar para realizar la primera venta.

c) La variancia de N.

d) La gráfica de la F.M.P. que corresponde a N en el intervalo $[1, 6]$.

Solución

a) Considerando como éxito realizar una venta $p = 0.5$, entonces: $q = 1 - 0.5 = 0.5$

con lo cual de la expresión (12):

$P_N(4) = (0.5)(0.5)^3 = (0.5)^4 = 0.063$

b) Haciendo $k = 1$ en la expresión (10):

$\mu_N = \frac{1}{0.5} = 2$ (entrevistas)

c) $\sigma_N^2 = \frac{0.5}{(0.5)^2} = 2$

d) Con la expresión (12) se obtienen los valores de la siguiente figura.

n	$P_N(n)$
1	0.500
2	0.250
3	0.125
4	0.063
5	0.031
6	0.016

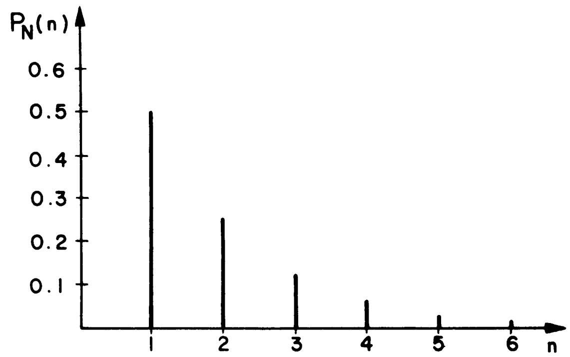


Figura III.3

III.1.5 DISTRIBUCION DE POISSON

Si se divide el intervalo de definición de una variable continua t que puede representar tiempo, distancia o cualquier otra variable continua en pequeños subintervalos iguales Δt, si en cualquier punto de t se pueden presentar dos posibles resultados denominados *éxito* y *fracaso*, entonces se puede considerar a cada subintervalo Δt como un ensayo de Bernoulli en el que puede ocurrir un éxito con probabilidad p (en cualquier Δt) o un fracaso con probabilidad $q = 1 - p$.

Cuando la ocurrencia de un éxito en un subintervalo es independiente de lo que ocurre previamente y no influye en lo que ocurre posteriormente, entonces el intervalo de definición de t representa un proceso de Bernoulli.

Si además, el tamaño de los subintervalos Δt tienden a cero y la *tasa promedio de éxitos por unidad de* t definida como $\lambda = n\,p$ permanece constante, se trata entonces de *un proceso Poisson.*

Algunos ejemplos de dicho proceso se presentan al considerar como variable continua el tiempo y como éxito la llegada de un cliente a un almacén, la descompostura de uno de los elementos de cierto mecanismo o la detección de un sismo.

Si una variable aleatoria Y representa *el número de éxitos en un proceso Poisson* con una tasa promedio de éxitos λ, entonces Y tiene distribución de Poisson con parámetro λ.

Para obtener la función masa de probabilidad de esta distribución considérese el proceso Poisson como un proceso de Bernoulli, es decir:

$$P_Y(y) = C(n, y)\, p^y (1 - p)^{n-y} \qquad \ldots (13)$$

donde n es el número de subintervalos Δt y la probabilidad de éxito p se obtiene a partir de la definición de λ como:

$$p = \frac{\lambda}{n} \qquad \ldots (14)$$

sustituyendo en la expresión (13) y considerando que:

$$C(n, y) = \frac{n!}{y!(n - y)!}$$

$$P_Y(y) = \frac{n!}{y!(n - y)!} \left(\frac{\lambda}{n}\right)^y \left(1 - \frac{\lambda}{n}\right)^{n-y}$$

$$= \frac{[n(n-1)(n-2)\ldots(n-y-1)]}{y!} \frac{\left(\frac{\lambda}{n}\right)^y \left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^y}$$

$$= \frac{n^y \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\ldots\left(1 - \frac{y+1}{n}\right)}{y!\left(1 - \frac{\lambda}{n}\right)^y} \frac{\lambda^y \left(1 - \frac{\lambda}{n}\right)^n}{n^y}$$

Obteniendo el límite cuando n tiende a infinito (lo que equivale a que los subintervalos $\Delta t \to 0$):

$$\lim_{n \to \infty} \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \ldots \left(1 - \frac{y+1}{n}\right) = 1$$

$$\lim_{n \to \infty} \left(1 - \frac{\lambda}{n}\right)^{-y} = 1$$

por la definición del número e:

$$\lim_{n \to \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

con lo cual la función masa de probabilidad queda:

$$\boxed{P_Y(y) = \begin{cases} \dfrac{\lambda^y e^{-\lambda}}{y!} & \text{para } y \geq 0 \\ 0 & y < 0 \end{cases}} \qquad \ldots (15)$$

MOMENTOS DE LA DISTRIBUCION DE POISSON

Siguiendo un procedimiento similar al de las distribuciones anteriores, la función generatriz de momentos de esta distribución es:

$$M_Y(s) = e^{\lambda(e^s - 1)}$$

de donde:

$$\mu_Y = \left.\frac{d\, M_Y(s)}{ds}\right|_{s=0} = \left[\lambda e^s \quad e^{\lambda(e^s - 1)}\right]_{s=0} = \lambda \qquad \ldots (16)$$

$$\sigma_Y^2 = \left.\frac{d^2 M_Y(s)}{ds^2}\right|_{s=0} - \mu_Y^2$$

$$= \lambda \left\{e^s \left[\lambda e^s \; e^{\lambda(e^s - 1)}\right] + e^s \; e^{\lambda(e^s - 1)}\right\}_{s=0} - \lambda^2$$

$$= \lambda^2 + \lambda - \lambda^2 = \lambda \qquad \ldots (17)$$

Ejemplo III.5

En una empresa textil se ha observado que la probabilidad de encontrar un defecto en un metro de cierto tipo de tela es 0.02. Considerando un rollo de 100 metros de dicha tela, cuál es:

a) El número promedio de defectos por rollo.

b) La probabilidad de que no exista ningún defecto en los 100 metros.

c) La F.M.P. en forma gráfica y tabular de la variable aleatoria que representa el número de defectos por rollo.

Solución

a) Considerando como éxito el encontrar un defecto en la tela, p = 0.02, con lo cual la tasa promedio de éxitos es:

$$\lambda = n\,p$$

$$= 100(0.02) = 2 \text{ defectos por rollo}$$

b) De la expresión (15):

$$P_Y(y) = \frac{(2)^0\ e^{-2}}{0!} = 0.1353$$

mediante la distribución binomial se puede aproximar este resultado como:

$$P_K(0) = C(100,\ 0)(0.02)^0(0.98)^{100} = 0.1326$$

El error que se comete con esta aproximación se debe a que se está discretizando la variable continua distancia, en intervalos de un metro, lo cual modifica las características del experimento.

c) De la expresión (15) se obtienen los valores de la función que se muestra en la siguiente figura.

y	$P_Y(y)$
0	0.135
1	0.270
2	0.270
3	0.180
4	0.090
5	0.036
6	0.012
7	0.003
8	0.001

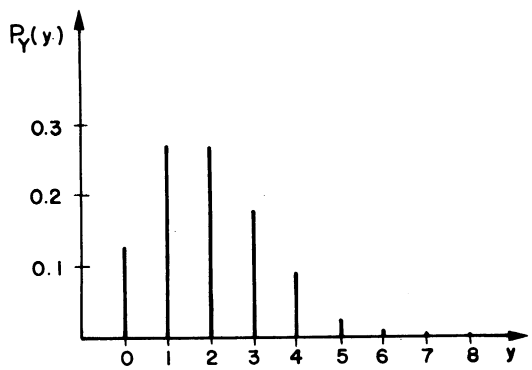


Figura III.4

Otra forma de expresar la F.M.P. de la distribución de Poisson, se obtiene al multiplicar una constante t, no negativa, por la tasa promedio de éxitos λ, con el fin de cambiar la escala de unidades. De este modo la expresión (15) queda como:

$$P_Y(y) = \begin{cases} \dfrac{(\lambda t)^y \, e^{-\lambda t}}{y!} & \text{para } y \geq 0 \\ \\ 0 & y \leq 0 \end{cases} \quad \ldots (18)$$

Ejemplo III.6

En el conmutador telefónico de una compañía se reciben en promedio dos llamadas por minuto. Suponiendo que dichas llamadas siguen un proceso de Poisson, cuál es la probabilidad de recibir:

a) Ocho llamadas en cinco minutos.

b) Ninguna llamada en tres minutos.

c) Menos de tres llamadas en dos minutos.

Solución

a) De la expresión (18) con $\lambda = 2$ y $t = 5$

$$P_Y(8) = \frac{[2(5)]^8 \; e^{-2(5)}}{8!} = 0.1126$$

b) Con $t = 3$, $\lambda t = 2(3) = 6$

$$P_Y(0) = \frac{6^0 \, e^{-6}}{0!} = 0.0024$$

c) Con $t = 2$, $\lambda t = 2(2) = 4$

$$P\,[Y < 3] = P_Y(0) + P_Y(1) + P_Y(2)$$

$$= \frac{(4)^0 \, e^{-4}}{0!} + \frac{(4)^1 \, e^{-4}}{1!} + \frac{(4)^2 \, e^{-4}}{2!}$$

$$= e^{-4} \left[1 + 4 + \frac{16}{2}\right]$$

$$= 0.2381$$

III.2 DISTRIBUCIONES TEORICAS DE VARIABLES ALEATORIAS CONTINUAS

Hasta ahora se han presentado modelos probabilísticos que se basan en los procesos de Bernoulli o Poisson y manejan variables aleatorias discretas. En esta sección se presentarán modelos probabilísticos para variables aleatorias continuas, que no se basan necesariamente en los procesos antes mencionados.

III.2.1 DISTRIBUCION UNIFORME O RECTANGULAR

Sea X una variable aleatoria que *puede tomar, con la misma probabilidad, cualquier valor en el intervalo* $[a, b]$, entonces X tiene distribución uniforme con parámetros a y b.

En la realidad existen muy pocos experimentos cuyos resultados tengan la misma probabilidad de ocurrencia ; sin embargo, en algunos casos se utiliza esta distribución cuando no se cuenta con la información necesaria y se tienen que asignar probabilidades a los eventos. Por ejemplo, la proporción de personas que consumirán un nuevo producto en una comunidad, se puede representar mediante una variable aleatoria con distribución uniforme y parámetros $a = 0$ y $b = 1$.

De acuerdo con la definición, la forma de la F.D.P. es como se muestra en la siguiente figura:

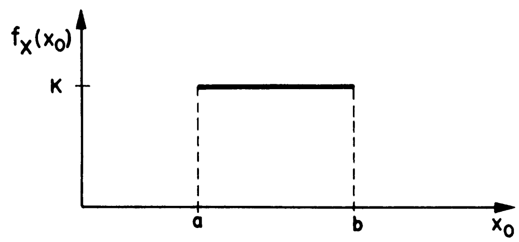


Figura III.5

donde el valor de k en términos de los parámetros a y b se obtiene a partir de la propiedad:

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

como en este caso:

$$f_X(x) = \begin{cases} k & a \leq x \leq b \\ 0 & \text{para cualquier otro caso} \end{cases}$$

entonces:

$$\int_a^b k\, dx = k\left[x\right]_a^b = k\left[b - a\right] = 1$$

con lo cual:

$$k = \frac{1}{b - a}$$

finalmente:

$$\boxed{f_X(x) = \begin{cases} \dfrac{1}{b - a} & \text{para } a \leq x \leq b \\ 0 & \text{para cualquier otro caso} \end{cases}} \quad \ldots (19)$$

MOMENTOS DE LA DISTRIBUCION UNIFORME

Utilizando la definición de la media:

$$\mu_X = E\left[X\right] = \int_{-\infty}^{\infty} x f_X(x)\, dx$$

$$= \frac{1}{b-a} \int_a^b x dx$$

$$= \frac{1}{2} \left[\frac{b^2 - a^2}{b - a}\right]$$

por lo tanto:

$$\mu_X = \frac{1}{2}(b + a) \qquad \ldots (20)$$

para la variancia:

$$\sigma_X^2 = E\left[X^2\right] - \mu_X^2$$

donde:

$$E\left[X^2\right] = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx = \frac{1}{b-a} \int_a^b x^2\, dx = \frac{b^3 - a^3}{3(b-a)}$$

$$= \frac{b^2 + ab + a^2}{3}$$

de manera que:

$$\sigma_X^2 = \frac{b^2 + ab + a^2}{3} - \left[\frac{b+a}{2}\right]^2$$

$$= \frac{(b-a)^2}{12} \qquad \ldots (21)$$

Ejemplo III.7

Un inversionista considera que al aportar cierta cantidad de dinero I en un negocio, puede perder todo lo invertido, recibir el triple de la inversión o quedar en una situación intermedia; todo con la misma posibilidad. ¿Cuál es:

a) La probabilidad de que reciba cuando menos 1.5 veces lo que invirtió.

b) La utilidad esperada (utilidad = cantidad recibida - cantidad aportada).

c) El coeficiente de variación de la cantidad que recibirá.

Solución

a) Si X representa la cantidad de dinero recibida, a = 0
b = 3I y de la expresión (19):

$$P(X \geq 1.5I) = \int_{1.5I}^{3I} \frac{1}{3I-0}\, dx$$

$$= \left[\frac{x}{3I}\right]_{1.5I}^{3I} = \frac{1.5I}{3I}$$

con lo cual:

$$P(X \geq 1.5I) = 0.5$$

b) Si la utilidad es la diferencia X - I

entonces:

$$E\left[\text{utilidad}\right] = E\left[X\right] - E\left[I\right]$$

y por la expresión 20:

$$E\left[X\right] = \frac{1}{2}(3I - 0) = 1.5I$$

por lo tanto:

$$E\left[\text{utilidad}\right] = 1.5I - I = 0.5I$$

c) Por definición:

$$CV = \frac{\sigma_X}{\mu_X}$$

y de la expresión (21):

$$\sigma_X^2 = \frac{(3I - 0)^2}{12} = \frac{9I^2}{12} = \frac{3I^2}{4}$$

entonces:

$$\sigma_X = \sqrt{\frac{3I^2}{4}} = 0.866I$$

con lo cual:

$$CV = \frac{0.866I}{0.5I} = 1.732$$

III.2.2 DISTRIBUCIONES NORMAL Y NORMAL ESTANDAR

La distribución normal o de Gauss es uno de los modelos probabilísticos de más utilidad debido entre otras razones, a que en la realidad muchas variables aleatorias pueden representarse adecuadamente mediante esta distribución, incluso otras distribuciones que se estudiarán en este capítulo toman como base a la distribución normal y además, bajo ciertas condiciones, la distribución normal se aproxima a otras distribuciones teóricas, como se verá en el capítulo V.

En forma general se dice que si la función densidad de probabilidad de una variable aleatoria X está dada por:

$$\boxed{f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(X-\mu)^2}{2\sigma^2}} \quad \text{para toda } x} \qquad \ldots (22)$$

entonces X tiene distribución normal, con parámetros μ y σ.

En la figura III.6 se muestra la forma de la F.D.P. para diferentes valores de σ. Por su forma esta curva se llama campana de Gauss.

Se puede observar que la gráfica de la función es simétrica con respecto a la media; su punto máximo está en $X = \mu$ y es asintótica con el eje X (en ambos lados).

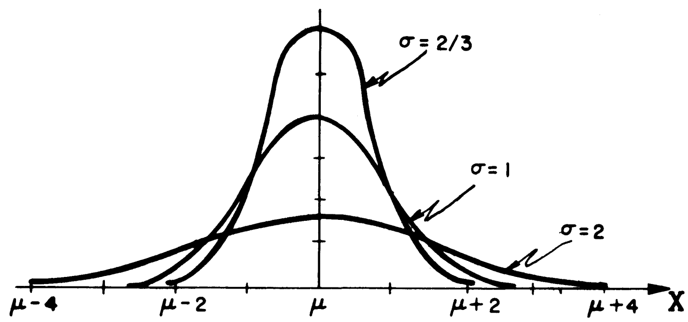


Figura III.6

MOMENTOS DE LA DISTRIBUCION NORMAL

Conforme a la definición, la función generatriz de esta distribución es:

$$M_X(s) = E\left[e^{sX}\right] = \int_{-\infty}^{\infty} e^{sx}\, \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/2\sigma^2}\, dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{sx - (x-\mu)^2/2\sigma^2}\, dx$$

al desarrollar el exponente:

$$sx - (x-\mu)^2/2\sigma^2 = -\frac{1}{2\sigma^2}\left[x^2 - 2\mu x + \mu^2 - 2\sigma^2 sx\right]$$

$$= -\frac{1}{2\sigma^2}\left[(x - \mu - \sigma^2 s)^2\right] + s\mu + \frac{\sigma^2 s^2}{2}$$

por lo cual:

$$M_X(s) = e^{s\mu + \frac{\sigma^2 s^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x - \mu - \sigma^2 s)^2}{2\sigma^2}} dx$$

Tómese en cuenta que la función que se está integrando es igual a la expresión (22) retrasada $\sigma^2 s$ unidades.

Su integral definida en el intervalo $(-\infty, \infty)$ es igual a uno; luego entonces:

$$M_X(s) = e^{\left[s\mu + \frac{\sigma^2 s^2}{2}\right]}$$

A partir de esta función, la media y la variancia de la distribución normal son respectivamente:

$$\mu_X = \left.\frac{d\, M_X(s)}{ds}\right|_{s=0} = \mu$$

$$\sigma_X^2 = \left.\frac{d^2\, M_X(s)}{ds}\right|_{s=0} - \mu^2 = \sigma^2$$

Es decir, que la media y la variancia de la distribución normal son respectivamente los parámetros μ y σ^2.

Para calcular la probabilidad de que la variable aleatoria X tome un valor dentro del intervalo $[x_1, x_2]$ se debe evaluar la integral definida:

$$P(x_1 \leq X \leq x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-(x-\mu)^2/2\sigma^2} dx$$

la cual presenta algunos problemas para obtenerla analíticamente.

Haciendo el cambio de variable (estandarizando la variable)

$$Z = \frac{x - \mu_X}{\sigma_X} \qquad \ldots (23)$$

Se puede demostrar que Z también tiene distribución normal con parámetros $\mu_Z = 0$ y $\sigma_Z = 1$. Con el cambio de variable, la F.D.P. de la expresión (22) se transforma en:

$$\boxed{f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad \text{para cualquier valor } Z} \qquad \ldots (24)$$

con lo cual:

$$P(x_1 \leq X \leq x_2) = \int_{\frac{x_1 - \mu_X}{\sigma_X}}^{\frac{x_2 - \mu_X}{\sigma_X}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

$$= P(z_1 \leq Z \leq z_2)$$

donde:

$$z_1 = \frac{x_1 - \mu_X}{\sigma_X} \quad \text{y} \quad z_2 = \frac{x_2 - \mu_X}{\sigma_X}$$

A la distribución de probabilidad de la variable Z, cuya F.D.P. está dada por la expresión (24), se le denomina normal estándar y su función de distribución acumulada es:

$$F_Z(z_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_0} e^{-\frac{z^2}{2}} dz \qquad \ldots (25)$$

la cual se encuentra tabulada para diferentes valores de Z_0 en la tabla A que se presenta en el apéndice de estos apuntes.

Para calcular la probabilidad de que Z tome un valor en cierto intervalo, se puede aprovechar la simetría de la función y el hecho de que el área bajo la curva que se encuentra a uno y otro lado del eje de simetría es 0.5.

Ejemplo III.8

Se ha observado que el tiempo necesario para que un viajero se traslade de una ciudad a otra tiene una distribución normal, con media $\mu = 13.5$ hrs., y desviación estándar $\sigma = 1$ hr. Calcular la probabilidad de que el recorrido tarde:

a) Cuando mucho 14 hrs.

b) Entre 12 y 16 hrs.

c) Menos de 12.5 o más de 15 hrs.

Solución

a) Normalizando el valor x = 14 horas con la expresión (23):

$$z = \frac{14 - 13.5}{1} = 0.5$$

con lo cual:

$$P(x \leq 14) = P(z \leq 0.5)$$

o bien:

$$F_X(14) = F_Z(0.5)$$

En la siguiente figura se muestra la relación entre las dos variables:

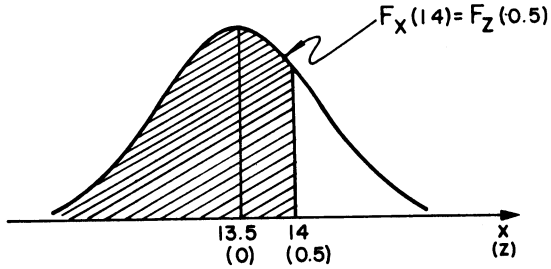


Figura III.7

buscando en la tabla A que se presenta en el apéndice de estos apuntes:

$$F_Z(0.5) = 0.6915$$

b) Estandarizando los valores:

$$z_1 = \frac{12 - 13.5}{1} = -1.5$$

$$z_2 = \frac{16 - 13.5}{1} = 2.5$$

En la siguiente figura se muestra la probabilidad deseada.

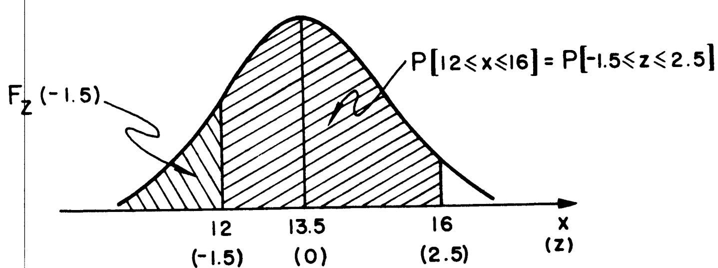


Figura III.8

$P\left[-1.5 \leq Z \leq 2.5\right] = F_Z(2.5) - F_Z(-1.5)$

buscando en la tabla A de la distribución normal estándar:

$P\left[-1.5 \leq Z \leq 2.5\right] = 0.9938 - 0.0668 = 0.927$

Debe notarse que por la simetría de la función:

$F_Z(-1.5) = 1 - F_Z(1.5)$

c) **Para este inciso:**

$$z_1 = \frac{12.5 - 13.5}{1} = -1$$

$$z_2 = \frac{15 - 13.5}{1} = 1.5$$

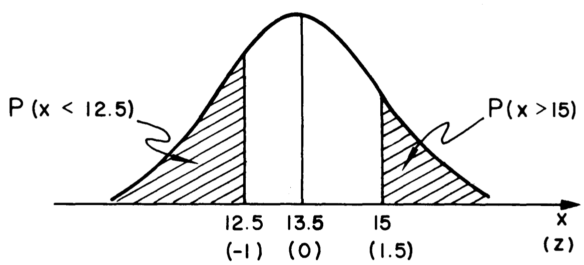


Figura III.9

De la figura III.9:

$P(z > 1.5) = 1 - P(z \leq 1.5)$

con lo cual:

$P(z < -1) + P(z > 1.5) = F_Z(-1) + \left[1 - F_Z(1.5)\right]$

buscando nuevamente en la misma tabla:

$P\left[z < -1\right] + P\left[z > 1.5\right] = 0.1587 + \left[1 - 0.9332\right]$

$= 0.2255$

III.2.3 DISTRIBUCION EXPONENCIAL

Sea T una variable aleatoria que representa el *tiempo que transcurre entre la ocurrencia de dos éxitos consecutivos de un proceso Poisson*, con una tasa promedio de éxitos λ; entonces T tiene distribución exponencial con parámetro λ.

Puesto que T es una variable continua y la función densidad de probabilidad no relaciona directamente probabilidades a los valores de la variable independiente, se obtendrá primero la función de distribución acumulada, que de acuerdo con la definición es:

$$F_T(t_0) = P(T \leq t_0)$$

o bien:

$$F_T(t_0) = 1 - P(T > t_0)$$

donde $P(T > t_0)$ se obtiene al considerar la probabilidad de que no ocurra ningún éxito en un intervalo de tiempo mayor a t_0, que equivale al valor $P_Y(0)$ calculado mediante la distribución de Poisson, con lo cual:

$$F_T(t_0) = \begin{cases} 1 - e^{-\lambda t_0} & ; \quad t_0 \geq 0 \\ 0 & ; \quad t_0 < 0 \end{cases} \qquad \ldots (26)$$

que define a la F.D.A. de la distribución exponencial.

A partir de esta función y conforme a lo visto en el capítulo anterior, la función densidad de probabilidad de la distribución exponencial es:

$$f_T(t) = \frac{d\, F_T(t)}{dt}$$

con lo cual:

$$\boxed{f_T(t) = \begin{cases} \lambda e^{-\lambda t} & \text{para} \quad t \geq 0 \\ 0 & \qquad\quad t < 0 \end{cases}} \qquad \ldots (27)$$

MOMENTOS DE LA DISTRIBUCION EXPONENCIAL

La función generatriz de momentos es de acuerdo con su definición:

$$M_T(s) = \int_{-\infty}^{\infty} e^{st} f_T(t)\, dt$$

Al sustituir la expresión (27):

$$M_T(s) = \lambda \int_0^{\infty} e^{-\lambda t}\, e^{st}\, dt$$

$$= \lambda \int_0^{\infty} e^{-t(\lambda - s)}\, dt$$

$$= \lambda \left[\frac{e^{-t(\lambda - s)}}{-(\lambda - s)} \right]_0^{\infty}$$

evaluando para los límites de la expresión:

$$M_T(s) = \frac{\lambda}{\lambda - s}$$

como se ha venido haciendo, a partir de $M_T(s)$ pueden calcularse la media y la variancia de esta distribución.

Para la media se tiene:

$$\mu_T = \left. \frac{d\, M_T(s)}{ds} \right|_{s=0} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda} \qquad \ldots (28)$$

y para la variancia:

$$\sigma_T^2 = E\,[X^2] - \mu_T^2 = \left. \frac{d^2 M_T(s)}{ds^2} \right|_{s=0} - \mu_T^2$$

donde:

$$\left. \frac{d^2 M_T(s)}{ds^2} \right|_{s=0} = \left. \frac{-\,2\lambda(\lambda - s)}{(\lambda - s)^4} \right|_{s=0} = \frac{2\lambda^2}{\lambda^4} = \frac{2}{\lambda^2}$$

con lo cual:

$$\sigma_T^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \qquad \ldots (29)$$

Ejemplo III.9

A un centro comercial llegan en promedio tres personas por minuto.

1. Calcular la probabilidad de que el tiempo entre dos llegadas consecutivas sea:

a) Menor de medio minuto.

b) Mayor de 2 minutos.

c) Exactamente de un minuto.

2. Obtener el tiempo promedio entre llegadas y su desviación estándar.

Solución

a) Como $\lambda = 3$ clientes por minuto; $t_0 = 0.5$ min y de la expresión (26):

$$F_T(0.5) = 1 - e^{-3(0.5)} = 0.7769$$

b) $P(T > 2) = 1 - F_T(2) = 1 - 1 + e^{-3\ (2)} = 0.0024$

c) Debido a que T es una variable continua:

$$P(T = 1) = 0$$

2. De (28) y (29):

$$\mu_T = \frac{1}{3} = 0.333 \text{ minutos}$$

$$\sigma_T^2 = \frac{1}{(3)^2} = 0.111$$

con lo cual:

$$\sigma_T = \sqrt{0.111} = 0.333 \text{ minutos}$$

Como se ha presentado la distribución exponencial, la variable aleatoria T representa el tiempo que transcurre hasta la primera ocurrencia de un evento o bien el tiempo que transcurre entre dos eventos consecutivos.

Por ese motivo se utiliza frecuentemente para representar el tiempo entre la llegada de clientes o el tiempo entre la ocurrencia de fallas en un sistema productivo. Sin embargo, también puede representar a otras variables continuas, como la distancia entre dos vehículos que circulan en una avenida.

III.2.4 DISTRIBUCION GAMMA O ERLANG-K

Si en un proceso Poisson, la variable aleatoria T representa el *tiempo que transcurre hasta que ocurren* k *éxitos*, con una tasa promedio de éxitos λ, entonces T tiene distribución gamma o Erlang-k con parámetros λ y k.

La F.D.P. de la distribución gamma se obtiene al considerar a T como la suma de K variables aleatorias independientes $T_1, T_2, .., T_K$ con distribución exponencial, esto es:

$$T = T_1 + T_2 + ... + T_K$$

Mediante un procedimiento que involucra la aplicación repetida de la integral de convolución se obtiene:

$$f_T(t) = \begin{cases} \dfrac{\lambda^k (\lambda t)^{k-1} e^{-\lambda t}}{(k-1)!} & \text{para} \quad t \geq 0 \\ 0 & t < 0 \end{cases} \qquad \text{... (30)}$$

En forma más general, la F.D.P. de la distribución gamma se presenta como:

$$\boxed{f_T(t) = \begin{cases} \dfrac{\lambda^k (\lambda t)^{k-1} e^{-\lambda t}}{\Gamma(k)} & \text{para} \quad t \geq 0 \\ 0 & t < 0 \end{cases}} \qquad \text{... (31)}$$

donde $\Gamma(k)$ es la función gamma definida como:

$$\Gamma(k) = \int_0^{\infty} e^{-u} \, u^{(k-1)} \, du$$

La constante k no se restringe a tomar valores enteros aunque sí debe ser positiva y la variable T no representa necesariamente el intervalo de tiempo pará que se presente el K-ésimo evento.

MOMENTOS DE LA DISTRIBUCION GAMMA

La media y la variancia se obtienen como el primer momento con respecto al origen y el segundo con respecto a la media, con lo cual:

$$\mu_T = \frac{k}{\lambda} \qquad \text{... (32)}$$

$$\sigma_T^2 = \frac{k}{\lambda^2} \qquad \text{... (33)}$$

Ejemplo III.10

En una región del país se han registrado 90 sismos de una intensidad considerable en los últimos 50 años. Suponiendo que las condiciones geológicas no han cambiado en forma significativa y que la ocurrencia de un sismo es independiente de lo que ha sucedido en el pasado, cuál es:

a) La gráfica de la F.D.P. de la variable que representa al tiempo necesario para que ocurran tres sismos.

b) La probabilidad de que se tenga que esperar cuando mucho tres años para registrar tres sismos.

Solución

a) La tasa promedio de sismos se puede estimar como:

$$\lambda = \frac{90}{50} = 1.8 \text{ sismos/año}$$

de la expresión (30):

$$f_T(t) = \frac{(1.8)^3\, t^{3-1}\, e^{-1.8t}}{(3-1)!} = 2.916\, t^2\, e^{-1.8t}$$

al variar t se obtiene la tabla y la figura siguiente:

t	$f_T(t)$
0	0
0.5	0.2964
1	0.4820
1.5	0.4409
2	0.3187
2.5	0.2024
3	0.1185
3.5	0.0656
4	0.0348

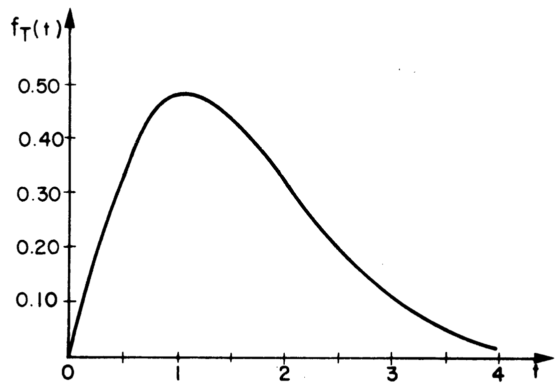


Figura III.10

b) Integrando la F.D.P. en el intervalo $[0, 3]$

$$F_T(3) = \int_0^3 2.916\, t^2\, e^{-1.8t}\, dt$$

Mediante la fórmula de integración numérica de Simpson:

$$= \frac{h}{3}\left[f(t_0) + f(t_n) + 2\,\Sigma \begin{array}{c}\text{ordenadas con}\\ \text{índice par}\end{array} + 4\,\Sigma \begin{array}{c}\text{ordenadas con}\\ \text{índice impar}\end{array}\right]$$

$$F_T(3) = \frac{(0.5)}{3}\Big[0 + 0.0348 + 2(0.4820 + 0.3187 + 0.1185) +$$

$$+ 4(0.2964 + 0.4409 + 0.2024 + 0.0656)\Big]$$

$$= \frac{0.5}{3}\left[0.0348 + 2(0.9192) + 4(1.0053)\right]$$

$$= 0.9824$$

La función de distribución acumulada de la distribución gamma, se encuentra tabulada para algunos valores de k, en la tabla correspondiente a la distribución ji - cuadrada (χ^2), que como se verá a continuación es un caso particular de la distribución gamma.

III.2.5 DISTRIBUCION JI-CUADRADA O CHI-CUADRADA

Cuando una variable aleatoria χ^2 está formada por la *suma de los cuadrados de* ν *variables aleatorias independientes con distribución normal*, entonces se dice que esta variable χ^2 tiene distribución ji-cuadrada con ν grados de libertad.

Su función densidad de probabilidad está definida por:

$$f_{\chi^2}(x) = \begin{cases} \dfrac{1/2\ (x/2)^{\nu/2\,-1}\ e^{-x/2}}{2^{\nu/2}\quad \Gamma(\nu/2)} & \text{para}\quad x \geq 0 \\ \\ 0 & \text{para}\quad x < 0 \end{cases} \qquad \ldots (34)$$

Comparando esta función con la expresión (31) se puede observar que la distribución ji-cuadrada es un caso particular de la distribución gamma, en la cual:

$$\lambda = \frac{1}{2} \quad y \quad k = \frac{\nu}{2}$$

Los momentos de la distribución ji-cuadrada se pueden obtener de las expresiones (32) y (33), haciendo los cambios mencionados, con lo cual:

$$\mu_{\chi^2} = \nu \qquad \ldots (35)$$

$$\sigma_{\chi^2} = 2\nu \qquad \ldots (36)$$

La función de distribución acumulada de esta distribución se encuentra tabulada para diferentes valores de ν en la tabla B del apéndice de estos apuntes.

La mayoría de las aplicaciones que tiene la distribución χ^2 se encuentran en la estadística y para facilitar su manejo en esa área, la tabla B del apéndice presenta los valores de la función de distribución acumulada $F_{\chi^2}(x)$.

De esta manera, si se desea encontrar el valor $\chi^2 = x$ tal que $F_{\chi^2}(x) = 0.05$ con $\nu = 10$ grados de libertad, se debe buscar en la tabla correspondiente la intersección entre la columna de valores encabezada por 0.05 y el renglón que muestra en el margen izquierdo el valor 10, donde se encuentra el valor $\chi^2 = 3.94$.

La forma de la distribución para algunos valores de ν se muestra en la siguiente figura en donde se observa que conforme aumenta ν, la distribución se aproxima a la forma de la distribución normal.

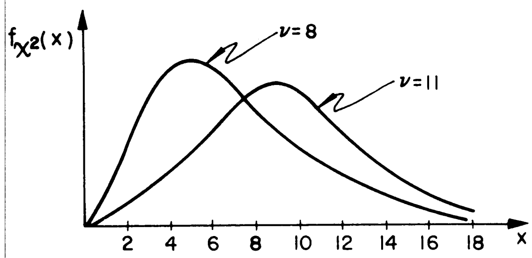


Figura III.11

III.2.6 DISTRIBUCION t DE STUDENT

Fue creada por W.S. Gosset bajo el seudónimo de *Student* (estudiante) y posteriormente se desarrolló de manera más rigurosa por R.A. Fisher.

Las aplicaciones más importantes de esta distribución también se encuentran en la inferencia estadística.

Una variable aleatoria T, tiene distribución t de student (o simplemente t), si está formada por el cociente:

$$T = \frac{Z}{V/\sqrt{\nu}} \qquad \dots (37)$$

donde Z es una variable aleatoria con distribución normal estándar; V es otra variable aleatoria con distribución ji-cuadrada y ν son los grados de libertad de la variable aleatoria V.

La función densidad de probabilidad de la distribución t de student, está definida por:

$$\boxed{f_T(t) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)\sqrt{\pi\nu}}\left(1+\frac{t^2}{\nu}\right)^{\left(-\frac{\nu+1}{2}\right)} \quad \text{para } -\infty < t < \infty} \qquad \dots (38)$$

En la figura III.12 se muestra la forma de esta distribución para diferentes valores de ν.

En ella se observa que conforme aumenta el parámetro ν la distribución tiende a la forma de la normal estándar.

La media y la variancia de esta distribución para $\nu > 2$ son:

$$\mu_T = 0 \qquad \ldots (39)$$

$$\sigma_T^2 = \frac{\nu}{\nu - 2} \qquad \ldots (40)$$

Como en la distribución ji-cuadrada, la función de distribución acumulada se encuentra tabulada en la tabla C del apéndice para hallar los valores de T a partir de los grados de libertad y del valor $F_T(t)$.

La manera de buscarlos es igual que en el caso de la distribución ji-cuadrada, con la única diferencia de que en esta tabla sólo se encuentran los valores de t para $F_T(t)$ entre 0.5 y 1, debido a la simetría de la curva.

Ejemplo III.11

Si se desea obtener el valor de t, tal que $F_T(t) = 0.975$ con $\nu = 15$ grados de libertad, se encontrará que $t = 2.13$.

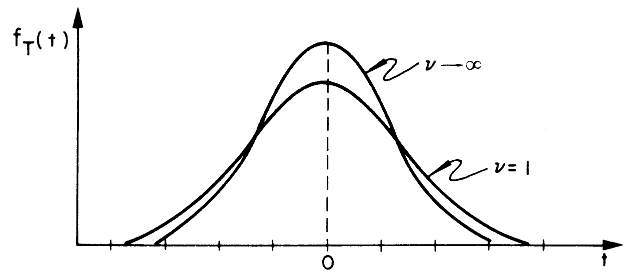


Figura III.12

Las aplicaciones de la distribución normal estándar, ji-cuadrada y t de student se presentarán en el capítulo VII que corresponde a inferencia estadística. También en el capítulo V se profundizará sobre algunos aspectos teóricos de las variables aleatorias que se forman a partir de una combinación lineal de otras.

CAPITULO IV VARIABLES ALEATORIAS CONJUNTAS

INTRODUCCION

En los capítulos anteriores se han estudiado las variables aleatorias de manera independiente; sin embargo en un sistema físico, económico o social, de naturaleza aleatoria, existen generalmente interrelaciones entre sus elementos.

Debido a estas interrelaciones, en los modelos probabilísticos deben tomarse en cuenta todas las variables relevantes en forma conjunta. Por ejemplo, si se estudia un sistema de control de inventarios para materias primas en determinada compañía, se deben considerar simultáneamente las variables que representan la demanda de los productos, su precio, el tiempo que tardan los proveedores en surtir un pedido y otras variables que permitan establecer la cantidad de productos que deben mantenerse en el almacén o el tamaño de los pedidos.

Este capítulo tiene por objeto generalizar los conceptos vistos en el capítulo II para el caso de dos o más variables aleatorias.

IV.1 FUNCIONES DE PROBABILIDAD CONJUNTA

Cuando a cada punto del espacio muestral se le asocian dos o más variables aleatorias, entonces dichas variables aleatorias se denominan *conjuntas*. Estas variables pueden ser discretas, continuas o mixtas, lo cual determina el tipo de función que representa su comportamiento.

Se estudiará primero el caso de dos variables aleatorias y posteriormente se establecerá el caso general de n variables.

FUNCIONES MASA DE PROBABILIDAD CONJUNTA Y MARGINAL

Sean X y Y dos variables aleatorias discretas, entonces la función masa de probabilidad conjunta de X y Y está definida como:

$$\boxed{P_{XY}(x, y) = P\,[(X = x) \cap (Y = y)] \;;\; \forall\, x, y} \qquad \ldots (1)$$

esto significa que la F.M.P. conjunta relaciona el punto (x, y) del espacio muestral con su probabilidad de ocurrencia.

Ejemplo IV.1

En la planeación de un proyecto se han encontrado dos actividades consecutivas, consideradas como las más importantes por la influencia que ejercen sobre el tiempo total del proyecto, para fines prácticos, dichos tiempos se pueden considerar en forma discreta, con una función masa de probabilidad conjunta dada por:

$$P_{XY}(x, y) = \begin{cases} \dfrac{(8 - x)(8 - y)}{144} & \text{para } x = 3, 4, 5 \quad [\text{semanas}] \\ & \text{para } y = 3,4,5 \\ 0 & \text{para cualquier otro caso} \end{cases}$$

a) Representar la F.M.P. conjunta en forma tabular y gráfica.

b) ¿Cuál es la probabilidad de que se terminen las dos actividades en ocho semanas?

Solución

a) Tabulando y graficando $P_{XY}(x, y)$ se tiene:

y \ x	3	4	5
3	0.1736	0.1389	0.1042
4	0.1389	0.1111	0.0833
5	0.1042	0.0833	0.0625

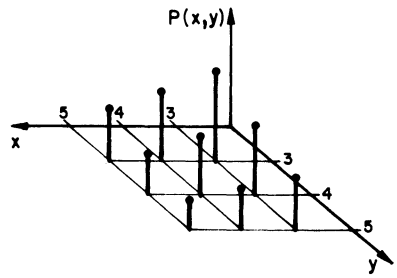


Figura IV.1

b) Para que las actividades se terminen en ocho semanas se pueden presentar los siguientes eventos:

$(x = 3, y = 5)$, $(x = 4, y = 4)$ ó $(x = 5, y = 3)$

por lo cual:

$$P(x + y = 8) = P_{XY}(3, 5) + P_{XY}(4, 4) + P_{XY}(5, 3)$$
$$= 0.1042 + 0.1111 + 0.1042 = 0.3195$$

PROPIEDADES DE LA FUNCION MASA DE PROBABILIDAD CONJUNTA

Para que una función discreta sea una F.M.P. conjunta es necesario que cumpla con las siguientes propiedades:

1) $P_{XY}(x, y) \geq 0$ ... (2)

2) $$\sum_{\forall x_i} \sum_{\forall y_i} p_{XY}(x_i, y_i) = 1 \qquad \ldots (3)$$

A las funciones masa de probabilidad de cada una de las variables aleatorias que se estudian en forma conjunta se les llama *función masa de probabilidad marginal.*

Para obtener estas funciones a partir de la F.M.P. conjunta se tiene que:

$$p_X(x) = P(X = x) = P\left[(x, y_1) \cup (x, y_2) \cup (x, y_3) \cup \cdots \cup (x, y_n)\right]$$

$$= p_{XY}(x, y_1) + p_{XY}(x, y_2) + p_{XY}(x, y_3) + \ldots + p_{XY}(x, y_n)$$

$$= \sum_{\forall y} p_{XY}(x, y) \qquad \ldots (4)$$

El conjunto de estas $p_X(x)$, para todos los valores de x, forma la F.M.P. marginal de X.

En forma similar, para Y

$$p_Y(y) = P(Y = y) = \sum_{\forall x} p_{XY}(x, y) \qquad \ldots (5)$$

Ejemplo IV.2

Obtener las funciones masa de probabilidad marginal de X y de Y para el ejemplo IV.1.

Solución

De la expresión (4):

$$p_X(x) = \sum_{y=3}^{5} \frac{(8 - x)(8 - y)}{144} = \frac{8 - x}{144} \sum_{y=3}^{5} (8 - y)$$

$$= \frac{8 - x}{144} \left[(8 - 3) + (8 - 4) + (8 - 5)\right] = \frac{(8 - x)12}{144} = \frac{8 - x}{12}$$

Para la variable Y, de la expresión (5):

$$p_Y(y) = \sum_{x=3}^{5} \frac{(8 - x)(8 - y)}{144}$$

siguiendo un razonamiento similar al de X:

$$p_Y(y) = \frac{8 - y}{12}$$

La F.M.P. marginal también se puede obtener en forma tabular a partir del cuadro de la figura IV.2, sumando las probabilidades por renglones y por columnas, esto es:

y \ x	3	4	5	$p_Y(y)$
3	0.1736	0.1389	0.1042	0.4167
4	0.1389	0.1111	0.0833	0.3333
5	0.1042	0.0833	0.0625	0.2500
$p_X(x)$	0.4167	0.3333	0.2500	1

Figura IV.2

FUNCIONES DENSIDAD DE PROBABILIDAD CONJUNTA Y MARGINAL

Sean X y Y dos variables aleatorias continuas. Entonces la función densidad de probabilidad conjunta de X y Y, denotada por $f_{XY}(x, y)$, es aquella que relaciona cada punto del plano X - Y con su correspondiente densidad de probabilidad.

Debe notarse que esta función no proporciona una probabilidad directamente, pues la probabilidad $P(X = x, Y = y)$ es prácticamente cero cuando las variables son continuas.

De manera similar al caso de una variable, en el cual la probabilidad de que X tome un valor en el intervalo $[a, b]$ se obtiene como:

$$P(a \leq X \leq b) = \int_a^b f_X(x)\, dx$$

para dos variables conjuntas X y Y, la probabilidad de que sus valores estén comprendidos en la región $R = \{(x, y)/a \leq X \leq b;\ c \leq Y \leq d\}$ se obtiene de la siguiente forma:

$$\boxed{P(a \leq X \leq b,\ c \leq Y \leq d) = \int_a^b \int_c^d f_{XY}(x, y)\, dy\, dx} \quad \ldots (6)$$

Gráficamente corresponde al volumen bajo la superficie descrita por la función densidad de probabilidad conjunta en la región R, como se muestra en la siguiente figura.

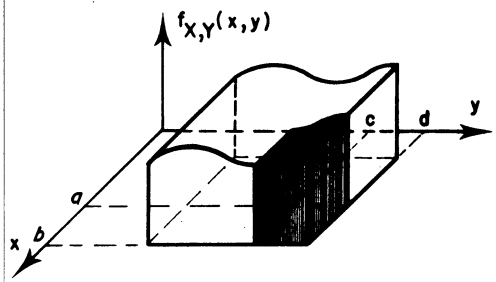


Figura IV.3

PROPIEDADES DE LA FUNCION DENSIDAD DE PROBABILIDAD CONJUNTA

Como en el caso de la función masa de probabilidad conjunta, la función densidad de probabilidad conjunta también tiene que estar definida de acuerdo con los axiomas básicos de la probabilidad, es decir:

$$f_{XY}(x, y) \geq 0 \quad \forall \text{ X y Y} \qquad \ldots (7)$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy\, dx = 1 \qquad \ldots (8)$$

Ejemplo IV.3

Un diseñador de muebles infantiles ha descubierto que la edad y estatura de los niños menores de dos años, se puede modelar con la siguiente función densidad de probabilidad conjunta:

$$f_{XY}(x, y) = \begin{cases} xy - 2y - x + 2 & \text{para: } 0 \leq x \leq 2 \quad \text{donde x: edad} \\ & \phantom{\text{para: }} 0 \leq y \leq 1 \quad \text{y: estatura} \\ 0 & \text{para cualquier otro caso} \end{cases}$$

a) Verificar que cumple con las expresiones (7) y (8).

b) ¿Cuál es la probabilidad de que un niño seleccionado al azar, sea menor de 1 año y tenga una estatura inferior a los 80 centímetros?

Solución

a) En la figura IV.4 se puede observar que se cumple la propiedad (7) y, por otra parte, como la integral de la función fuera de la región definida es cero:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy\, dx = \int_0^2 \int_0^1 (xy - 2y - x + 2)\, dy\, dx$$

resolviendo la doble integral:

$$\int_0^2 \left[\frac{xy^2}{2} - \frac{2y^2}{2} - xy + 2y\right]_0^1 dx = \int_0^2 \left(\frac{x}{2} - 1 - x + 2\right) dx$$

$$= \left[\frac{x^2}{4} - \frac{x^2}{2} + x\right]_0^2 = \frac{4}{4} - \frac{4}{2} + 2 = 1$$

lo cual verifica la segunda propiedad.

b) De la expresión (6):

$$P(0 \leq x \leq 1;\ 0 \leq y \leq 0.8) = \int_0^1 \int_0^{0.8} (xy - 2y - x + 2)\, dy\, dx$$

$$= \int_0^1 \left[\frac{xy^2}{2} - \frac{2y^2}{2} - xy + 2y\right]_0^{0.8} dx$$

$$= \int_0^1 (0.32x - 0.64 - 0.8x + 1.6)\, dx$$

$$= \int_0^1 (0.96 - 0.48x)\, dx$$

$$= \left[0.96x - \frac{0.48x^2}{2}\right]_0^1 = 0.72$$

Gráficamente se puede observar en la siguiente figura:

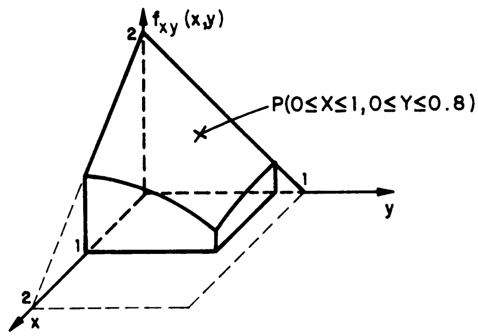


Figura IV.4

Para el caso de variables aleatorias continuas, las funciones densidad de probabilidad marginales se pueden obtener a partir de la F.D.P. conjunta, en forma análoga al caso discreto, esto es:

$$\boxed{f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy} \qquad \dots (9)$$

$$\boxed{f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx} \qquad \dots (10)$$

Ejemplo IV.4

Obtener las funciones densidad de probabilidad marginal para las variables aleatorias X y Y del ejemplo IV.3.

Solución

Para X, de la expresión (9):

$$f_X(x) = \int_{-\infty}^{\infty} (xy - 2y - x + 2)dy = \int_0^1 (xy - 2y - x + 2)dy$$

$$= \left[\frac{xy^2}{2} - \frac{2y^2}{2} - xy + 2y\right]_0^1$$

$$= \frac{x}{2} - x + 1 = 1 - \frac{x}{2}$$

entonces:

$$f_X(x) = \begin{cases} 1 - \frac{x}{2} & \text{para } 0 \leq x \leq 2 \\ 0 & \text{para cualquier otro caso} \end{cases}$$

para Y, de la expresión (10):

$$f_Y(y) = \int_{-\infty}^{\infty} (xy - 2y - x + 2)dy = \int_0^2 (xy - 2y - x + 2)dx =$$

$$= \left[\frac{x^2y}{2} - 2xy - \frac{x^2}{2} + 2x\right]_0^2 = 2y - 4y - 2 + 4 = 2 - 2y$$

entonces:

$$f_Y(y) = \begin{cases} 2(1 - y) & \text{para } 0 \le y \le 1 \\ 0 & \text{para cualquier otro caso} \end{cases}$$

IV.2 FUNCION DE DISTRIBUCION ACUMULADA CONJUNTA

También puede hacerse la generalización del caso de una variable para definir la función de distribución acumulada conjunta a partir del evento:

$$F_{XY}(x_0, y_0) = P(X \le x_0, Y \le y_0)$$

con lo cual se tiene para el caso discreto que:

$$\boxed{F_{XY}(x_0, y_0) = \sum_{\forall x \le x_0} \sum_{y \le y_0} P_{XY}(x_0, y_0)} \quad \ldots (11)$$

y en el caso continuo:

$$\boxed{F_{XY}(x_0, y_0) = \int_{-\infty}^{x_0} \int_{-\infty}^{y_0} f_{XY}(x, y)dy\, dx} \quad \ldots (12)$$

Ejemplo IV.5

Obtener la función de distribución acumulada del ejemplo IV.2.

Solución

Como es una función discreta, de la expresión (11):

$$F_{XY}(x_0, y_0) = \sum_{x=3}^{x_0} \sum_{y=3}^{y_0} P_{XY}(x, y)$$

dando valores a x_0 y y_0 se obtienen los valores mostrados en la tabla de la figura IV.5; por ejemplo si $x_0 = 4$ y $y_0 = 4$

$$F_{XY}(4, 4) = \sum_{x=3}^{4} \sum_{y=3}^{4} P_{XY}(x, y)$$

$$= \sum_{x=3}^{4} \left[P_{XY}(x,\ 3) + P_{XY}(x,\ 4) \right]$$

$$= P_{XY}(3,\ 3) + P_{XY}(3,\ 4) + P_{XY}(4,\ 3) + P_{XY}(4,\ 4)$$

$$= 0.1736 + 0.1389 + 0.1389 + 0.1111 = 0.5625$$

En la figura IV.5 se puede observar una representación gráfica de la función de distribución acumulada conjunta.

y \ x	3	4	5
3	0.1736	0.3125	0.4167
4	0.3125	0.5625	0.7500
5	0.4167	0.7500	1.000

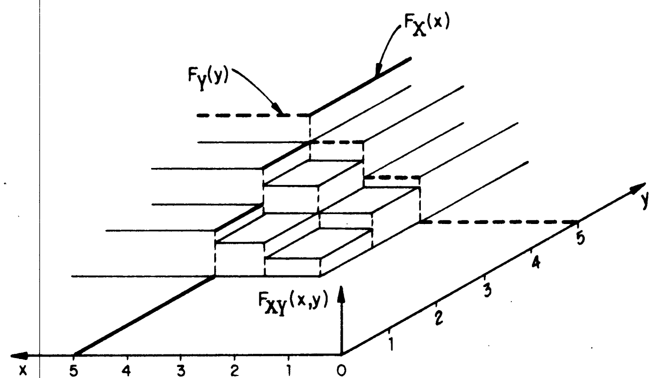


Figura IV.5

PROPIEDADES DE LA FUNCION DE DISTRIBUCION ACUMULADA CONJUNTA

Entre las principales propiedades de esta función se encuentran las siguientes*:

1. Es monótona no decreciente en X y en Y. ... (13)

2. $\lim_{x \to -\infty} F_{XY}(x,\ y) = \lim_{y \to \infty} F_{XY}(x,\ y) = 0$ ... (14)

* La demostración de estas propiedades se encuentra en el libro de Iván Obregón Sanín, *Teoría de la probabilidad*, Editorial Limusa, México, 1977, páginas 157 y 158.

3. $\lim\limits_{\substack{x \to \infty \\ y \to \infty}} F_{XY}(x, y) = 1$ ... (15)

En la figura IV.5 se observan las propiedades anteriores puesto que la función vale cero cuando cualquiera de las dos variables sea menor que tres; vale uno cuando las dos variables son mayores que cinco y no decrece cuando aumentan los valores de las variables:

$$F_{XY}(x, y) = 0 \quad \forall \; x, y < 3$$

$$F_{XY}(x, y) = 1 \quad \forall \; x, y > 5$$

$$F_{XY}(x_1, y_1) \leq F_{XY}(x_2, y_2) \quad \forall \; \begin{matrix} x_1 \leq x_2 \\ y_1 \leq y_2 \end{matrix}$$

FUNCIONES DE DISTRIBUCION ACUMULADA MARGINALES

Para obtener las F.D.A. marginales de dos variables aleatorias conjuntas X y Y, continuas o discretas, a partir de la F.D.A. conjunta, se utilizan las siguientes propiedades:

$$\boxed{\lim_{x \to \infty} F_{XY}(x, y) = F_Y(y)} \quad \text{... (16)}$$

$$\boxed{\lim_{y \to \infty} F_{XY}(x, y) = F_X(x)} \quad \text{... (17)}$$

En la figura IV.5 se observa que $F_{XY}(x, 5) = F_X(x)$ y $F_{XY}(5, y) = F_Y(y)$

Gráficamente $F_X(x)$ corresponde a la curva seccionalmente continua que se define por la intersección del plano $y = 5$ y la función $F_{XY}(x, y)$. En forma análoga $F_Y(y)$, corresponde a la curva seccionalmente continua definida por la intersección del plano $x = 5$ y la función $F_{XY}(x, y)$.

Ejemplo IV.6

A partir de la siguiente función densidad de probabilidad conjunta:

$$f_{XY}(x, y) = \begin{cases} 0.1 & \text{para } \begin{matrix} 0 \leq x \leq 5 \\ 0 \leq y \leq 2 \end{matrix} \\ 0 & \text{para cualquier otro caso} \end{cases}$$

obtener:

a) La F.D.A. conjunta.

b) La probabilidad de que las variables X y Y, tomen un valor comprendido en la región.

$$A = \{(x, y) / \; x \leq 3 \;;\; y \leq 1.5\}$$

c) Las F.D.A. marginales de X y de Y.

Solución

a) De la expresión (12):

$$F_{XY}(x_0,\ y_0) = \int_{-\infty}^{x_0}\int_{-\infty}^{y_0} 0.1\ dy\ dx = \int_0^{x_0}\int_0^{y_0} 0.1\ dy\ dx$$

resolviendo la integral:

$$F_{XY}(x_0,\ y_0) = 0.1\ x_0\ y_0$$

b) Por la definición de la F.D.A. conjunta:

$$P(X \leq 3,\ Y \leq 1.5) = F_{XY}(3,\ 1.5) = 0.1(3)(1.5) = 0.45$$

c) De las propiedades (16) y (17):

$$F_X(x) = \lim_{y \to \infty} (0.1\ xy) \qquad F_Y(y) = \lim_{x \to \infty} (0.1\ xy)$$

como $F_{XY}(x,\ y)$ no cambia para valores de y mayores que dos o valores de x mayores que cinco:

$$F_X(x) = (0.1)(2)\ x = 0.2x \qquad F_Y(y) = (0.1)(5)\ y = 0.5y$$

La figura IV.6 muestra la función de distribución acumulada conjunta y en ella se señalan las funciones marginales (en las aristas).

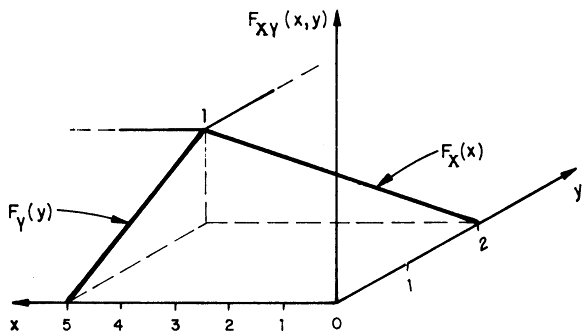


Figura IV.6

IV.3 DISTRIBUCIONES CONDICIONALES

La probabilidad condicional de que ocurra un evento A, dado que ha ocurrido otro evento B, fue definida en el capítulo I como:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Al considerar dos variables aleatorias discretas X, Y y dos eventos A, B tales que $A = \{X = x\}$ y $B = \{Y = y\}$, se define la función masa de probabilidad condicional de X dado Y como:

$$P_{X/Y}(x, y) = P(X = x / Y = y) = \frac{P_{XY}\left[(X = x) \cap (Y = y)\right]}{P_Y(Y = y)} = \frac{P_{XY}(x, y)}{P_Y(y)} \quad \ldots (18)$$

también se define la función masa de probabilidad condicional de Y dado X como:

$$P_{Y/X}(y, x) = P(Y = y / X = x) = \frac{P_{XY}(x, y)}{P_X(x)} \quad \ldots (19)$$

de manera semejante, para las variables continuas se tienen las siguientes funciones densidad de probabilidad condicional:

$$f_{X/Y}(x, y) = \frac{f_{XY}(x, y)}{f_Y(y)} \quad \ldots (20)$$

$$f_{Y/X}(y, x) = \frac{f_{XY}(x, y)}{f_X(x)} \quad \ldots (21)$$

Ejemplo IV.7

La función masa de probabilidad conjunta de dos variables aleatorias X, Y, está dada por:

$$P_{XY}(x, y) = \begin{cases} \dfrac{2x + y}{42} & \text{para} \quad x = 0, 1, 2 \\ & \qquad\quad\; y = 0, 1, 2, 3 \\ 0 & \text{para cualquier otro caso} \end{cases}$$

a) Obtener la función $P_{Y/X}(y, 2)$

b) Calcular el valor de $P_{X/Y}(1, 3)$

Solución

a) Para poder calcular la función masa de probabilidad condicional, es necesario conocer la función $P_X(x)$, por lo cual de la expresión (4):

$$P_X(x) = \sum_{y=0}^{3} \frac{2x + y}{42} = \frac{2}{42} \sum_{y=0}^{3} x + \frac{1}{42} \sum_{y=0}^{3} y$$

$$= \frac{1}{21}\left[4x\right] + \frac{1}{42}\left[6\right] = \frac{8x + 6}{42}$$

con este resultado, de la expresión (19):

$$P_{Y/X}(y, x) = \frac{(2x + y)/42}{(8x + 6)/42} = \frac{2x + y}{8x + 6}$$

b) en el caso particular en el cual x = 2:

$$p_{Y/X}(y,\ 2) = \frac{2(2) + y}{8(2) + 6} = \frac{4 + y}{22}$$

b) Para obtener $p_Y(y)$ de la expresión (5):

$$p_Y(y) = \sum_{x=0}^{2} \frac{2x + y}{42} = \frac{2}{42} \sum_{x=0}^{2} x + \frac{1}{42} \sum_{x=0}^{2} y$$

$$= \frac{6 + 3y}{42} = \frac{2 + y}{14}$$

de la expresión (18) se obtiene la forma general de la función como:

$$p_{X/Y}(x,\ y) = \frac{\dfrac{2x + y}{42}}{\dfrac{2 + y}{14}} = \frac{2x + y}{3(2 + y)} = \frac{2x + y}{6 + 3y}$$

para el caso particular en el cual (x = 1, y = 3)

$$p_{X/Y}(1,\ 3) = \frac{2(1) + 3}{6 + 3(3)} = 0.333$$

VARIABLES ALEATORIAS INDEPENDIENTES

En el capítulo I se definió la independencia entre eventos. Ahora se extenderá el concepto al caso de variables aleatorias.

En general se dice que dos variables aleatorias son independientes, cuando la probabilidad de que una de ellas tome cierto valor o se encuentre en un intervalo dado (en el caso contínuo) es independiente del valor que tome la otra variable aleatoria; es decir, que las funciones masa o densidad de probabilidad condicionales son iguales a las marginales, de esta manera, para las variables aleatorias discretas se tiene que:

$$p_{X/Y}(x,\ y) = p_X(x) \qquad \ldots (22)$$

$$p_{Y/X}(y,\ x) = p_Y(y) \qquad \ldots (23)$$

retomando las expresiones (18) y (19):

$$p_{XY}(x,\ y) = p_{X/Y}(x/y)\ p_Y(y) \qquad \ldots (24)$$

$$p_{XY}(x,\ y) = p_{Y/X}(y/x)\ p_X(x) \qquad \ldots (25)$$

al sustituir las expresiones (22) y (23) en (24) y (25) respectivamente, se tiene:

$$\boxed{p_{XY}(x,\ y) = p_X(x)\ p_Y(y)} \qquad \ldots (26)$$

$$\boxed{p_{XY}(x,\ y) = p_Y(y)\ p_X(x)} \qquad \ldots (27)$$

lo que significa que cuando las variables aleatorias discretas X y Y son independientes, la función masa de probabilidad conjunta es igual al producto de las funciones marginales de X y de Y.

Para el caso de variables continuas, se tiene también que cuando, x y Y son variables aleatorias independientes:

$$f_{X/Y}(x, y) = f_X(x) \ , \quad f_{Y/X}(y, x) = f_Y(y) \qquad \ldots (28)$$

$$f_{XY}(x, y) = f_X(x)\, f_Y(y) \qquad \ldots (29)$$

Ejemplo IV.8

Comprobar si las variables aleatorias de los ejemplos IV.6 y IV.7 son independientes.

Solución

Para el problema IV.6 se puede utilizar la expresión (29) en la cual se necesita primero calcular $f_X(x)$ y $f_Y(y)$.

Como:

$$f_{XY}(x, y) = \begin{cases} 0.1 & \text{para } \begin{matrix} 0 \leq x \leq 5 \\ 0 \leq y \leq 2 \end{matrix} \\ 0 & \text{para cualquier otro caso} \end{cases}$$

y de las expresiones (9) y (10):

$$f_X(x) = \int_0^2 0.1 \, dy = \left[0.1y\right]_0^2 = 0.2$$

$$f_Y(y) = \int_0^5 0.1 \, dx = \left[0.1x\right]_0^5 = 0.5$$

entonces de la expresión (29):

$$\begin{aligned} f_X(x)\ f_Y(y) &= (0.2)(0.5) \\ &= 0.1 \\ &= f_{XY}(x, y) \end{aligned}$$

por lo tanto se puede concluir que las variables sí son independientes.

Para el ejemplo IV.7 se puede observar que:

$$P_{X/Y}(y, x) \neq P_Y(y) \quad \text{y} \quad P_{X/Y}(x, y) \neq P_X(x)$$

es decir que:

$$\frac{2x + y}{8x + 6} \neq \frac{2 + y}{14} \quad \text{y} \quad \frac{2x + y}{6 + 3y} \neq \frac{8x + 6}{42}$$

por lo tanto las variables aleatorias no son independientes.

IV.4 ESPERANZA DE FUNCIONES CONJUNTAS

Frecuentemente se presentan variables aleatorias que están en función de otras dos o más variables. Por ejemplo, en el diseño de una subestación eléctrica para una planta industrial, la capacidad puede determinarse por la suma de las demandas de cada uno de sus departamentos; dichas demandas pueden considerarse como variables aleatorias y la determinación de su comportamiento es seguramente más sencilla que para la demanda total. En este caso la capacidad total C puede expresarse como la suma de las n demandas D_i, donde $i = 1, 2, \ldots, n$, es decir:

$$C = D_1 + D_2 + \ldots + D_n$$

Por ahora se tratará el caso de dos variables aleatorias y se definirá la esperanza de ciertos tipos de función que se utilizan para obtener los momentos con respecto al origen y los momentos con respecto a las medias de las distribuciones de probabilidad conjunta. También se obtendrá la media y la variancia de una combinación lineal de variables aleatorias.

La definición del operador esperanza, vista en el capítulo II, puede ampliarse para el caso de una función de dos variables aleatorias $H(X, Y)$ como:

$$\boxed{E\left[H(X, Y)\right] = \sum_{\forall x} \sum_{\forall y} H(x, y) P_{XY}(x, y) \quad \text{caso discreto}} \ldots (30)$$

$$\boxed{E\left[H(X, Y)\right] = \int_{-\infty}^{\infty}\int_{\infty}^{\infty} H(x, y)\, f_{XY}(x, y)\, dy\, dx \quad \text{caso continuo}} \ldots (31)$$

MOMENTOS DE UNA FUNCION CONJUNTA

A partir del concepto de esperanza se pueden definir los momentos con respecto al origen, al hacer $H(X, Y) = X^k\, Y^r$ donde k y r indican el grado del momento.

Para este curso tienen especial importancia los siguientes casos: cuando $(k = 1\ ,\ r = 0)$, $(k = 0\ ,\ r = 1)$, $(k = 1\ ,\ r = 1)$ y $(k = 2\ ,\ r = 2)$. El último de estos casos se usará más adelante y los primeros tres, se pueden interpretar como sigue:

$$E\left[X\, Y^0\right] = E\left[X\right] = \mu_X$$

$$E\left[X^0\, Y\right] = E\left[Y\right] = \mu_Y$$

$$E\left[X\, Y\right] = \mu_{XY}$$

este último momento se puede interpretar como el valor promedio del producto de las variables aleatorias X y Y.

Teorema IV.1

Si dos variables aleatorias X, Y son independientes, la esperanza conjunta

$$E\left[X\, Y\right] = E\left[X\right]\ E\left[Y\right] \qquad \ldots (32)$$

Demostración:

Considerando variables continuas, de la expresión (31):

$$E\left[X\ Y\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\ f_{XY}(x,\ y)\ dy\ dx$$

como X y Y son independientes:

$$f_{XY}(x,\ y) = f_X(x)\ f_Y(y)$$

entonces:

$$E\left[X\ Y\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\ f_X(x)\ f_Y(y)\ dy\ dx$$

$$= \int_{-\infty}^{\infty} x\ f_X(x)\ dx \int_{-\infty}^{\infty} y\ f_Y(y)\ dy$$

finalmente:

$$E\left[X\ Y\right] = E\left[X\right]\ E\left[Y\right]$$

De manera semejante se puede probar con variables discretas.

Ejemplo IV.9

Una empresa está elaborando el presupuesto que destinará al departamento de mantenimiento para el próximo año. En los archivos de la empresa se encontraron registros de gastos mensuales, con el número de reparaciones que se hicieron en cada ocasión. Del análisis de los mismos se obtuvo que es posible modelar el comportamiento de las variables con la siguiente F.D.P. conjunta.

Y \ X	5	6	7	8	9
1	0.08	0.04	0	0	0
1.5	0.04	0.12	0.08	0.04	0
2	0.04	0.08	0.12	0.12	0
2.5	0	0.04	0.08	0.08	0.04

donde:

X = Número de reparaciones por mes.

Y = Costo unitario de reparación en decenas de miles de pesos, (considerado en forma discreta).

Cuál es:

a) El número esperado de reparaciones por mes.

b) El costo esperado por una reparación.

c) El costo mensual esperado por concepto de reparaciones.

Solución

a) Para obtener la esperanza $E[X]$ se puede determinar primero la F.M.P. marginal de X y posteriormente calcular la esperanza tal como se vio en el capítulo II, o bien utilizar la expresión (30). Siguiendo este último procedimiento:

$$E[X] = \sum_{x=5}^{9} \sum_{y=1}^{2.5} x\, P_{XY}(x, y) = \sum_{x=5}^{9} x \sum_{y=1}^{2.5} P_{XY}(x, y)$$

$$= 5(0.08 + 0.04 + 0.04) + 6(0.04 + 0.12 + 0.08 + 0.04)$$

$$+ 7(0.08 + 0.12 + 0.08) + 8[0.04 + 0.12 + 0.08] + 9(0.04)$$

$$= 5(0.16) + 6(0.28) + 7(0.28) + 8(0.24) + 9(0.04) = 6.72 \simeq 7 \text{ reparaciones}$$

b) De manera semejante:

$$E[Y] = \sum_{x=5}^{9} \sum_{y=1}^{2.5} y\, P_{XY}(x, y) = \sum_{y=1}^{2.5} y \sum_{x=5}^{9} P_{XY}(x, y)$$

$$= 1(0.08 + 0.04) + 1.5(0.04 + 0.12 + 0.08 + 0.04)$$

$$+ 2(0.04 + 0.08 + 0.12 + 0.12) + 2.5(0.04 + 0.08 + 0.08 + 0.04)$$

$$= 1(0.12) + 1.5(0.28) + 2(0.36) + 2.5(0.24) = 1.86$$

como son decenas de miles de pesos:

$$E[Y] = \$\ 18{,}600.00$$

c) Considerando que el costo mensual es igual al producto XY:

$$E[XY] = \sum_{x=5}^{9} \sum_{y=1}^{2.5} xy\, P_{XY}(x, y) = \sum_{x=5}^{9} x \sum_{y=1}^{2.5} y\, P_{XY}(x, y)$$

$$= 5[1(0.08) + 1.5(0.04) + 2(0.04)] +$$

$$+ 6\,[1(0.04) + 1.5(0.12) + 2(0.08) + 2.5(0.04)] +$$

$$+ 7[1.5(0.08) + 2(0.12) + 2.5(0.08)] +$$

$$+ 8[1.5(0.04) + 2(0.12) + 2.5(0.08)] + 9[2.5(0.04)]$$

$$= 5(0.22) + 6 + 7(0.56) + 8(0.5) + 9(.1) = 12.8$$

$$E[XY] = \$\ 128{,}000.00$$

Se puede observar que:

$E\boxed{X}\ E\boxed{Y} = 6.72(1.86) = 12.499 \neq E\boxed{XY}$

lo cual implica que las variables aleatorias X y Y no son independientes.

Existen también momentos centrales (con respecto a las medias) que se obtienen al hacer $H(X, Y) = (X - \mu_X)^k (Y - \mu_Y)^r$, donde k y r representan nuevamente el grado que tiene el momento. En especial los casos que más se utilizan son cuando:

$(k = 2,\ r = 0),\ \ (k = 0,\ r = 2)$ y $(k = 1,\ r = 1)$

Con los dos primeros, se obtiene la variancia de cada una de las variables; es decir:

$$E[(X - \mu_X)^2 (X - \mu_X)^0] = \sigma_X^2 \quad y \quad E[(X - \mu_X)^0 (X - \mu_Y)^2] = \sigma_Y^2$$

al tercer caso se le llama *covariancia de* XY y se representa por σ_{XY} ó cov $\boxed{XY}$, de esta manera:

$$\boxed{\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]} \qquad \ldots (33)$$

Otra forma de calcular la covariancia se obtiene al desarrollar la expresión anterior en la forma:

$$E\left[(X - \mu_X)(Y - \mu_Y)\right] = E\left[XY - Y_{\mu_X} - X\,\mu_Y + \mu_X\,\mu_Y\right]$$

que por las propiedades del operador esperanza:

$$E\,[(X - \mu_X)(Y - \mu_Y)] = E\,\boxed{XY} - \mu_X\,E\,\boxed{Y} - E\,\boxed{X}\,\mu_Y + \mu_X\,\mu_Y$$

además como:

$$E\,\boxed{X} = \mu_X \quad y \quad E\,\boxed{Y} = \mu_Y$$

$$E\left[(X - \mu_X)(Y - \mu_Y)\right] = E\left[XY\right] - 2\,\mu_X\,\mu_Y + \mu_X\,\mu_Y$$

finalmente:

$$E\,[(X - \mu_X)(Y - \mu_Y)] = E\,\boxed{XY} - \mu_X\,\mu_Y \qquad \ldots (34)$$

La covariancia tiene la característica de tomar un valor nulo cuando las variables aleatorias son independientes. Esto se comprueba fácilmente al observar la expresión (34), en donde $E\,\boxed{XY} = E\,[X]\,E\,[Y] = \mu_X\,\mu_Y$ cuando X y Y son independientes, lo cual implica que la covariancia es cero.

También tiene la propiedad de tomar valores negativos cuando la relación entre las variables es inversamente proporcional y positivos cuando es directamente proporcional.

Las características anteriores hacen de la covariancia un indicador importante de la relación lineal que existe entre las variables, aunque tiene la desventaja de que no se pueden comparar covariancias de diferentes experimentos aleatorios, puesto que su valor depende de las diferentes unidades de cada experimento.

Existe otro indicador que es el *coeficiente de correlación*, el cual es una normalización de la covariancia y se define como el cociente de la covariancia entre el producto de las desviaciones estándar de las dos variables en estudio, esto es:

$$\boxed{\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \, \sigma_Y}} \qquad \ldots (35)$$

lo cual lo convierte en un indicador adimensional.

Teorema IV.2

Para cualquier par de variables aleatorias X, Y

$$-1 \leq \rho_{XY} \leq 1$$

Demostración:

Sea a un número real cualquiera y la función:

$$H(X, Y) = [(X - \mu_X) - a(Y - \mu_Y)]^2$$

donde:

$$\mu_X = E[X] \quad y \quad \mu_Y = E[Y]$$

obteniendo la esperanza de $H(X, Y)$:

$$E[H(X, Y)] = E[(X - \mu_X)^2 - 2a\,(X - \mu_X)(Y - \mu_Y) + a^2(Y - \mu_Y)^2]$$

por las propiedades del operador esperanza:

$$E[H(X, Y)] = E[(X - \mu_X)^2] - 2a\,E[(X - \mu_X)(Y - \mu_Y)] + a^2\,E[(Y - \mu_Y)^2]$$

$$= \sigma_X^2 - 2a\,\sigma_{XY} + a^2\,\sigma_Y^2 \quad \text{para toda } a$$

haciendo $a = \dfrac{\sigma_{XY}}{\sigma_Y^2}$

$$E[H(X, Y)] = \sigma_X^2 - 2\,\frac{\sigma_{XY}^2}{\sigma_Y^2} + \frac{\sigma_{XY}^2}{\sigma_Y^4}\,\sigma_Y^2$$

$$= \sigma_X^2 - \frac{\sigma_{XY}^2}{\sigma_Y^2}$$

como H(x, y) no puede tomar valores negativos, entonces:

$$\sigma_X^2 - \frac{\sigma_{XY}^2}{\sigma_Y^2} \geq 0$$

por lo que:

$$\frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2} \leq 1 \qquad \text{y} \qquad \left|\frac{\sigma_{XY}}{\sigma_X \sigma_Y}\right| \leq 1$$

como $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$ implica que:

$$-1 \leq \rho_{XY} \leq 1$$

Cuando $|\rho| = 1$ se dice que existe una correlación perfecta entre X y Y; es decir, que se puede expresar una de ellas como una función lineal de la otra.

Cuando $\rho = 0$, se puede afirmar que las variables son linealmente independientes; sin embargo, en la mayoría de los casos tomará un valor intermedio que solamente nos puede servir de indicador, puesto que un coeficiente de correlación de 0.8 no implica que la correlación es dos veces mayor que otra de 0.4. Tampoco se puede asegurar que dos variables que tengan un coeficiente de correlación muy pequeño, son independientes ya que pueden tener otro tipo de relación que no sea lineal, como:

$$Y = ax^b \quad \text{ó} \quad Y = ae^{bx}$$

Ejemplo IV.10

En una cuenca donde se piensa construir una presa, se unen dos ríos cuyo caudal de agua se puede modelar con la siguiente F.D.P. conjunta:

$$f_{XY}(x, y) = \begin{cases} \frac{x + y}{24} & \text{para } 0 \leq x \leq 2 \text{ [metros cúbicos/segundo]} \\ & \phantom{\text{para }} 0 \leq y \leq 4 \\ 0 & \text{para cualquier otro caso} \end{cases}$$

Calcular:

a) El valor esperado de cada uno de los flujos.

b) La covariancia de la distribución.

c) El coeficiente de correlación.

Solución

a) De la expresión (31) para H(x, y) = x

$$E\,[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f_{XY}(x, y)\, dy\, dx = \int_{0}^{2}\int_{0}^{4} \frac{x(x+y)}{24}\, dy\, dx$$

$$= \int_{0}^{2}\int_{0}^{4} \frac{x^2+xy}{24}\, dy\, dx \quad = \frac{1}{24}\int_{0}^{2}\left[yx^2 + \frac{xy^2}{2}\right]_0^4 dx$$

$$= \frac{1}{24}\int_{0}^{2}(4x^2 + 8x)\, dx \quad = \frac{1}{24}\left[\frac{4x^3}{3} + 4x^2\right]_0^2 = 1.111$$

Para Y, H(x, y) = y, de manera que:

$$E\,[Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f_{XY}(x, y)\, dy\, dx = \frac{1}{24}\int_{0}^{2}\int_{0}^{4} \frac{y(x+y)}{24}\, dy\, dx$$

$$= \int_{0}^{2}\int_{0}^{4} \frac{xy+y^2}{24}\, dy\, dx \quad = \frac{1}{24}\int_{0}^{2}\left[\frac{y^2x}{2} + \frac{y^3}{3}\right]_0^4 dx$$

$$= \frac{1}{24}\int_{0}^{2}\left(8x + \frac{64}{3}\right) dx \quad = \frac{1}{24}\left[\frac{8x^2}{2} + \frac{64x}{3}\right]_0^2 = 2.444$$

b) Obteniendo primero $E\,[XY]$ para utilizar la expresión (34):

$$E\,[XY] = \frac{1}{24}\int_{0}^{2}\int_{0}^{4} xy(x+y)\, dy\, dx = \frac{1}{24}\int_{0}^{2}\left[\frac{x^2y^2}{2} + \frac{xy^3}{3}\right]_0^4 dx$$

$$= \frac{1}{24}\int_{0}^{2}\left(8x^2 + \frac{64x}{3}\right) dx = \frac{1}{24}\left[\frac{8x^3}{3} + \frac{64x^2}{6}\right]_0^2 = 2.666$$

sustituyendo en la expresión (34):

$\mathrm{cov}\,[X, Y] = 2.666 - 1.111(2.444) = -\,0.0494$

c) Para calcular el coeficiente de correlación es necesario obtener primero la variancia y posteriormente la desviación estándar de cada variable, para lo cual:

$$\sigma_X^2 = E\,[X^2] - \mu_X^2$$

donde:

$$E\left[X^2\right] = \frac{1}{24}\int_0^2\int_0^4 x^2(x+y)\,dy\,dx = \frac{1}{24}\int_0^2\left[x^3\,y + \frac{y^2\,x^2}{2}\right]_0^4 dx$$

$$= \frac{1}{24}\int_0^2 (4x^3+8x^2)\,dx = \frac{1}{24}\left[\frac{4x^4}{4} + \frac{8x^3}{3}\right]_0^2 = 1.556$$

con lo cual:

$$\sigma_X^2 = 1.556 - (1.111)^2 = 0.322 \quad y \quad \sigma_X = \sqrt{0.322} = 0.567$$

para Y:

$$E\left[Y^2\right] = \frac{1}{24}\int_0^2\int_0^4 y^2(x+y)\,dy\,dx = \frac{1}{24}\int_0^2\left[\frac{x\,y^3}{3} + \frac{y^4}{4}\right]_0^4 dx$$

$$= \frac{1}{24}\int_0^2\left[\frac{64}{3}\,x + 64\right] dx = \frac{64}{24}\left[\frac{x^2}{6} + x\right]_0^2 = 7.111$$

con lo cual:

$$\sigma_Y^2 = 7.111 - (2.444)^2 = 1.138 \quad y \quad \sigma_Y = \sqrt{1.138} = 1.067$$

finalmente de (35):

$$\rho = \frac{-\ 0.0494}{(0.567)(1.067)} = -\ 0.081$$

ESPERANZA Y VARIANCIA DE FUNCIONES LINEALES DE DOS VARIABLES ALEATORIAS

Cuando una variable aleatoria Z está formada por una combinación lineal de otras dos variables aleatorias X , Y en la forma $Z = aX + bY$, donde a y b son constantes; se puede conocer la media y la variancia de Z, a partir de la función masa o densidad de probabilidad conjunta de X , Y. De esta manera, para variables discretas, la media está dada por:

$$E\left[Z\right] = E\left[aX + bY\right] = \sum_{\forall x}\sum_{\forall y}\left[aX + bY\right] P_{XY}(x,\ y)$$

por las propiedades de la sumatoria:

$$E\left[aX + bY\right] = a\sum_{\forall x}\sum_{\forall y} x\,P_{XY}(x,\ y) + b\sum_{\forall x}\sum_{\forall y} y\,P_{XY}(x,\ y)$$

como:

$$E[X] = \sum_{\forall x} \sum_{\forall y} x\, P_{XY}(x, y) \quad y \quad E[Y] = \sum_{\forall x} \sum_{\forall y} y\, P_{XY}(x, y)$$

entonces:

$$\boxed{E[aX + bY] = a\,E[X] + b\,E[Y]} \qquad \ldots (36)$$

de manera semejante se puede demostrar para el caso continuo.

La variancia, por su misma definición:

$$\begin{aligned}\sigma_Z^2 &= E\,[(aX + bY) - E(aX + bY)]^2 \\ &= E\,[aX - a\mu_X + bY - b\mu_Y]^2 \\ &= E\,[a^2(X - \mu_X)^2 + 2ab(X - \mu_X)(Y - \mu_Y) + b^2(Y - \mu_Y)^2] \\ &= a^2\,E\,[(X-\mu_X)^2] + 2\,ab\,E\,[(X - \mu_X)(Y - \mu_Y)] + b^2\,E\,[(Y - \mu_Y)^2]\end{aligned}$$

como:

$$E\,[(X - \mu_X)^2] = \sigma_X^2 \;; \qquad E\,[(Y - \mu_Y)^2] = \sigma_Y^2 \quad y$$

$$E\,[(X - \mu_X)(Y - \mu_Y)] = \sigma_{XY}$$

$$\boxed{\sigma_Z^2 = a^2\,\sigma_X^2 + b^2\,\sigma_Y^2 + 2ab\,\sigma_{XY} \quad \text{donde} \quad Z = aX + bY} \qquad \ldots (37)$$

se debe observar que si X y Y son independientes $\sigma_{XY} = 0$

Ejemplo IV.11

En una empresa maquiladora se está diseñando una nueva planta industrial. Actualmente, la planta en operación tiene dos departamentos A y B, de los cuales se conoce la F.D.P. del consumo de energía eléctrica (en KVA):

para el departamento A:

$$f_X(x) = \begin{cases} \frac{3}{16}\, x^{1/2} & \text{para} \quad 0 \leq x \leq 4 \quad [KVA] \\ 0 & \text{para cualquier otro caso} \end{cases}$$

para el departamento B:

$$f_Y(y) = \begin{cases} 0.1 + 0.04y & \text{para} \quad 0 \leq y \leq 5 \quad [KVA] \\ 0 & \text{para cualquier otro caso} \end{cases}$$

En la nueva planta se tendrán dos departamentos semejantes (A' y B'), pero se calcula que A' será 2.5 veces más grande que A y B' será el triple de B. Suponiendo que el consumo de energía eléctrica entre los departamentos es independiente, cuál es:

a) El valor esperado de la demanda de energía en la nueva planta.

b) La desviación estándar de la misma.

Solución

La demanda total de energía (D) se puede representar como:

$$D = 2.5X + 3Y$$

a) De la expresión (36):

$$E[D] = 2.5\,E[X] + 3\,E[Y] \qquad \ldots (38)$$

calculando primero $E[X]$ y $E[Y]$:

$$E[X] = \int_0^4 x\left[\frac{3x^{1/2}}{16}\right] dx = \frac{3}{16}\left[\frac{x^{5/2}}{5/2}\right]_0^4 = 2.4$$

$$E[Y] = \int_0^5 y(0.1 + 0.04y)\,dy = (0.1)\left[\frac{y^2}{2} + \frac{0.4y^3}{3}\right]_0^5 = 2.917$$

sustituyendo en la expresión (38):

$$E[D] = 2.5(2.4) + 3(2.917) = 14.751\ [KVA]$$

b) Tomando en cuenta que las variables aleatorias son independientes $\sigma_{XY} = 0$, de manera que de la expresión (37):

$$\sigma_D^2 = (2.5)^2\ \sigma_X^2 + (3)^2\ \sigma_Y^2 \qquad \ldots (39)$$

calculando primero σ_X^2 y σ_Y^2:

$$\sigma_X^2 = E[X^2] - \mu_X^2 = \int_0^4 x^2\left[\frac{3x^{1/2}}{16}\right] dx - (2.4)^2$$

$$= \frac{3}{16}\left[\frac{x^{7/2}}{7/2}\right]_0^4 - (2.4)^2 = 6.875 - 5.76 = 1.097$$

$$\sigma_Y^2 = E\left[Y^2\right] - \mu_Y^2 = \int_0^5 y^2\ (0.1 + 0.04y)\ dy - (2.917)^2$$

$$= 0.1\left[\frac{y^3}{3} + \frac{0.4y^4}{4}\right]_0^5 - (2.917)^2 = 10.417 - 8.509 = 1.908$$

sustituyendo en (39):

$$\sigma_D^2 = (2.5)^2(1.097) + (3)^2(1.908) = 24.028$$

con lo cual:

$$\sigma_D = \sqrt{24.028} = 4.902 \quad [\text{KVA}]$$

IV.5 VECTORES ALEATORIOS

Todos los conceptos anteriores que se refieren a dos variables aleatorias, se pueden extender al caso general de n variables aleatorias. En ingeniería existen numerosos problemas en los que es necesario manejar más de dos variables aleatorias en forma conjunta, de manera que a cada posible resultado de un experimento se le asocian una serie de valores como temperatura, presión, intensidad de corriente, etc., que pueden representarse mediante un vector $(X_1, X_2, \ldots, X_n)$ y como cada una de las $X_i (i = 1, 2, \ldots, n)$ es una variable aleatoria, entonces a $(X_1, X_2, \ldots, X_n)$ se le denomina *vector aleatorio*. También se dice que el resultado del experimento es una variable aleatoria de dimensión n.

FUNCIONES CONJUNTAS DE n VARIABLES ALEATORIAS

Con el objeto de simplificar la representación de un vector y manteniendo la misma notación de las funciones conjuntas, se utilizará $\overline{X}$ en lugar de $(X_1, X_2, \ldots, X_n)$. De esta manera la F.M.P. conjunta será:

$$p_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = p_{\overline{X}}(\overline{x})$$

que como en el caso de dos variables:

$$p_{\overline{X}}(\overline{x}) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) \qquad \ldots (40)$$

Esta función cumple también con las propiedades (2) y (3). La F.M.P. marginal de una variable X_K o de un grupo de m variables (donde $m < n$), se puede obtener sumando la F.M.P. sobre todo el espectro de las n-1 ó n-m variables restantes, por ejemplo:

$$p_{X_1, X_3}(x_1, x_3) = \sum_{\forall x_2} \sum_{\forall x_4} \ldots \sum_{\forall x_n} p_{\overline{X}}(\overline{x})$$

Para las variables continuas se puede representar la función densidad de probabilidad como:

$$f_{\overline{X}}(\overline{x}) = f_{X_1,X_2,\ldots,\ X_n}\ (x_1,\ x_2,\ \ldots,\ x_n)$$

y la probabilidad de que el vector aleatorio $\overline{X}$ de dimensión n se encuentre en cierta región A del espacio muestral, se puede obtener en la siguiente forma:

$$P(\overline{X} \in A) = \int_A \int \cdots \int f_{\overline{X}}(\overline{x})\ dx_1,\ dx_2,\ \ldots,\ dx_n \qquad \ldots\ (41)$$

De manera análoga al caso discreto, las F.D.P. marginales de una variable o de un grupo de m variables continuas, se obtienen integrando la función $F_{\overline{X}}(x)$ en todo el espectro de las $n-1$ ó $n-m$ variables restantes, por ejemplo:

$$f_{X_2,X_3}\ (x_2,\ x_3) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\overline{X}}(x)\ dx_1,\ dx_4,\ \ldots,\ dx_n \qquad \ldots\ (42)$$

La función de distribución acumulada conjunta, en el caso de vectores aleatorios, se define de la siguiente manera:

$$F_{\overline{X}}(\overline{x}^c) = P\ [(X_1 \leq x_{1,0})\quad (X_2 \leq x_{2,0})\quad \ldots\quad (X_n \leq x_{n,0})]$$

la cual se relaciona con la función masa o densidad de probabilidad conjunta de la siguiente forma:

$$F_{\overline{X}}(\overline{x}_0) = \int_{-\infty}^{x_{1,0}} \int_{-\infty}^{x_{2,0}} \cdots \int_{-\infty}^{x_{n,0}} f_{\overline{X}}(\overline{x})\ dx_1\, dx_2\ \ldots\ dx_n \ \text{(caso continuo)} \qquad \ldots\ (43)$$

$$F_{\overline{X}}(\overline{x}_0) = \sum_{x_1=0}^{x_{1,0}} \sum_{x_2=0}^{x_{2,0}} \cdots \sum_{x_n=0}^{x_{n,0}} P_{\overline{X}}(\overline{x}) \quad \text{(caso discreto)} \qquad \ldots\ (44)$$

Existen también funciones condicionales de $x_1,\ x_2,\ \ldots,\ x_k$ dado $x_{k+1},\ \ldots,\ x_n$, que en el caso continuo, por ejemplo, se define como:

$$f_{X_1,\ldots,\ X_k/X_{k+1},\ldots,\ X_n}\ (x_1,\ x_2,\ldots,\ x_n) = \frac{f_{X_1,\ldots,\ X_n}(x_1,\ldots,\ x_n)}{f_{X_{k+1},\ldots,\ X_n}(x_{k+1},\ldots,\ x_n)} \qquad \ldots\ (45)$$

Se puede demostrar fácilmente que si las variables $X_1, X_2, \ldots X_n$ son mutuamente independientes:

$$p_{\overline{X}}(\overline{x}) = \prod_{i=1}^{n} p_{X_i}(x_i) \quad \forall\ X_i \quad \text{(caso discreto)} \qquad \ldots (46)$$

$$f_{\overline{X}}(\overline{x}) = \prod_{i=1}^{n} f_{X_i}(x_i) \quad \forall\ X_i \quad \text{(caso continuo)} \qquad \ldots (47)$$

lo cual indica que si las variables aleatorias son independientes, se puede conocer el comportamiento del vector aleatorio a partir del producto de las funciones marginales de las variables que lo componen.

Ejemplo IV.12

En una gasolinera se ha observado que el número de vehículos que llegan a cada una de las tres estaciones de servicio (X_1, X_2, X_3) en un minuto, tiene distribución Poisson con parámetro $\lambda = 2$ vehículos/minuto. Si estas variables se consideran independientes. Cuál es:

a) La F.M.P. conjunta del número de vehículos que llega a cada estación en un minuto.

b) La probabilidad de que lleguen tres vehículos a la primera estación, dos a la segunda y uno a la tercera durante un mismo minuto.

c) La probabilidad de que lleguen dos o más vehículos a cada estación en el mismo minuto.

Solución

a) Como las variables son independientes, de la expresión (46):

$$p_{X_1X_2X_3}(x_1, x_2, x_3) = p_{X_1}(x_1)\, p_{X_2}(x_2)\, p_{X_3}(x_3)$$

sustituyendo la F.M.P. de la distribución Poisson:

$$p_{X_1X_2X_3}(x_1, x_2, x_3) = \left[\frac{2^{x_1} e^{-2}}{x_1!}\right] \left[\frac{2^{x_2} e^{-2}}{x_2!}\right] \left[\frac{2^{x_3} e^{-2}}{x_3!}\right]$$

$$= \frac{2^{x_1+x_2+x_3}}{x_1!\,x_2!\,x_3!}\, e^{-6}$$

b) Haciendo $x_1 = 3,\ x_2 = 2,\ x_3 = 1$

$$p_{X_1X_2X_3}(3, 2, 1) = \frac{2^{3+2+1}}{3!\,2!\,1!}\, e^{-6} = 0.0132$$

c) Considerando que $P(X_1 \geq 2, X_2 \geq 2, X_3 \geq 2) = 1 - P(X_1 < 2, X_2 < 2, X_3 < 2)$

$$P(X_1 < 2, X_2 < 2, X_3 < 2) = P_{X_1X_2X_3}(0,0,0) + P_{X_1X_2X_3}(0,0,1) + P_{X_1X_2X_3}(0,1,0)$$

$$+ P_{X_1X_2X_3}(1,0,0) + P_{X_1X_2X_3}(0,1,1) + P_{X_1X_2X_3}(1,0,1)$$

$$+ P_{X_1X_2X_3}(1,1,0) + P_{X_1X_2X_3}(1,1,1)$$

$$= e^{-6}\left[\frac{2^0}{0!0!0!} + 3\ \frac{2^1}{0!0!1!} + 3\ \frac{2^2}{0!1!1!} + \frac{2^3}{1!1!1!}\right]$$

$$P(X_1 < 2, X_2 < 2, X_3 < 2) = e^{-6}\ [1 + 3(2) + 3(4) + 8] = 0.3346$$

entonces:

$$P(X_1 \geq 2;\ X_2 \geq 2,\ X_3 \geq 2) = 1 - 0.3346 = 0.6654$$

ESPERANZA DE FUNCIONES CONJUNTAS DE n VARIABLES ALEATORIAS

Para una función de n variables aleatorias $H(X_1, X_2, \ldots, X_n)$ se define su esperanza (en el caso continuo) como:

$$E\ [H(X_1, X_2, \ldots, X_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} H(x_1, x_2, \ldots, x_n)\, f_{\bar{X}}(\bar{x})\, dx_1 dx_2 \ldots dx_n \qquad \ldots (48)$$

o bien en el caso discreto como:

$$E\ [H(X_1, X_2, \ldots, X_n)] = \sum_{\forall x_1} \cdots \sum_{\forall x_n} H(X_1, X_2, \ldots, X_n)\ P_{\bar{X}}(\bar{x}) \qquad \ldots (49)$$

de esta manera es posible encontrar la media y variancia de cada una de las variables aleatorias, al hacer:

$$H(X_1, X_2, \ldots, X_n) = X_k \quad ó$$

$$H(X_1, X_2, \ldots, X_k) = \left(X_k - \mu_{X_k}\right)^2$$

para $k = 1, 2, \ldots, n$

Otra aplicación importante de las expresiones (48) y (49) es la matriz de covariancia o variancia del vector aleatorio.

Esta matriz se obtiene como:

$$E\ \{[\bar{X} - E(\bar{X})]^t [\bar{X} - E(\bar{X})]\}$$

donde:

$$E(\bar{X}) = [E(X_1), E(X_2), \ldots, E(X_n)]$$

y el superíndice t significa que es la transpuesta de la matriz $[\bar{X} - E(\bar{X})]$, esta matriz es simétrica y tendrá la forma:

$$\begin{bmatrix} \sigma_{X_1X_1} & \sigma_{X_1X_2} & \cdots & \sigma_{X_1X_n} \\ \sigma_{X_2X_1} & \sigma_{X_2X_2} & \cdots & \sigma_{X_2X_n} \\ \vdots & \vdots & & \\ \sigma_{X_nX_1} & \sigma_{X_nX_2} & \cdots & \sigma_{X_nX_n} \end{bmatrix}$$

donde $\sigma_{X_iX_j}$ es la variancia de X_i cuando $i = j$ y es la covariancia entre las dos variables si $i \neq j$.

El uso de vectores aleatorios de dimensión mayor que dos tiene sus principales aplicaciones en la estadística y como se verá en los siguientes capítulos muchas variables aleatorias están formadas por una función de otras n variables aleatorias.

CAPITULO V TEOREMAS SOBRE CASOS LIMITE

INTRODUCCION

En este capítulo se presentan algunos teoremas basados en los conceptos anteriores que resaltan la importancia que tienen las distribuciones normal y normal estándar.

Estos teoremas reafirman la validez de la interpretación frecuentista de la probabilidad, además vinculan los conceptos de probabilidad con la inferencia estadística.

V.1 TEOREMA O DESIGUALDAD DE CHEBYSHEV

Se ha visto que los parámetros descriptivos, tales como la media y la variancia, pueden proporcionar información valiosa para determinar el comportamiento de una variable aleatoria.

En algunos casos es necesario estimar la probabilidad de que una variable aleatoria X, cuya distribución de probabilidad se desconoce, esté comprendida dentro o fuera de cierto intervalo cuyo centro es μ_X y cuyos extremos estén separados k desviaciones estándar del centro (donde k es una constante positiva), como se muestra en la figura V.1. En ella se observa que si $|X - \mu_X| \leq k\,\sigma_X$, la variable X está comprendida en el intervalo central $[\mu_X - k\,\sigma_X,\ \mu_X + k\,\sigma_X]$.

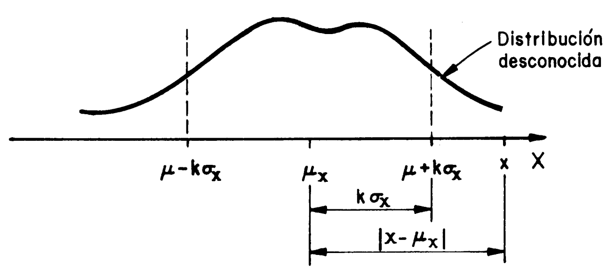


Figura V.1

Obtener esta probabilidad resulta sencillo cuando se conoce la distribución de la variable aleatoria; sin embargo, cuando únicamente se conoce la media y la variancia, se puede estimar mediante el teorema denominado de Chebyshev, el cual se presenta a continuación.

Teorema V.1

Sea X una variable aleatoria, con distribución de probabilidad desconocida, con media μ_X y variancia σ_X^2 ; entonces la probabilidad de que la diferencia, en valor absoluto, de los posibles valores de la variable y su media sea cuando mucho $k\ \sigma_X$, donde k es una constante positiva, es cuando menos

$$1 - \frac{1}{k^2},$$

esto es:

$$\boxed{P(|X - \mu_X| < k\ \sigma_X) \geq 1 - \frac{1}{k^2}} \qquad \ldots (1)$$

o bien, mediante su complemento:

$$\boxed{P(|X - \mu_X| \geq k\ \sigma_X) \leq \frac{1}{k^2}} \qquad \ldots (2)$$

Demostración:

A partir de la definición de variancia:

$$\sigma_X^2 = E\left[(X - \mu_X)^2\right]$$

que en el caso continuo:

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2\ f_X(x)\ dx$$

separando el intervalo de integración en tres subintervalos, como los mostrados en la figura V.1:

$$\sigma_X^2 = \int_{-\infty}^{\mu_X - k\ \sigma_X} (x - \mu_X)^2\ f_X(x)\ dx + \int_{\mu_X - k\ \sigma_X}^{\mu_X + k\ \sigma_X} (x - \mu_X)^2\ f_X(x)\ dx +$$

$$\int_{\mu_X + k\ \sigma_X}^{\infty} (x - \mu_X)^2\ f_X(x)\ dx$$

Como en el segundo término $(x - \mu_X)^2 \geq 0$ y $f_X(x) \geq 0$ para toda x, la integral es no negativa, por lo cual:

$$\sigma_X^2 \geq \int_{-\infty}^{\mu_X - k\,\sigma_X} (x-\mu_X)^2\, f_X(x)\, dx + \int_{\mu_X + k\,\sigma_X}^{\infty} (x-\mu_X)^2\, f_X(x)\, dx \quad \ldots \quad (a)$$

Observando la figura V.1, se puede notar que para cualquier valor x de los dos intervalos de integración:

$$|x - \mu_X| \geq k\,\sigma_X$$

de donde:

$$(x - \mu_X)^2 \geq k^2\,\sigma_X^2$$

sustituyendo en la expresión (a):

$$\sigma_X^2 \geq \int_{-\infty}^{\mu_X - k\,\sigma_X} k^2\,\sigma_X^2\, f_X(x)\, dx + \int_{\mu_X + k\,\sigma_X}^{\infty} k^2\,\sigma_X^2\, f_X(x)\, dx$$

como $k^2\,\sigma_X^2$ es constante, se sigue que:

$$\sigma_X^2 \geq k^2\,\sigma_X^2 \left[\int_{-\infty}^{\mu_X - k\,\sigma_X} f_X(x)\, dx + \int_{\mu_X + k\,\sigma}^{\infty} f_X(x)\, dx \right]$$

donde las dos integrales representan la probabilidad del evento:

$$\left[(-\infty < X \leq \mu_X - k\,\sigma_X) \cup (\mu_X + k\,\sigma_X \leq X < \infty) \right]$$

y de acuerdo con la figura V.1 equivale a:

$$P\left[|x - \mu_X| \geq k\,\sigma_X \right]$$

con lo cual:

$$\sigma_X^2 \geq k^2\,\sigma_X^2\, P\left[|x - \mu_X| \geq k\,\sigma_X \right]$$

dividiendo entre $k^2\,\sigma_X^2$:

$$P\left[|x - \mu_X| \geq k\,\sigma_X \right] \leq \frac{1}{k^2}$$

cuyo complemento es:

$$P(|x - \mu_X| < k\,\sigma_X) > 1 - \frac{1}{k^2}$$

Para el caso discreto, la demostración es muy similar.

Ejemplo V.1

Para atender adecuadamente a las personas que deben presentar su declaración anual de impuestos, el jefe de una de las oficinas receptoras ha diseñado un sistema, de tal manera que se pueda dar servicio al número promedio de personas que se presentan en un día, más tres desviaciones estándar.

Si no se conoce la distribución de probabilidad de la variable aleatoria X que representa al número de contribuyentes que asisten a presentar su declaración, es posible usar el teorema de Chebyshev para calcular la máxima probabilidad de que en un día cualquiera se presenten más de $\mu_X + 3\ \sigma_X$ personas o menos de $\mu_X - 3\ \sigma_X$.

Solución

En este caso k = 3 y de la expresión (2):

$$P(|X - \mu_X| \geq 3\ \sigma_X) \leq \frac{1}{(3)^2} = 0.1111$$

Ejemplo V.2

En una de las oficinas de la Procuraduría Federal del Consumidor se reciben en promedio 150 quejas con una desviación estándar de 25.

a) ¿Cuál es la mínima probabilidad de recibir entre 100 y 200 quejas?

b) ¿Cuál es el intervalo más corto alrededor de la media, que contiene cuando menos al 90% del número de quejas que se reciben en un día?

Solución

a) Los límites del intervalo se encuentran a 2 desviaciones estándar de la media, por lo cual k = 2 y de la expresión (1):

$$P(|X - \mu_X| \leq 50) \geq 1 - \frac{1}{(2)^2}$$

$$= 0.75$$

lo cual significa que cuando menos en el 75% de los casos, el número de quejas en un día no difiere de su media en más de 50.

b) Para que contenga cuando menos el 90% de las quejas que se reciben en un día:

$$1 - \frac{1}{k^2} = 0.90$$

despejando k:

$$k = \sqrt{\frac{1}{0.1}} = 3.16$$

con lo cual el intervalo buscado es:

$$\left[150 - 3.16(25) \leq X \leq 150 + 3.16(25)\right] = \left[71 \leq X \leq 229\right]$$

V.2 CONVERGENCIA DE VARIABLES ALEATORIAS

Para poder estudiar con propiedad la ley de los grandes números y el teorema del límite central, es necesario definir el concepto de *convergencia* de una sucesión de variables aleatorias.

Cuando se trata de una sucesión de números $a_1, a_2, \ldots, a_n, \ldots$, se dice que tal sucesión converge hacia a (o tiende al número a) cuando dado un número $\varepsilon > 0$, se puede encontrar un entero k, tal que $|a_n - a| < \varepsilon$ para todo $n > k$.

También se puede expresar la convergencia como:

$$\lim_{n \to \infty} a_n = a \qquad \text{ó} \qquad a_n \to a$$

ahora bien, cuando se trata de una sucesión de variables aleatorias, $X_1, X_2, \ldots, X_n, \ldots$, que converge a X, no es posible asegurar que la diferencia absoluta $|X_n - X|$ sea menor que un valor ε, puesto que aun para valores muy grandes de n, X y X_n son variables aleatorias y su diferencia es otra variable aleatoria.

Por lo anterior, existen otras formas de definir la convergencia de sucesiones de variables aleatorias, entre las cuales están las siguientes:

CONVERGENCIA EN PROBABILIDAD O ESTOCASTICA

Se dice que una sucesión X_n (para $n = 1, 2, \ldots$) converge en probabilidad a X, si dado un valor arbitrario $\varepsilon > 0$, la probabilidad de que $|X_n - X| > \varepsilon$ tiende a cero cuando n tiende a infinito, es decir:

$$P\left[|X_n - X| > \varepsilon\right] \to 0 \text{ cuando } n \to \infty \qquad \ldots (3)$$

CONVERGENCIA CON PROBABILIDAD UNO O CON CASI CERTEZA

Una sucesión de variables aleatorias X_n(para $n = 1, 2, \ldots$) converge a X con probabilidad uno si la probabilidad de que $X_n \to X$ es igual a uno, cuando n tiende a infinito, esto es:

$$P\left[X_n \to X\right] = 1 \qquad \text{cuando } n \to \infty \qquad \ldots (4)$$

A partir de las definiciones anteriores se puede demostrar que si la sucesión X_n converge en probabilidad a X, entonces las funciones de distribución acumulada de las variables X_n convergen punto por punto a la función de distribución acumulada de X; además las funciones generatrices de momentos de las variables X_n tienden a la función generatriz de momentos de X.

V.3 LEY DE LOS GRANDES NUMEROS

Para obtener una estimación de la media de una variable aleatoria, de la cual solamente se conocen algunos valores que se obtienen al observar, en forma independiente los resultados del experimento que representa la variable, se puede hacer mediante el promedio aritmético de dichas observaciones.

Esta afirmación se puede respaldar mediante la ley de los grandes números, si se considera que el resultado de cada una de las observaciones, es una variable aleatoria que por haberse obtenido del mismo experimento tiene la misma distribución de probabilidad y por consiguiente la misma media y la misma variancia. Formalmente esta ley se puede establecer como sigue:

LEY DE LOS GRANDES NUMEROS

Sean $X_1, X_2, \ldots, X_n$ una sucesión de variables aleatorias independientes con la misma distribución de probabilidad y por lo tanto con la misma media μ_X y la misma variancia σ_X^2. Entonces, la sucesión de variables aleatorias, definidas como:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \quad \text{donde} \quad n = 1, 2, 3, \ldots$$

converge en probabilidad a μ_X; esto es:

$$\boxed{P\left[|\bar{X}_n - \mu_X| > \varepsilon\right] \to 0 \quad \text{cuando} \quad n \to \infty} \qquad \ldots (5)$$

Demostración:

Por la desigualdad de Chebyshev:

$$P\left[\left|\bar{X}_n - E\left[\bar{X}_n\right]\right| > k\,\sigma_{\bar{X}_n}\right] \le \frac{1}{k^2} \qquad \text{para} \quad k > 0 \quad \ldots (a)$$

donde:

$$E\left[\bar{X}_n\right] = \frac{1}{n} E\left[\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n}\left\{E\left[X_1\right] + E\left[X_2\right] + \ldots + E\left[X_n\right]\right\}$$

pero:

$$E\left[X_1\right] = E\left[X_2\right] = \ldots = E\left[X_n\right] = \mu_X$$

con lo cual:

$$E\left[\bar{X}_n\right] = \frac{n\,\mu_X}{n} = \mu_X \qquad \ldots (6)$$

por otra parte $\bar{X}_n$ es una combinación lineal de las variables independientes X_i por lo cual, siguiendo un procedimiento similar al efectuado en la expresión (37) del capítulo IV, se obtiene:

$$\sigma^2_{\bar{X}_n} = \left(\frac{1}{n}\right)^2 \sigma^2_{X_1} + \left(\frac{1}{n}\right)^2 \sigma^2_{X_2} + \ldots + \left(\frac{1}{n}\right)^2 \sigma^2_{X_n}$$

$$= n\left(\frac{1}{n}\right)^2 \sigma^2_X = \frac{\sigma^2_X}{n}$$

de manera que:

$$\sigma_{\bar{X}_n} = \frac{\sigma_X}{\sqrt{n}} \qquad \ldots (7)$$

sustituyendo las expresiones (6) y (7) en (a):

$$P(|\bar{X}_n - \mu_X| > k\frac{\sigma_X}{\sqrt{n}}) \leq \frac{1}{k^2} \qquad \text{para} \quad k > 0$$

haciendo $k = \frac{\varepsilon\sqrt{n}}{\sigma_X}$ se tiene que:

$$P(|\bar{X}_n - \mu_X| > \varepsilon) \leq \frac{\sigma^2_X}{\varepsilon^2 n}$$

con lo cual:

$$\lim_{n \to \infty} P(|\bar{X}_n - \mu_X| > \varepsilon) \leq \lim_{n \to \infty} \frac{\sigma^2_X}{\varepsilon^2 n} = 0$$

La ley de los grandes números también se puede usar para respaldar la interpretación frecuentista de la probabilidad, al definir una sucesión de variables aleatorias en la forma:

$$Y_n = \sum_{i=1}^{n} \frac{X_i}{n}$$

donde $X_i (i = 1, 2, \ldots, n)$ son variables aleatorias con distribución de Bernoulli que representan el número de éxitos o de veces que se presenta un evento A en un ensayo (observación) y Y_n es la frecuencia relativa del evento A cuando se hacen n observaciones.

En el capítulo III correspondiente a los modelos probabilísticos comunes, se demostró que la media de una variable con distribución de Bernoulli es la probabilidad de éxito p y de acuerdo con la ley de los grandes números:

$$\boxed{P\left[|Y_n - p| > \varepsilon\right] \to 0 \quad \text{cuando} \quad n \to \infty}$$

lo cual reafirma que la frecuencia relativa de un evento, tiende a la probabilidad de ocurrencia del mismo, cuando el número de observaciones tiende a infinito.

V.4 TEOREMA DEL LIMITE CENTRAL

Como se mencionó en el capítulo III, la distribución normal es una de las distribuciones de probabilidad más importantes, entre otras razones, por la gran cantidad de fenómenos físicos que con ella se pueden representar en forma aceptable.

Esta afirmación se fundamenta en el teorema del límite central, que en una de sus versiones dice lo siguiente:

Teorema V.2

Sean $X_1, X_2, \ldots, X_n$ una sucesión de variables aleatorias independientes con la misma distribución de probabilidad y, por consiguiente, con la misma media μ_X y la misma variancia σ_X^2. Entonces la sucesión de variables aleatorias $Z_1, Z_2, \ldots, Z_n$ definidas como:

$$Z_n = \frac{\overline{X}_n - \mu_X}{\sigma_X/\sqrt{n}} \qquad \text{para} \quad n = 1, 2, \ldots$$

en donde:

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

tiende a la distribución normal estándar cuando n tiende a infinito.

Demostración:

Debido a que la función generatriz de momentos de una distribución de probabilidad es única, se demostrará que la función generatriz de momentos de Z_n converge a la función generatriz de momentos de la distribución normal estándar.

De la definición de función generatriz de momentos:

$$M_Z(s) = E\left[e^{sZ}\right]$$

$$= E\left[e^{s(\overline{X}_n - \mu_X)/\sigma_X/\sqrt{n}}\right]$$

como $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

$$M_Z(s) = E\left[\prod_{i=1}^{n} e^{s(X_i - \mu_X)/\sigma_X/\sqrt{n}}\right]$$

considerando que las variables X_i tienen la misma distribución:

$$M_Z(s) = E\left[M_{\frac{X - \mu_X}{\sigma_X}}\left(\frac{s}{\sqrt{n}}\right)\right]^n$$

obteniendo logaritmos naturales en ambos miembros:

$$\ln M_Z(s) = n \ln M_{\frac{X-\mu_X}{\sigma_X}}\left(\frac{s}{\sqrt{n}}\right)$$

expandiendo en serie de Taylor $M_{\frac{X-\mu_X}{\sigma_X}}\left(\frac{s}{\sqrt{n}}\right)$

$$\ln M_Z(s) = n \ln \left[1 + m_1\left(\frac{s}{\sqrt{n}}\right) + \frac{m_2}{2!}\left(\frac{s}{\sqrt{n}}\right)^2 + \ldots + \frac{m_k}{k!}\left(\frac{s}{\sqrt{n}}\right)^k + \ldots\right]$$

para $|s| < 1$

donde $m_k (k = 1, 2, 3, \ldots)$ son los k-ésimos momentos con respecto a la media.

De ellos $m_1 = 0$ y $m_2 = 1$ puesto que $\frac{X - \mu_X}{\sigma_X}$ es una variable normalizada, de manera que:

$$\ln M_Z(s) = n \ln \left[1 + \frac{s^2}{2n} + \ldots + \frac{m_k}{k!}\left(\frac{s}{\sqrt{n}}\right)^k + \ldots\right]$$

haciendo:

$$P(s) = \frac{s^2}{2n} + \ldots + \frac{m_k}{k!}\left(\frac{s}{\sqrt{n}}\right)^k + \ldots$$

se tiene que:

$$\ln M_Z(s) = n \ln [1 + P(s)]$$

por otra parte, se sabe que:

$$\ln [1 + b] = b - \frac{b^2}{2} + \ldots + \frac{(-1)^{k-1} b^k}{k} \quad \text{para } |b| < 1$$

entonces:

$$\ln M_Z(s) = n \left[P(s) - \frac{P(s)^2}{2} + \ldots + \frac{(-1)^{k-1} P(s)^k}{k}\right] ;$$

para $|P(s)| < 1$

obteniendo el límite cuando n tiende a infinito:

$$\lim_{n \to \infty} n P(s) = \lim_{n \to \infty} n \left[\frac{s^2}{2n} + \ldots + \frac{m_k}{k!}\left(\frac{s}{\sqrt{n}}\right)^k\right]$$

$$= \frac{s^2}{2}$$

y para los demás términos:

$$\lim_{n \to \infty} n \frac{(-1)^{k-1} P(s)^k}{k} = 0 \quad \text{para} \quad k = 2, 3, \ldots$$

con lo cual:

$$\lim_{n \to \infty} \ell n\, M_Z(s) = \frac{s^2}{2}$$

tomando antilogaritmos:

$$\lim_{n \to \infty} M_Z(s) = e^{\frac{s^2}{2}}$$

Esta función generatriz de momentos corresponde a una variable con distribución normal estándar, con lo cual se demuestra que cuando $n \to \infty$, $\overline{X}_n$ tiende a ser una variable con distribución normal estándar.

Otra versión del teorema del límite central es la siguiente.

Teorema V.3

Sean $X_1, X_2, \ldots, X_n$ una sucesión de variables aleatorias independientes, con la misma distribución de probabilidad, la misma media y la misma variancia. Entonces la sucesión de variables aleatorias $Z_1, Z_2, \ldots, Z_n$; definida como:

$$Z_n = \frac{Y_n - n\,\mu_X}{\sigma_X \sqrt{n}} \qquad \text{para} \quad n = 1, 2, \ldots$$

en donde:

$$Y_n = \sum_{i=1}^{n} X_i$$

tiende a la distribución normal estándar cuando n tiende a infinito.

La demostración de esta versión es muy similar a la anterior. Su importancia radica en que Y_n es la suma de las variables X_i y no el promedio, como en el caso anterior.

También existen otras versiones en las cuales se debilitan las premisas de los teoremas anteriores, en el sentido de que las variables X_i no necesitan tener la misma distribución (aunque tampoco deben existir variables dominantes), o que no todas sean independientes entre sí.

Por todo lo anterior, la distribución normal se puede utilizar para representar el comportamiento de variables aleatorias que estén formadas por la suma de un gran número de componentes, como por ejemplo la carga total de un autobús, si se considera que dicha carga es el peso total de todas las personas que transporta.

Ejemplo V.3

La empresa Ferrocarriles Nacionales de México, ha determinado que con un nuevo equipo S, requiere en promedio 6 horas para renovar un kilómetro de vía de ferrocarril, con una desviación estándar de 3 horas.

¿Cuál es la probabilidad de que se tarden más de 550 hrs., en reemplazar 85 km de vía?

Solución

Si Y es la variable aleatoria que representa el tiempo total y $X_i (i = 1, 2, \ldots, 85)$ representa el tiempo necesario para reemplazar todas las vías hasta el kilómetro i, por el teorema del límite central, la variable:

$$Y = X_1 + X_2 + \ldots + X_{85}$$

tiene una distribución aproximadamente normal con media $\mu_Y = n\,\mu_X$ y desviación estándar $\sigma_Y = \sqrt{n}\;\sigma_X$, con lo cual:

$$P[Y > 600] = P[Z > Z_c] = 1 - P[Z \leq Z_c]$$

donde:

$$Z_c = \frac{550 - 85(6)}{\sqrt{85}\,(3)} = 1.45$$

buscando en la tabla de la distribución normal estándar:

$$\begin{aligned} P[Y > 600] &= 1 - P(Z \leq 1.45) \\ &= 1 - (0.92648) \\ &= 0.07352 \end{aligned}$$

V.5 APROXIMACION DE LA DISTRIBUCION DE POISSON A LA BINOMIAL

Como se vio en el capítulo III la distribución binomial tiende a la de Poisson cuando $n \to \infty$, $p \to 0$ y λ permanece constante. Por esta razón, cuando n es suficientemente grande se puede usar una de ellas, como una aproximación de la otra. Generalmente la distribución de Poisson se utiliza como una aproximación de la binomial, debido a que el cálculo de las combinaciones $C(n, x)$ en la distribución binomial se complica, conforme aumenta el parámetro n.

Empíricamente se ha observado que la aproximación es aceptable cuando $n > 20$ y $P < 0.05$, sin embargo dependiendo del problema, se pueden modificar estos criterios.

Ejemplo V.4

El encargado de un taller mecánico asegura que, en promedio, recibe reclamaciones de sus clientes en el 2% de las reparacio-

nes. Si en una semana se repararon 140 vehículos y se considera que las reclamaciones son independientes, ¿cuál es la probabilidad de que se reciban cinco reclamaciones?

Solución

Considerando las reclamaciones como éxitos, la variable X que representa el número de reclamaciones tiene distribución binomial con parámetros $n = 140$ y $p = 0.02$, con lo cual:

$$
\begin{aligned}
p_X(5) &= C(n, x)\, p^x\, q^{n-x} \\
&= C(140, 5)(0.02)^5(0.98)^{135} \\
&= \frac{(140)(139)(138)(137)(136)}{5!}(0.02)^5(0.98)^{135} \\
&= 0.08725
\end{aligned}
$$

Aproximando con la distribución de Poisson:

$$\lambda = n\,p = (140)(0.02) = 2.8$$

$$
\begin{aligned}
p_X(5) &= \frac{\lambda^x e^{-\lambda}}{x!} \\
&= \frac{(2.8)^5 e^{-2.8}}{5!} = 0.08721
\end{aligned}
$$

Debe notarse que la distribución de Poisson es más fácil de calcular cuando el valor de n es grande, y en este caso, el error debido a la aproximación es sólo de cuatro cienmilésimas.

V.6 APROXIMACION A LA DISTRIBUCION BINOMIAL Y POISSON POR MEDIO DE LA NORMAL

La distribución binomial se ha definido como el número de éxitos en una secuencia de n ensayos de Bernoulli; sin embargo, también se puede definir como la suma de n variables aleatorias X_i, $(i = 1, 2, \ldots, n)$, con distribución de Bernoulli, de igual media p y variancia $p(1 - p)$. Se sigue entonces que si

$$Y_n = \sum_{i=1}^{n} X_i$$

tiene distribución binomial, por el teorema del límite central, la distribución de probabilidad de Y_n tiende a la forma normal cuando n tiende a infinito.

De manera análoga se puede definir la distribución de Poisson como la suma de n variables $X_i (i = 1, 2, \ldots, n)$, que representen el número de éxitos en cada uno de los n pequeños subintervalos Δt de la variable continua t. Con lo cual se puede afirmar que la distribución Poisson tiende a la normal cuando $\lambda = n\,p$ tiende a infinito.

Empíricamente se ha encontrado que la aproximación es buena cuando $\lambda = n\,p > 5$.

Ejemplo V.5

En relación al ejemplo V.4, ¿cuál es la probabilidad de que lleguen cuando mucho 5 clientes a reclamar por los servicios de una misma semana?

Solución

Usando la distribución binomial sería:

$$P(0 \le X \le 5) = p_X(0) + p_X(1) + p_X(2) + p_X(3) + p_X(4) + p_X(5)$$

$$= 0.0591 + 0.1688 + 0.2395 + 0.2248 + 0.1572 + 0.0873$$

$$= 0.9387$$

Mediante la distribución normal:

$$P(0 \le X \le 5) \approx P(z_1 \le Z \le z_2)$$

donde:

$$z_1 = \frac{0 - \mu_X}{\sigma_X}$$

$$z_2 = \frac{5 - \mu_X}{\sigma_X}$$

como X tiene distribución binomial:

$$\mu_X = n\,p = (140)(0.02) = 2.8$$

$$\sigma_X = \sqrt{n\,p(1 - p)} = \sqrt{2.8(0.98)} = 1.657$$

entonces:

$$z_1 = \frac{-\,2.8}{1.657} = -\,1.69$$

$$z_2 = \frac{5 - 2.8}{1.657} = 1.33$$

Buscando en la tabla A de la distribución normal:

$$P(-1.69 \le Z \le 1.33) = 0.90824 - 0.04551 = 0.86273$$

El error que se comete con esta aproximación se puede reducir al considerar que la distribución binomial es discreta, la normal es continua, y que la primera está definida para valores mayores o iguales que cero, mientras la segunda se define para cualquier valor de Z.

En la figura V.2 se puede observar que a cada valor $p_X(x_0)$ de la figura (a) le debe corresponder un área equivalente de la (b), para ello se puede calcular el área, tomando media unidad antes y media unidad después de x_0.

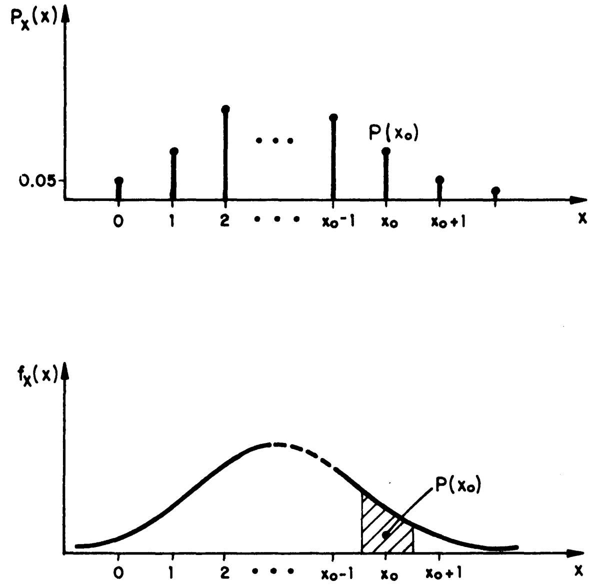


Figura V.2

Por otra parte, en $p_X(0)$ se puede considerar que se concentra la probabilidad de que la variable aleatoria continua tome un valor menor o igual a cero.

Aceptando estas dos consideraciones, la aproximación del ejemplo V.3 será:

$$P(0 \leq x \leq 5) \approx P(Z \leq z_0)$$

donde:

$$z_0 = \frac{x + \frac{1}{2} - \mu_X}{\sigma_X} = \frac{5.5 - 2.8}{1.657} = 1.63$$

La corrección se hizo sumando media unidad, porque la probabilidad buscada incluye a $x = 5$. Sin embargo, si no estuviera incluido, se debería restar media unidad.

Por lo tanto, de la tabla A se tiene $P(z \leq 1.63) = 0.9484$.

Al ajuste anterior se le llama *ajuste por continuidad* y es recomendable que siempre se realice para mejorar la aproximación.

Ejemplo V.6

El número de vehículos que pasan por la caseta de cobro de una carretera, obedece a una distribución Poisson, con parámetros $\lambda = 45$ vehículos por minuto. Si actualmente se pueden atender hasta 55 vehículos por minuto, ¿cuál es la probabilidad de que se sobrepase la capacidad de la caseta?

Solución

Si X tiene distribución de Poisson:

$$\mu_X = \lambda = 45$$

$$\sigma_X = \sqrt{\lambda} = \sqrt{45} = 6.71$$

Usando la distribución normal, como aproximación de la Poisson

$$P(x > 55) \simeq P(z \geq z_0)$$

donde:

$$z_0 = \frac{(x - 0.5) - \mu_X}{\sigma_X} = \frac{55 - 0.5 - 45}{6.71} = 1.42$$

Obsérvese que en este caso la corrección por continuidad se hizo restando 0.5 al valor crítico de x; porque la probabilidad de que X sea mayor de 55, se aproxima mediante el área que se encuentra a la derecha de z_0. Esto se muestra en la siguiente figura.

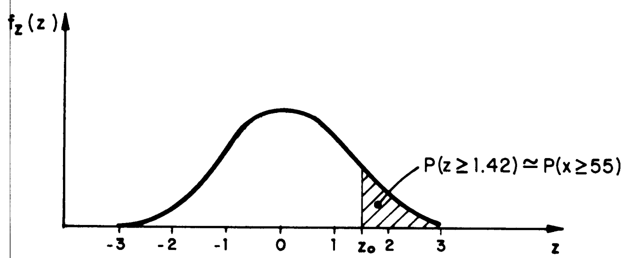


Figura V.3

Buscando en la tabla A de la distribución normal:

$$P(Z \geq 1.42) = 1 - P(Z \leq 1.42) = 1 - 0.9222 = 0.0778$$

Esto significa que en el 7.78% del tiempo, el número de vehículos es mayor o igual que 55.

CAPITULO VI ESTADISTICA DESCRIPTIVA

INTRODUCCION

En la actualidad la estadística ha llegado a ser un instrumento de uso cotidiano para todos los profesionistas que están en contacto con fenómenos de naturaleza aleatoria, y que a partir del conocimiento de ciertos datos cuantitativos del fenómeno, deben tomar decisiones sobre su comportamiento general.

Algunas de las aplicaciones de la estadística son:

a) Presentar en forma ordenada y resumida la información registrada en una encuesta, entrevista, cuestionario, etc.

b) Pronosticar el tamaño de la población en la zona metropolitana del Valle de México para el año 2,000.

c) Probar la eficacia de una nueva terapia o producto farmacéutico.

d) Establecer controles de calidad en la producción de artículos.

Para su estudio, la estadística se divide en: *estadística descriptiva* e *inferencia estadística*.

La primera se encarga de la recopilación, organización, resumen y presentación de los datos numéricos obtenidos de la observación de un fenómeno; mientras que la segunda tiene por objeto, obtener conclusiones probables sobre el comportamiento general del fenómeno, a partir de algunas observaciones particulares del mismo.

En este capítulo se estudiará la estadística descriptiva y en el siguiente se tratará lo correspondiente a la inferencia estadística.

VI.1 CONCEPTOS GENERALES

Para el estudio de la estadística descriptiva, se requiere definir los siguientes conceptos:

POBLACION

Es un conjunto de objetos, llamados comúnmente elementos, que tienen en común una o varias características particulares que se desean estudiar.

MUESTRA

Es un subconjunto de la población.

De manera más específica se puede considerar que algunos de los elementos de la población (o muestra) representan las características de los objetos más que a los objetos mismos.

Así por ejemplo, una población se puede definir como el conjunto formado por todos los vehículos que circulan en la zona metropolitana de la Ciudad de México. La característica específica de estudio podría ser la cantidad de contaminante que arroja cada vehículo a la atmósfera durante una hora, y una muestra sería un pequeño número de vehículos *seleccionados al azar* a los que se les mide la cantidad de contaminante que arrojan.

La selección de una muestra es una etapa muy importante dentro del estudio estadístico, debido a que la información que presenta la muestra es la base para hacer suposiciones o inferencias sobre lo que ocurre en la población.

Si en el ejemplo anterior se selecciona una muestra de 10 vehículos relativamente nuevos, posiblemente el estudio que se realice a partir de ella señale que no existen problemas graves de contaminación, lo cual sería una suposición equivocada porque el muestreo que se siguió no fue el adecuado y, en consecuencia, la muestra no sería *representativa* de la población.

Para que una muestra sea representativa de la población, se debe establecer un proceso de muestreo en el que todos los elementos de la población tengan la misma posibilidad de ser seleccionados y, cuando sea posible, que la selección de cada elemento sea independiente de las demás.

Lo anterior significa que al elegir los elementos de una muestra no debe haber preferencia por algunos de ellos, ni deben seleccionarse en función de lo que se observe en los anteriores.

Para obtener una muestra representativa, existen diferentes técnicas, entre las cuales se encuentran las siguientes:

MUESTREO ALEATORIO

Consiste en formar una lista de todos los elementos de la población, enumerarlos y hacer la selección mediante la generación de números aleatorios con distribución uniforme.

Para generar números aleatorios con distribución uniforme se puede usar una tabla de dígitos aleatorios como la que se muestra en el apéndice de estos apuntes, o mediante fórmulas de recurrencia diseñadas con ese propósito. En este último caso, los números obtenidos se denominan seudoaleatorios, debido a que con el mismo valor inicial, se obtiene la misma secuencia de números.

El muestreo aleatorio es recomendable cuando la población es suficientemente pequeña, como para ser enumerada.

MUESTREO SISTEMATICO

En este tipo de muestreo también se elabora una lista con los elementos de la población; pero en lugar de seleccionarlos de manera aleatoria, se recorre la lista y se va seleccionando cada k-ésimo elemento, iniciando aleatoriamente con uno de los primeros k.

El muestreo sistemático es más sencillo de aplicar que el anterior. Sin embargo, tiene la limitación de no poderse aplicar a poblaciones demasiado grandes, ni tampoco cuando los datos presentan periodicidad, puesto que ésta puede coincidir con el período de selección k.

MUESTREO ESTRATIFICADO

En esta técnica, la población se divide en clases o estratos para hacer posteriormente una selección, que puede ser aleatoria o sistemática dentro de cada estrato. La definición de cada clase debe ser suficientemente clara para evitar que uno de los elementos se pueda ubicar en dos clases diferentes.

El número de elementos que se seleccionan de cada clase puede ser proporcional al tamaño del estrato cuando la diferencia entre ellos es muy grande, o pueden ser iguales cuando el tamaño de los estratos es semejante.

MUESTREO POR CONGLOMERADOS

Es semejante al muestreo estratificado, en el sentido de definir grupos de elementos; sin embargo, esta técnica se aplica cuando la población es homogénea y existen grupos ya definidos.

Debido a la homogeneidad de la población no se requiere seleccionar elementos de todos los conglomerados y, en ocasiones, es suficiente con seleccionar uno de los conglomerados con todos sus elementos.

Para realizar la selección por conglomerados, se puede utilizar el muestreo aleatorio, considerando grupos en lugar de elementos individuales.

VI.2 CLASIFICACION Y ORGANIZACION DE LOS DATOS DE UNA MUESTRA

Cuando los datos de una muestra se encuentran desordenados es difícil obtener información directamente de ellos. Una forma natural de ordenarlos es de manera ascendente o descendente, sobre todo cuando la muestra es pequeña. Sin embargo, cuando se trata de una muestra grande, el procedimiento anterior se hace muy laborioso y una vez ordenada la muestra es difícil manejarla.

Mediante un ejemplo se presentará el procedimiento que generalmente se utiliza para presentar los datos de una muestra grande en forma resumida y ordenada.

Ejemplo VI.1

Los siguientes datos representan la longitud en centímetros de 80 piezas de madera que fueron seleccionadas en una bodega. Se desea presentar los datos en forma resumida y ordenada.

50.1	50.6	51.1	50.8	52.2	51.9	51.2	52.0
50.6	49.1	51.8	51.0	50.8	51.8	51.1	49.7
50.7	51.4	51.9	50.4	51.7	51.0	49.5	52.0
51.1	51.8	50.3	51.5	51.7	50.3	49.9	49.7
52.0	51.3	51.1	50.8	49.4	50.3	51.1	51.2
50.8	51.5	51.1	51.2	50.3	51.3	51.7	51.8
51.4	51.0	51.7	50.1	52.1	51.0	52.8	51.1
49.9	50.9	50.2	51.5	51.0	50.2	49.6	51.3
51.8	50.3	50.5	51.7	51.7	50.4	49.6	51.2
51.3	51.2	51.6	51.5	51.9	51.6	53.1	51.8

Como primer paso, se debe calcular el *rango de la muestra* que se define como la diferencia entre el mayor y el menor de los elementos. Así, para los datos anteriores:

$$\text{rango} = 53.1 - 49.1 = 4.0$$

Para ordenar los ochenta datos de la muestra, se clasifican en varios intervalos, denominados *intervalos de clase*.

La fijación de los intervalos de clase depende del criterio del analista, aunque debe tomar en cuenta lo siguiente:

El número de intervalos que se puede establecer depende de la cantidad de datos que contiene la muestra (tamaño de la muestra) y de la dispersión de los mismos.

En general se recomienda establecer entre 5 y 15 intervalos, procurando que no queden intervalos vacíos dentro del rango de valores.

También debe tenerse cuidado en la fijación de los límites de cada intervalo para evitar, por un lado, la posibilidad de que un mismo elemento pertenezca a dos intervalos diferentes y, por otro, que la magnitud de los intervalos sea difícil de manejar.

Si todos los datos comprendidos en un intervalo se distribuyen de manera uniforme, se considera que el punto medio del intervalo puede representar a todos los valores de la muestra que se encuentran en él. A dicho punto se le denomina *marca de clase*.

En las dos primeras columnas de la tabla VI.1 se muestran los intervalos y marcas de clase (t_r) de la muestra que se está ordenando.

Intervalos r	Límites del intervalo	Marcas de clase t_r	Frecuencia f_r	Frecuencia relativa f'_r	Frec. relativa acumulada F'_r
1	49.0-49.9	49.5	9	0.11	0.11
2	50.0-50.9	50.5	20	0.25	0.36
3	51.0-51.9	51.5	44	0.55	0.91
4	52.0-52.9	52.5	6	0.08	0.99
5	53.0-53.9	53.5	1	0.01	1.00

Tabla VI.1

Al número de elementos de la muestra que pertenece a un intervalo de clase r se le llama *frecuencia del intervalo* y se representa como f_r. La suma de las frecuencias deberá ser igual al número total de elementos de la muestra de tamaño n, o sea:

$$\sum_{r=1}^{m} f_r = n$$

donde m es el número de intervalos de clase (para el ejemplo $m = 5$).

Al cociente de la frecuencia entre el número total de datos muestrales se le llama *frecuencia relativa* y se representa como f'_r, esto es:

$$\boxed{f'_r = \frac{f_r}{n}} \qquad \ldots (1)$$

La frecuencia relativa de un intervalo de clase se puede interpretar como la proporción de los datos que se encuentran en el intervalo correspondiente.

Finalmente, se define *la frecuencia relativa acumulada* F'_r como la suma de las frecuencias relativas hasta el r-ésimo intervalo, con lo cual:

$$\boxed{F'_r = \sum_{j=1}^{r} f'_j} \qquad \ldots (2)$$

En la tabla VI.1 se presenta la información de la muestra en forma agrupada (resumida y ordenada) y en ella se encuentran los resultados de los conceptos definidos anteriormente.

VI.3 DISTRIBUCIONES EMPIRICAS

Para determinar la distribución de probabilidad de una variable aleatoria que representa un fenómeno o el modelo probabilístico teórico más aproximado a ella, es útil construir la gráfica de las frecuencias, frecuencias relativas o frecuencias relativas acumuladas.

Para representar las frecuencias o las frecuencias relativas se usa generalmente *el histograma* o *el polígono de frecuencias*. En el histograma, la frecuencia se considera constante en todos los puntos de cada intervalo de clase, por lo que se representa como una sucesión de rectángulos del mismo ancho y cuyas alturas corresponden a las frecuencias o a las frecuencias relativas acumuladas de los intervalos correspondientes.

La siguiente figura muestra el histograma del ejemplo que se está analizando.

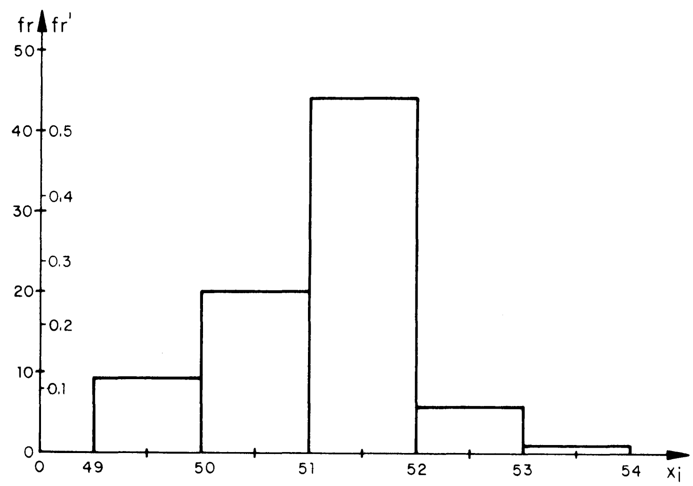


Figura VI.1

El polígono de frecuencias es otra gráfica que muestra la distribución de frecuencias. Para construirla se marca sobre el mismo sistema de ejes del histograma una sucesión de puntos, cuyas abscisas son las marcas de clase y las ordenadas son las frecuencias o las frecuencias relativas correspondientes. Posteriormente se unen mediante rectas todos los puntos consecutivos. Para cerrar el polígono de frecuencias en los extremos, se consideran otros intervalos con frecuencia cero.

El polígono de frecuencias del ejemplo VI.1 se muestra en la siguiente figura:

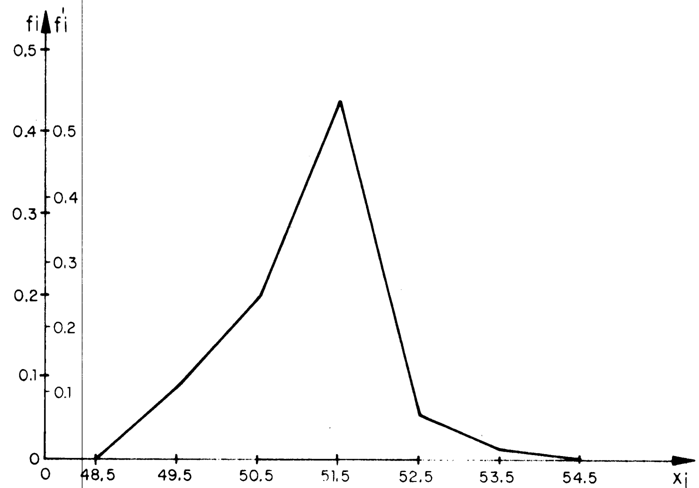


Figura VI.2

Por la ley de los grandes números, la frecuencia relativa de cada intervalo de clase se puede considerar como una aproximación de la probabilidad de que la variable aleatoria tome un valor dentro del intervalo de clase y, por consiguiente, el polígono de frecuencias se puede considerar como una aproximación de la función densidad de probabilidad de dicha variable aleatoria.

Otra gráfica de uso frecuente para estimar la distribución de probabilidad de la variable aleatoria es el *polígono de frecuencias relativas acumuladas u ojiva.*

La manera de graficarlo consiste en unir, mediante rectas, una sucesión de puntos cuyas abscisas son los límites superiores de cada intervalo de clase y las ordenadas son las frecuencias relativas acumuladas F_i' (ó F_r).

En la figura VI.3 se muestra el polígono de frecuencias relativas acumuladas que corresponde al ejemplo sobre el que se está trabajando.

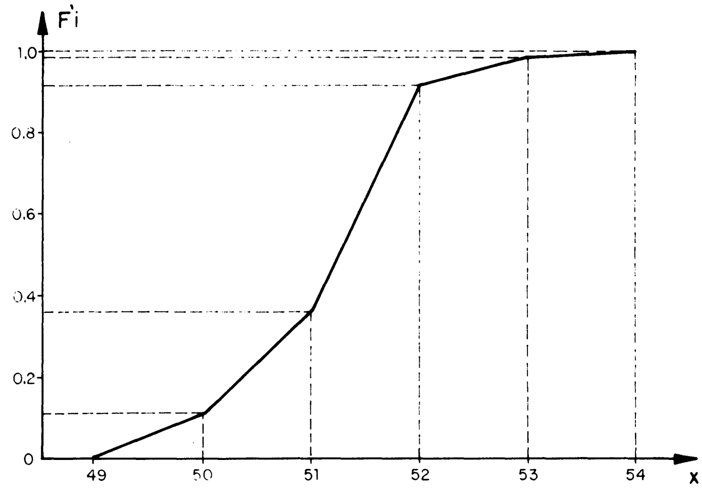


Figura VI.3

El polígono de frecuencias relativas acumuladas puede considerarse como una aproximación de la función de distribución acumulada de la variable aleatoria que representa a la población.

VI.4 MEDIDAS DESCRIPTIVAS DE UNA DISTRIBUCION DE FRECUENCIAS

En el capítulo II se presentaron algunos indicadores que describen a una distribución de probabilidad; así, la media representa la abscisa del centroide de la función masa o densidad de probabilidad, y la variancia es una medida que describe la dispersión de los posibles valores de X con respecto a la media. Estos indicadores se representan mediante letras griegas como μ, σ, ρ, etc., y se les llama *parámetros*.

En el caso de una distribución de frecuencias, también se pueden establecer medidas descriptivas y, para distinguirlas de los parámetros, se usan letras latinas como $\bar{x}$, s, r, etc.

Todas las medidas descriptivas que se definen a continuación para una distribución de frecuencias, se pueden definir también de manera semejante para una distribución de probabilidad.

Como se verá en el siguiente capítulo, las medidas descriptivas de una distribución de frecuencia se utilizan para estimar los parámetros correspondientes que describen a la población, debido a que si la muestra es representativa, sus características son semejantes a las de la población.

Generalmente se utilizan cuatro tipos de medidas descriptivas que son:

a) De tendencia central.

b) De dispersión.

c) De sesgo o asimetría.

d) De curtosis o apuntamiento.

VI.4.1 MEDIDAS DE TENDENCIA CENTRAL

Con estas medidas se busca un valor que pueda representar a toda la muestra, por encontrarse en el centro de ella, desde diferentes puntos de vista.

Las más importantes medidas de tendencia central son las siguientes:

MEDIA ($\bar{x}$)

Se define como el promedio aritmético de los datos de una muestra. Para diferenciarla de la media de la población se representa con $\bar{x}$ en lugar de μ_X, de esta manera:

$$\boxed{\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i} \qquad \ldots (3)$$

donde x_i es i-ésimo elemento de la muestra.

Si los datos de la muestra están agrupados en una tabla de frecuencias, se considera que la marca de clase (t_r) representa a los valores x_i que se encuentran dentro del intervalo r-ésimo, de manera que la suma de todos los valores x_i que se encuentran en un intervalo de clase, equivale a sumar f_r veces t_r, con lo cual la expresión (3) se puede escribir como:

$$\bar{x} = \frac{1}{n} \sum_{r=1}^{m} t_r f_r$$

donde m es el número de intervalos de clase.

Para simplificar la expresión se puede tomar en cuenta que la frecuencia relativa:

$$f'_r = \frac{f_r}{n},$$

entonces:

$$\boxed{\bar{x} = \sum_{r=1}^{m} t_r\, f'_r} \qquad \ldots (4)$$

Ejemplo VI.2

El número de automóviles que se vende semanalmente en una agencia distribuidora, se ha registrado durante diez semanas con los siguientes resultados:

2,5,3,4,4,4,5,1,1,3

¿Cuál es la media muestral?

Solución

Como la muestra es pequeña, no es necesario agrupar los datos en una tabla de frecuencias y se puede usar directamente la expresión (3), con lo cual:

$$\bar{x} = \frac{1}{10}\left[2+5+3+4+4+4+5+1+1+3\right] = 3.2 \text{ automóviles}$$

Ejemplo VI.3

Encontrar la media de los datos del problema VI.1.

Solución

Debido a que los datos están agrupados, de la expresión (4):

$$\bar{x} = 49.5(0.11) + 50.5(0.25) + 51.5(0.55) + 52.5(0.08) + 53.5(0.01)$$

$$= 51.13 \text{ cm.}$$

MEDIANA (me)

Es el valor que corresponde a la mitad de los datos ordenados de una muestra.

Para encontrar la mediana de una muestra pequeña, se ordenan sus datos en forma ascendente o descendente y si el tamaño de la muestra (n) es un número impar, entonces la mediana será el elemento que ocupe la posición (n + 1)/2 de la muestra ordenada, debido a que dicho valor la divide en dos partes, con el mismo número de elementos en cada una. Sin embargo, cuando el número de elementos es par, no hay un valor que se encuentre exactamente a la mitad de los elementos, en este caso se pueden promediar los dos valores más cercanos a la mitad.

Ejemplo VI.4

¿Cuál es la mediana de la muestra del ejemplo VI.2?

Solución

Ordenando los elementos en forma ascendente, se obtiene:

1,1,2,3,3,4,4,4,5,5

como son diez elementos, la mediana se encuentra entre el quinto (3) y el sexto elemento (4), por lo cual:

$$me = \frac{3 + 4}{2} = 3.5$$

Cuando la muestra es grande y sus elementos se encuentran agrupados, la mediana puede obtenerse determinando primero el intervalo que contiene a la mediana, el cual se distingue porque es el primero que tiene una frecuencia relativa acumulada mayor o igual a 0.5 y, posteriormente, mediante una interpolación lineal se encuentra el valor de x que corresponde a la frecuencia relativa acumulada de 0.5.

Con el polígono de frecuencias relativas acumuladas se puede aproximar el valor de la mediana, trazando una línea horizontal que parte de F' = 0.5 hasta cruzar el polígono y, posteriormente, con otra recta vertical que parte del punto de la intersección, se encuentra, en el eje de las abscisas, el valor de la mediana.

La figura VI.4 muestra la localización gráfica de la mediana sobre un polígono de frecuencias relativas acumuladas.

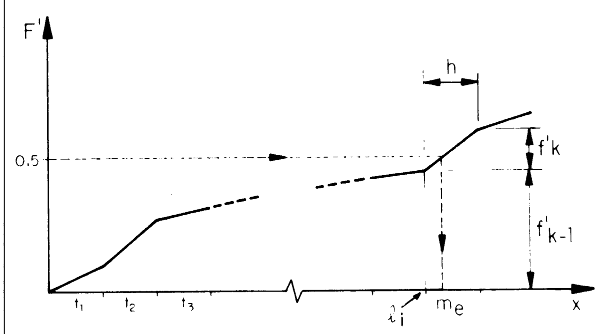


Figura VI.4

Para hacer la interpolación lineal, se usa la ecuación de una recta, de la forma:

$$y = y_0 + m(x - x_0)$$

que en este caso, de acuerdo con la figura VI.4:

$$y = 0.5; \quad y_0 = F'_{k-1}; \quad m = \frac{f'_k}{h}; \quad x = me$$

y

$$x_0 = \ell_i$$

donde:

k = al subíndice que corresponde al intervalo de clase que contiene a la mediana,

F'_{k-1} = a la frecuencia relativa acumulada hasta el intervalo $k - 1$,

f'_k = a la frecuencia relativa del intervalo k,

h = al tamaño del intervalo de clase y,

ℓ_i = al límite inferior del intervalo k.

Sustituyendo los valores:

$$0.5 = F'_{k-1} + \frac{f'_k}{h}(me - \ell_i)$$

despejando a la mediana:

$$\boxed{me = \ell_i + \frac{h}{f'_k}(0.5 - F'_{k-1})} \qquad \ldots (5)$$

Ejemplo VI.5

A partir de los datos de la tabla VI.1, encontrar la mediana de la muestra correspondiente.

Solución

En la tabla se puede observar que la mediana se encuentra en el intervalo comprendido de 51 a 51.9, debido a que es el primero cuya frecuencia relativa acumulada es mayor de 0.5, con lo cual de la expresión (5):

$$me = 51 + \frac{1}{0.55}[0.5 - 0.36]$$

$$= 51.25$$

MODA O MODO (mo)

Se define como el elemento de la muestra que tiene la máxima frecuencia, es decir, aquél que más se repite.

Una muestra puede tener dos o más modas, en cuyo caso se dice que es bimodal o multimodal. Cuando todos los elementos de la muestra son diferentes, entonces no tiene sentido hablar de ella, porque puede considerarse que todos los elementos son la moda o que no existe, por estas características, su uso está muy limitado.

Cuando la muestra es pequeña, la moda se determina directamente por inspección, mientras que en muestras grandes, con datos agrupados, se puede aproximar con la marca de clase del intervalo modal, que es el que tiene la máxima frecuencia.

En algunos casos se puede mejorar la aproximación, considerando que la moda es la abscisa del máximo de una curva hipotética que pasa por las marcas de clase, como lo muestra la siguiente figura.

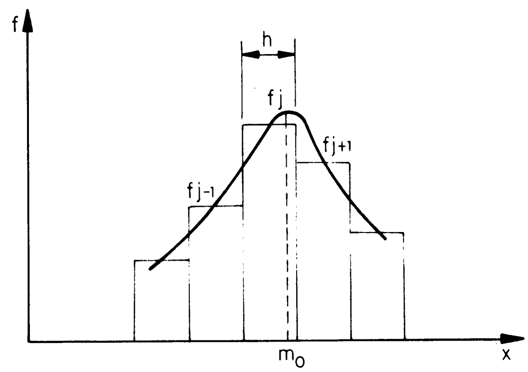


Figura VI.5

De acuerdo con esto, se puede considerar que la moda debe pertenecer al intervalo de clase con máxima frecuencia, pero proporcionalmente más cercano al intervalo adyacente que le siga en frecuencia, de esta manera, se puede plantear la proporción:

$$\frac{m_0 - \ell_i}{h} = \frac{f_{j+1}}{f_{j-1} + f_{j+1}}$$

donde:

j = al subíndice que corresponde al intervalo de clase modal,

ℓ_j = al límite inferior del intervalo j,

despejando la moda:

$$\boxed{mo = \ell_j + h \frac{f_{j+1}}{f_{j-1} + f_{j+1}}} \qquad \dots (6)$$

Para la muestra del ejemplo VI.2 puede verse que el valor que más se repite es cuatro, por lo cual, esa es la moda de la muestra.

En el ejemplo VI.1 se observa que la moda de la muestra es 51.1 (se presenta siete veces); sin embargo, para ilustrar el uso de la expresión (6), se puede estimar la moda como:

$$\text{moda} = 51 + (1) \frac{6}{20 + 6} = 51.23$$

VI.4.2 MEDIDAS DE DISPERSION

Como su nombre lo indica, las medidas de dispersión reflejan la separación o alejamiento de los elementos de una muestra.

RANGO

Al principio del capítulo fue definido como una medida de dispersión, puesto que indica la máxima separación entre los datos.

VARIANCIA

En el capítulo II se definió la variancia de una variable aleatoria como el segundo momento con respecto a la media. De manera análoga, para una distribución de frecuencias se define el momento k-ésimo con respecto a la media como:

$$\boxed{m_k = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^k} \qquad \dots (7)$$

y en consecuencia, la *variancia muestral* se define como:

$$S_x^2 = m_2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Para simplificar el cálculo de la variancia, se puede desarrollar el binomio en la forma:

$$S_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i^2 - 2x_i \bar{x} + \bar{x}^2)$$

por las propiedades de la sumatoria:

$$S_x^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \frac{2\bar{x}}{n} \sum_{i=1}^{n} x_i + \frac{1}{n} \sum_{i=1}^{n} \bar{x}^2$$

como:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

entonces:

$$S_x^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - 2\bar{x}^2 + \frac{n\bar{x}^2}{n}$$

sumando los términos semejantes:

$$\boxed{S_x^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2} \qquad \ldots (8)$$

Si los datos están agrupados, cada marca de clase representa a los valores que se encuentran dentro del intervalo correspondiente, con lo cual:

$$S_x^2 = \frac{1}{n} \sum_{r=1}^{m} f_r t_r^2 - \bar{x}^2$$

donde:

n = al tamaño de la muestra,

m = al número de intervalos de clase,

f_r = a la frecuencia del intervalo r y

t_r = a la marca de clase del intervalo r.

como la frecuencia relativa $f_r' = \frac{f_r}{n}$, entonces:

$$\boxed{S_x^2 = \sum_{r=1}^{m} f_r' t_r^2 - \bar{x}^2} \qquad \ldots (9)$$

DESVIACION ESTANDAR

Es otra medida de dispersión que para una distribución de frecuencias, como en una distribución de probabilidad, es la raíz cuadrada de la variancia. Se representa con s_x, de esta manera, de la expresión (8):

$$s_x = \sqrt{s_x^2} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2} \qquad \ldots (10)$$

y cuando los datos están agrupados, de (9):

$$\boxed{s_x = \sqrt{\sum_{i=1}^{m} f_r' \, t_r^2 - \bar{x}^2}} \qquad \ldots (11)$$

COEFICIENTE DE VARIACION

Para una distribución de frecuencias, se define como el cociente de la desviación estándar muestral entre la media muestral; esto es:

$$\boxed{C.V. = \frac{s_x}{\bar{x}}} \qquad \ldots (12)$$

El significado y utilidad de los tres indicadores anteriores son los mismos que en las distribuciones de probabilidad tratadas en el capítulo II.

Ejemplo VI.6

La producción de mineral de fierro en siete países latinoamericanos durante 1980, fue la siguiente:

1,9.7,0.5,8.7,7,5,9,16 (miles de toneladas).

Obtener:

a) La media.

b) La desviación estándar.

c) El coeficiente de variación.

Solución

a) Como la muestra es pequeña, de la expresión (3):

$$\bar{x} = \frac{1 + 9.7 + 0.5 + 8.7 + 7 + 5.9 + 16}{7}$$

$$= 6.97 \text{ (miles de toneladas)}$$

b) De (8):

$$S_x = \sqrt{\frac{1}{7}\left[1^2 + 9.7^2 + 0.5^2 + 8.7^2 + 7^2 + 5.9^2 + (16)^2\right] - (6.97)^2}$$

$$= 4.94 \text{ (miles de toneladas)}$$

c) Por la expresión (12):

$$C.V. = \frac{4.94}{6.97} = 0.71$$

VI.4.3 MEDIDAS DE ASIMETRIA

Uno de los objetivos más importantes de la estadística descriptiva es proporcionar información sobre la muestra, que pueda ser de utilidad para determinar las características de toda la población.

Cuando se desea ajustar un modelo probabilístico, como los vistos en el capítulo III, a un fenómeno particular es conveniente comparar la forma del histograma o polígono de frecuencias de una muestra del fenómeno, con la de la función de probabilidad del modelo teórico.

Para describir la forma de la distribución de frecuencias de una muestra, se usa, entre otros indicadores, la *asimetría o sesgo*. Una distribución de frecuencias es simétrica si el tercer momento de la muestra con respecto a la media es igual a cero ($m_3 = 0$); en tal caso la media divide en dos partes iguales a la distribución de frecuencias y además, cualquiera de las partes es un reflejo de la otra.

Si una distribución de frecuencias es simétrica, la media, la mediana y la moda coinciden en el mismo punto, como lo muestra la figura VI.6(a). Sin embargo, cuando la figura no es simétrica se puede presentar una *asimetría positiva*, siendo $m_3 > 0$; o bien una *asimetría negativa* donde $m_3 < 0$.

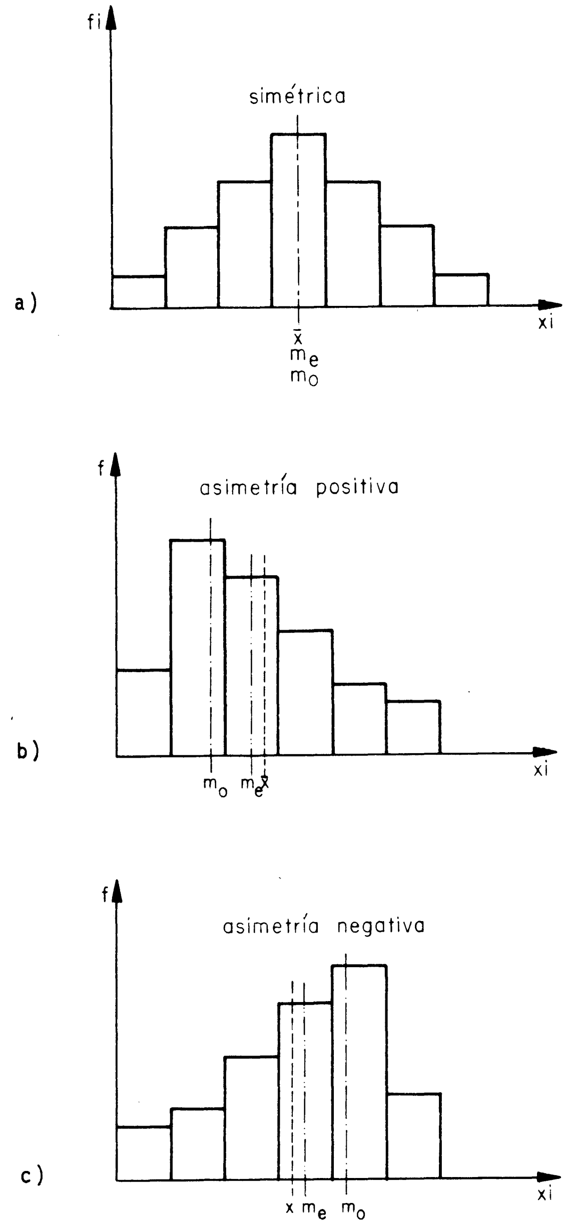


Figura VI.6

En las gráficas (b) y (c) de la figura anterior se presenta un ejemplo de cada tipo de asimetría.

Para medir en forma adimensional la asimetría de una distribución de frecuencias, se utiliza el *coeficiente de asimetría por momentos* (a_3), que se define como el cociente del tercer momento con respecto a la media entre la raíz cuadrada del segundo momento con respecto a la media elevada al cubo; esto es:

$$\boxed{a_3 = \frac{m_3}{\sqrt{m_2^3}}} \qquad \ldots (13)$$

Debe aclararse que si los datos están agrupados, se pueden aproximar los momentos con respecto a la media, con la expresión:

$$\boxed{m_k = \sum_{r=1}^{m} f_r'(t_r - \bar{x})^k} \qquad \ldots (14)$$

Otra forma de medir la asimetría de una distribución es mediante el *coeficiente de Pearson* (s) que se define como:

$$\boxed{s = \frac{\text{media} - \text{moda}}{S_x}}$$

Este indicador tiene la desventaja de que sólo se aplica cuando la distribución es unimodal y se puede demostrar que $-1 \leq s \leq 1$.

VI.4.4 CURTOSIS

Es otra característica que permite describir la forma de la distribución de frecuencias, también conocida como *apuntamiento, exceso o aplanamiento*. Este último nombre es tal vez el menos indicado, pues el significado de curtosis es contrario al de aplanamiento y por lo tanto, una curtosis grande implica poco aplanamiento y viceversa.

En la figura VI.7 se muestra un ejemplo de los tres tipos en los cuales se clasifican las distribuciones de frecuencia de acuerdo con su curtosis.

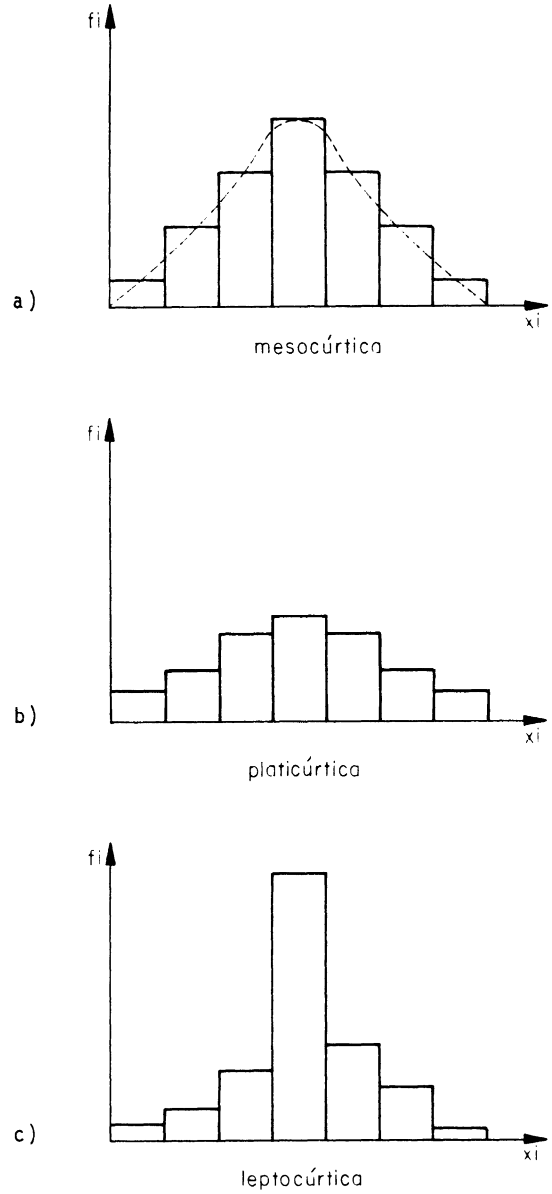


Figura VI.7

La referencia que se usa para establecer la curtosis media o *mesocúrtica*, como la mostrada en la figura VI.7(a), es la gráfica de la función densidad de probabilidad de la distribución normal estándar.

La gráfica (b) es más aplanada y su curtosis es menor que la mesocúrtica, mientras que en (c) la curtosis es mayor y consecuentemente su forma es más apuntada.

A estos dos casos se les llama respectivamente platicúrtica y leptocúrtica.

Para medir la curtosis se puede usar el cuarto momento con respecto a la media (m_4); sin embargo, para que la curtosis no dependa de las unidades de la variable en estudio, se define el *coeficiente de curtosis* como:

$$\boxed{a_4 = \frac{m_4}{m_2^2} - 3} \qquad \ldots (15)$$

Cuando la distribución es mesocúrtica se cumple que $\frac{m_4}{m_2^2} = 3$ por lo cual en la expresión (15) se resta este valor para que la referencia se encuentre en cero. De esta manera si $a_4 < 0$, la distribución es platicúrtica y si $a_4 > 0$, se trata de una distribución leptocúrtica.

Ejemplo VI.7

A partir de los datos del ejemplo VI.2, determinar por momentos:

a) El coeficiente de asimetría.

b) El coeficiente de curtosis.

Solución

Para obtener la media y los momentos con respecto a la media que se requieren para calcular los coeficientes de asimetría y curtosis, se utilizan las expresiones (4) y (14) respectivamente y mediante una tabla, como la siguiente, se pueden facilitar y ordenar las operaciones.

		$\overline{x}$	m_1	m_2	m_3	m_4
t_r	f'_r	$t_r\ f'_r$	$(t_r - \overline{x})$	$f'_r(t_r - \overline{x})^2$	$f'_r(t_r - \overline{x})^3$	$f'_r(t_r - \overline{x})^4$
9.5	0.050	0.475	-4.330	0.937	-4.059	17.576
10.5	0.067	0.704	-3.330	0.743	-2.474	8.239
11.5	0.067	0.771	-2.330	0.364	-0.848	1.975
12.5	0.250	3.125	-1.330	0.442	-0.588	0.782
13.5	0.133	1.796	-0.330	0.014	-0.005	0.002
14.5	0.133	1.929	0.670	0.060	0.040	0.027
15.5	0.116	1.798	1.670	0.324	0.540	0.902
16.5	0.067	1.106	2.670	0.478	1.275	3.405
17.5	0.050	0.875	3.670	0.673	2.472	9.071
18.5	0.050	0.925	4.670	1.090	5.092	23.781
19.5	0.017	0.331	5.670	0.547	3.099	17.570
		13.835	7.370	5.672	4.544	82.970

a) De la tabla anterior y de la expresión (13) el coeficiente de asimetría es:

$$a_3 = \frac{4.544}{\sqrt{(5.672)^3}} = 0.336$$

b) De (15) el coeficiente de curtosis es:

$$a_4 = \frac{82.97}{(5.672)^2} - 3 = -0.421$$

Por los resultados anteriores se concluye que la distribución de frecuencias tiene asimetría positiva y es más aplanada que la distribución normal estándar.

VI.4.5 OTRAS MEDIDAS DESCRIPTIVAS (LOS FRACTILES)

En algunas ocasiones resulta conveniente dividir una muestra ordenada en dos o más grupos con el mismo número de elementos cada uno. *Los fractiles, cuantiles o percentiles* son valores que pueden o no pertenecer a la muestra y cumplen con esa función. Así, la mediana que se presentó anteriormente es un fractil que divide la muestra en dos partes iguales.

Los cuartiles, que son otro tipo de fractiles, dividen a la muestra ordenada en cuatro partes iguales, de tal manera que el *primer cuartil o fractil 25%*, es mayor o igual que el 25% de los datos; el segundo cuartil coincide con la mediana y el tercer cuartil o fractil 75% es mayor o igual que el 75% de los elementos de la muestra.

De manera semejante se definen los *deciles*, que dividen a la muestra en grupos que contienen 10% de los datos cada uno.

A la diferencia entre el primer y el tercer cuartil se le llama *rango intercuartílico* y se puede considerar como una medida de dispersión, pues entre esos cuartiles están el 50% de los elementos centrales de la muestra.

También se define el *rango interdecílico* como la diferencia entre el primer y el noveno decil, el cual considera el 80% de los valores centrales de la muestra.

Para calcular los fractiles de una muestra, se puede usar la expresión (5) que se obtuvo para calcular la mediana (fractil 50%) de una muestra, considerando que en lugar de 0.5, debe aparecer la fracción que corresponde al fractil que se desea (F^*), de esta manera se tiene:

$$\boxed{\text{fractil } F^* = \ell_k + \frac{h}{f'_k}\left[F^* - F'_{k-1}\right]} \qquad \ldots (16)$$

donde k es el subíndice que corresponde al intervalo de clase que contiene al fractil F^*.

Ejemplo VI.8

De los registros del Servicio Meteorológico Nacional se han obtenido y ordenado las velocidades máximas del viento (en m/seg.) de cada mes, durante cinco años. La tabla de frecuencias es la siguiente:

Intervalos de clase	t_i	f_i	f'_i	F'_i
9 - 9.9	9.5	3	0.050	0.050
10 - 10.9	10.5	4	0.067	0.117
11 - 11.9	11.5	4	0.067	0.184
12 - 12.9	12.5	15	0.250	0.434
13 - 13.9	13.5	8	0.133	0.567
14 - 14.9	14.5	8	0.133	0.700
15 - 15.9	15.5	7	0.116	0.816
16 - 16.9	16.5	4	0.067	0.883
17 - 17.9	17.5	3	0.050	0.933
18 - 18.9	18.5	3	0.050	0.983
19 - 19.9	19.5	1	0.017	1.000

Calcular:

a) La mediana.

b) El rango intercuartílico.

Solución

a) El intervalo de clase que contiene la mediana es de 13 a 13.9, pues su frecuencia relativa acumulada es la primera superior a 0.5.

De manera que por la expresión (5):

$$me = \text{fractil } 50\% = 13 + \frac{1}{0.133}(0.5 - 0.434) = 13.496 \left[m/seg\right]$$

b) Calculando primero los cuartiles uno (q_1) y tres (q_3) con la expresión (16):

$$q_1 = \text{fractil } 25\% = 12 + \frac{1}{0.25}\left[0.25 - 0.184\right]$$
$$= 12.264$$

$$q_3 = \text{fractil } 75\% = 15 + \frac{1}{0.116}\left[0.75 - 0.70\right]$$
$$= 15.431$$

con lo cual el rango intercuartílico es:

$$15.431 - 12.264 = 3.167$$

Los fractiles en general no se pueden clasificar en un solo tipo de indicador, pues la mediana es una medida de tendencia central; los rangos intercuartílicos e interdecílicos son medidas de dispersión y los cuartiles, deciles y demás fractiles no se pueden clasificar en ninguna de ellas.

CAPITULO VII INFERENCIA ESTADISTICA

INTRODUCCION

En cualquier actividad profesional, se presenta frecuentemente la necesidad de describir el comportamiento de fenómenos aleatorios, tales como la resistencia de un material, la respuesta de cierto sistema físico o la demanda de un producto. Para lograrlo se puede partir de la información obtenida de una muestra.

En estadística se le llama *inferencia* al proceso de inducir o bien deducir las características o parámetros poblacionales a partir de la información muestral, midiendo con probabilidades la incertidumbre inherente.

Si no existiera la inferencia estadística, no sería posible diseñar reglas confiables para un control de calidad. Por ejemplo, prácticamente sería imposible determinar si todas las varillas que serán utilizadas en una construcción satisfacen las especificaciones de diseño, y mucho menos, si para probarlas se someten a pruebas destructivas.

Para hacer una inferencia estadística, se pueden estimar los parámetros descriptivos de la población a partir de la información obtenida en una muestra, ya sea como un valor puntual, como un intervalo, o bien establecer valores hipotéticos de los parámetros y probar estadísticamente si son válidas las hipótesis.

A los dos primeros casos se les llama *estimación estadística puntual o por intervalos*, y al segundo se le denomina *prueba de hipótesis*.

VII.1 DISTRIBUCIONES MUESTRALES

Antes de presentar las técnicas para realizar la inferencia estadística es necesario definir el concepto de estadístico muestral, dado que es la base para hacer cualquier tipo de inferencia.

Un *estadístico muestral* es una variable aleatoria que está en función de uno o más elementos de la muestra.

Como ejemplos se pueden mencionar las medidas descriptivas de una muestra ($\bar{X}$, S_X ó S_X^2), las cuales fueron presentadas en el capítulo anterior.

Para comprender por qué se dice que los estadísticos son variables aleatorias, considérese un conjunto de muestras representativas de una población, todas con el mismo número de elementos. Si en cada muestra se calcula un mismo estadístico como $\bar{X}$, es casi seguro que se encontrarán valores distintos, sin embargo la diferencia puede ser pequeña; además, como antes de calcularlo no es posible predecir con certeza el valor que tomará, entonces se concluye que el estadístico es una variable aleatoria.

Como variable aleatoria, cada estadístico tiene una distribución de probabilidad denominada *distribución muestral*.

No debe confundirse la distribución muestral de un estadístico con la distribución de la variable que representa a la población.

DISTRIBUCION MUESTRAL DE LA MEDIA

Los valores que pueden tomar los elementos de una muestra de tamaño n se representan mediante variables aleatorias $X_1, X_2, \ldots, X_n$. Si la muestra es representativa de la población, todas estas variables son independientes entre sí y tienen la misma distribución de probabilidad por pertenecer a la misma población, entonces por el teorema del límite central la media muestral $\bar{X}$ definida por:

$$\boxed{\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i}$$

tiende en probabilidad a una variable con distribución normal, cuando n tiende a infinito. Empíricamente se ha establecido que una muestra es suficientemente grande cuando $n > 30$.

Como $\bar{X}$ es una combinación lineal de las variables X_i con, $i = 1, 2, \ldots, n$. entonces, de acuerdo con lo visto en el capítulo V:

$$\boxed{\mu_{\bar{X}} = \mu_X} \quad \text{y} \quad \boxed{\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}} \qquad \ldots (1)$$

Ejemplo VII.1

El peso promedio de los melones que se producen en una huerta es de 1.7 kilogramos, con una desviación estándar de 0.4 kg. ¿Cuál es la probabilidad de que un posible comprador tome una muestra aleatoria de 40 unidades y encuentre que el peso promedio de la muestra es menor de 1.55 kg?

Solución

En este caso la variable aleatoria es el peso promedio de la muestra $\overline{X}$ y de la expresión (1):

$$\mu_{\overline{x}} = 1.7 \text{ kg} \quad \text{y} \quad \sigma_{\overline{x}} = \frac{0.4}{\sqrt{40}} = 0.063 \text{ kg}$$

con lo cual:

$$P(\overline{X} < 1.55) = P(Z < z_c)$$

donde:

$$z_c = \frac{\overline{X} - \mu_{\overline{x}}}{\sigma_{\overline{x}}} = \frac{1.55 - 1.7}{0.063} = -2.38$$

buscando en la tabla A del apéndice:

$$P(\overline{X} < 1.55) = 0.0087$$

esta probabilidad se muestra en la siguiente figura:

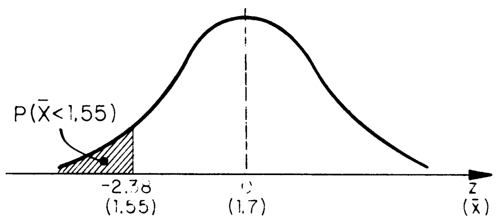


Figura VII.1

Cuando la población es finita y la selección de las muestras se hace sin reemplazar las unidades observadas, la desviación estándar (o error estándar) del estadístico $\overline{X}$ se altera ligeramente, debido a que conforme crece la muestra, cambia la probabilidad de que una unidad sea seleccionada. Para reducir esta alteración, se debe multiplicar la desviación estandar de la muestra por el factor:

$$\sqrt{\frac{N - n}{N - 1}} \qquad \ldots (2)$$

donde:

N es el tamaño de la población

Debe notarse que cuando N tiende a infinito el factor tiende a uno y cuando la muestra y la población son iguales, el factor se hace nulo.

Ejemplo VII.2

Considerando el enunciado del problema VII.1, ¿cuál es la probabilidad de que el peso promedio sea menor de 1.55 kg, si el muestreo se efectúa sin reemplazo y en la huerta se tienen únicamente 500 melones?

Solución

Como la población es finita (N = 500) y n = 40, de las expresiones (1) y (2):

$$\sigma_{\overline{x}} = \frac{0.4}{\sqrt{40}} \left[\frac{500 - 40}{500 - 1}\right]^{1/2} = 0.06$$

con lo cual:

$$z_c = \frac{1.55 - 1.7}{0.06} = -2.5$$

por lo tanto:

$$P(\overline{X} < 1.55) = 0.0062$$

Comparando este resultado con el anterior, se nota el efecto que provocan el tamaño de la población y el tipo de muestreo.

DISTRIBUCION DE UNA FUNCION DE LA VARIANCIA MUESTRAL

Otro de los estadísticos que tienen gran utilidad en la inferencia estadística es la variancia muestral S_X^2. Como variable aleatoria la distribución muestral de S_X^2 tiene escasa importancia. Sin embargo, para realizar inferencias estadísticas sobre variancias o desviaciones estándar se usa la variable aleatoria:

$$\boxed{\chi^2 = \frac{n\, S_X^2}{\sigma_X^2}} \qquad \ldots (3)$$

que tiene distribución ji-cuadrada, con $\nu = n - 1$ grados de libertad, siempre y cuando la población sea normal con media μ_X y variancia σ_X^2. Para demostrarlo se partirá de la definición de variancia muestral:

$$S_X^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

multiplicando ambos miembros por n y haciendo:

$$X_i - \bar{X} = X_i - \mu_X - \bar{X} + \mu_X$$

$$n\, S_X^2 = \sum_{i=1}^{n} \left[(X_i - \mu_X) - (\bar{X} - \mu_X)\right]^2$$

desarrollando el binomio:

$$n\, S_X^2 = \sum_{i=1}^{n} \left[(X_i - \mu_X)^2 - 2(X_i - \mu_X)(\bar{X} - \mu_X) + (\bar{X} - \mu_X)^2\right]$$

$$= \sum_{i=1}^{n} (X_i - \mu_X)^2 - \sum_{i=1}^{n} 2(X_i - \mu_X)(\bar{X} - \mu_X) + \sum_{i=1}^{n} (\bar{X} - \mu_X)^2$$

por las propiedades de la sumatoria:

$$n\, S_X^2 = \sum_{i=1}^{n} (X_i - \mu_X)^2 - 2(n\bar{X} - n\mu_X)(\bar{X} - \mu_X) + n(\bar{X} - \mu_X)^2$$

con lo cual:

$$S_X^2 = \frac{1}{n} \left[\sum_{i=1}^{n} (X_i - \mu_X)^2 - n(\bar{X} - \mu_X)^2\right] \qquad \ldots (4)$$

sustituyendo en la expresión (3):

$$\chi^2 = \frac{1}{\sigma_X^2} \left[\sum_{i=1}^{n} (X_i - \mu_X)^2 - n(\bar{X} - \mu_X)^2\right]$$

$$= \frac{\sum_{i=1}^{n} (X_i - \mu_X)^2}{\sigma_X^2} - \frac{(\bar{X} - \mu_X)^2}{\sigma_X^2/n}$$

De esta manera χ^2 está formada por la suma de los cuadrados de $n - 1$ variables aleatorias independientes, con distribución normal estándar; por lo tanto, de acuerdo con la definición de la distribución ji-cuadrada:

$$\chi^2 = \frac{n\, S_X^2}{\sigma_X^2}$$

tiene distribución ji-cuadrada con $\nu = n - 1$ grados de libertad.

Ejemplo VII.3

El espesor de las placas de acero que produce una empresa tiene una desviación estándar de 0.243 mm. Si dicho espesor obedece a una distribución normal, ¿cuál es la probabilidad de que en una muestra de 20 unidades se obtenga una desviación estándar menor de 0.240?

Solución

Usando la variable $\chi^2 = \frac{n\, S_X^2}{\sigma_X^2}$ para calcular probabilidades relacionadas con S_X, se tiene que para $S_X = 0.240$

$$\chi^2 = \frac{20(0.240)^2}{(0.243)^2} = 19.51$$

entonces:

$$P(S_X < 0.240) = P(\chi^2 < 19.51)$$

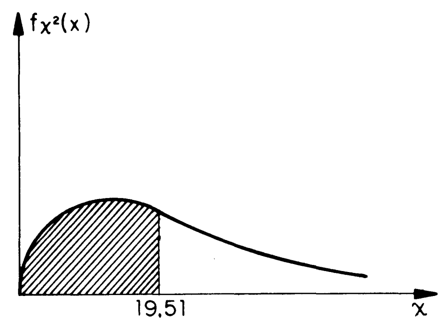


Figura VII.2

Buscando en la tabla B que se encuentra en el apéndice, la distribución ji-cuadrada con $\nu = 20 - 1 = 19$ y de acuerdo con la figura VII.2, se obtiene que $F(18.34) = 0.5$ y $F(22.72) = 0.75$.

Interpolando linealmente:

$$F(19.51) = 0.5 + \frac{(0.75 - 0.5)}{(22.72 - 18.34)} (19.51 - 18.34) = 0.5667$$

Esto significa que en una muestra de 20 unidades:

$$P(S_X < 0.24) = 0.5667$$

VII.2 ESTIMADORES ESTADISTICOS

Como se mencionó al inicio del capítulo, en la estadística clásica existen dos formas para estimar los parámetros:

Estimación puntual, cuando solamente se da un valor del parámetro desconocido.

Estimación por intervalos de confianza, cuando se fijan dos valores entre los cuales se encuentra el parámetro desconocido, con cierta probabilidad.

A continuación se presentan cada una de estas modalidades.

VII.2.1 ESTIMADORES PUNTUALES

Las medidas descriptivas de una muestra, vistas en el capítulo anterior, se usan generalmente como estimadores puntuales de los parámetros poblacionales; sin embargo, otro aspecto importante se refiere a la estimación de parámetros como p cuando se sabe que la población es binomial, o λ si la población tiene distribución de Poisson.

En esta sección se definen algunas características de los estimadores puntuales que se utilizan para seleccionar, entre varios de ellos, el más conveniente. También se presenta un método para obtenerlos.

Entre las características más importantes de un estimador se encuentran el sesgo, la eficiencia y la consistencia. A continuación se define cada una de ellas.

ESTIMADOR INSESGADO

Se dice que un estimador $\hat{\theta}$ del parámetro θ es insesgado si $E\left[\hat{\theta}\right] = \theta$, es decir que el valor esperado o medio del estimador es igual al parámetro que se quiere conocer; en caso contrario se dice que es sesgado. Así por ejemplo, el estadístico:

$$\hat{\theta} = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

es un estimador insesgado del parámetro $\theta = \mu_X$ puesto que $E\left[\overline{X}\right] = \mu_{\overline{X}} = \mu_X$, como se vio al estudiar la distribución muestral de la media.

Ejemplo VII.4

Demostrar si el estadístico $S_X^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$ es un estimador insesgado de σ_X^2

Solución

De la expresión (4):

$$S_X^2 = \frac{1}{n} \left[\sum_{i=1}^{n} (X_i - \mu_X)^2 - n(\overline{X} - \mu_X)^2 \right]$$

aplicando el operador esperanza en ambos miembros:

$$E\left[S_X^2\right] = \frac{1}{n} \left\{ \sum_{i=1}^{n} E\left[(X_i - \mu_X)^2\right] - n\, E\left[(\overline{X} - \mu_X)^2\right] \right\}$$

sin embargo, como $\mu_{\overline{X}} = \mu_X$ y $\sigma_{\overline{X}}^2 = \sigma_X^2/n$

$$E\left[(X_i - \mu_X)^2\right] = \sigma_X^2 \quad \text{para} \quad i = 1, 2, \ldots$$

$$E\left[(\overline{X} - \mu_X)^2\right] = \frac{\sigma_X^2}{n}$$

con lo cual:

$$E\left[S_X^2\right] = \frac{1}{n}\left[n\, \sigma_X^2 - n\, \frac{\sigma_X^2}{n}\right]$$

de donde:

$$E\left[S_X^2\right] = \frac{(n-1)\, \sigma_X^2}{n}$$

Con esto se demuestra que $E\left[S_X^2\right] \neq \sigma_X^2$ y por lo tanto es sesgado.

Se puede probar que si se define otro estadístico $\hat{S}_X^2$ como

$$\hat{S}_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \mu_X)^2$$

se tendrá un estimador insesgado de σ_X^2.

También se puede demostrar que, a pesar de que la variancia muestral es sesgada, la desviación estándar de una muestra es un estimador insesgado de la desviación estándar de la población.

ESTIMADOR EFICIENTE

Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados del parámetro θ, entonces $\hat{\theta}_1$ es más eficiente que $\hat{\theta}_2$ si la variancia de $\hat{\theta}_1$ es menor que la de $\hat{\theta}_2$; es decir $\sigma^2_{\hat{\theta}_1} < \sigma^2_{\hat{\theta}_2}$.

En capítulos anteriores se vio que la variancia es una medida de la dispersión que tienen los datos con respecto a su media. En el caso de un estimador insesgado, su media es igual al parámetro deseado y, por lo tanto, es conveniente tener estimadores que difieran poco del parámetro que se desea estimar.

Se puede dar el caso de tener un estimador insesgado y otro u otros que no lo sean, pero con menor variancia que el primero. Cuando esto ocurre, se tendrá que seleccionar el más adecuado de acuerdo con el uso que se le dará.

Ejemplo VII.5

Sean X_1, X_2, X_3 y X_4 variables aleatorias independientes que representan los posibles valores de una muestra representativa de una población con distribución normal, cada una con media μ_X y variancia σ_X^2. ¿Cuál de los siguientes estimadores ($\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$) se debe seleccionar para estimar la media de la población?

$$\hat{\theta}_1 = X_1$$

$$\hat{\theta}_2 = X_2 + X_3$$

$$\hat{\theta}_3 = \frac{X_1 + 2X_2 + 2X_3}{5}$$

Solución

Para determinar si son insesgados:

$$E\left[\hat{\theta}_1\right] = E\left[X_1\right] = \mu_X$$

$$E\left[\hat{\theta}_2\right] = E\left[X_2\right] + E\left[X_3\right] = 2\,\mu_X$$

$$E\left[\hat{\theta}_3\right] = \frac{E\left[X_1\right] + 2\,E\left[X_2\right] + 2\,E\left[X_3\right]}{5} = \frac{5\,\mu_X}{5} = \mu_X$$

De esta manera $\hat{\theta}_1$ y $\hat{\theta}_3$ son insesgados, mientras que $\hat{\theta}_2$ no lo es.

Para comparar la eficiencia, se debe obtener la variancia de cada estimador, considerando que son una combinación lineal de variables aleatorias:

$$\text{var}\left[\hat{\theta}_1\right] = \text{var}\left[X_1\right] = \sigma_X^2$$

$$\text{var}\left[\hat{\theta}_2\right] = \text{var}\left[X_1\right] + \text{var}\left[X_2\right]$$

$$= \sigma_X^2 + \sigma_X^2 = 2\,\sigma_X^2$$

$$\text{var}\left[\hat{\theta}_3\right] = \left(\frac{1}{5}\right)^2 \text{var}\left[X_1\right] + \left(\frac{2}{5}\right)^2 \text{var}\left[X_2\right] + \left(\frac{2}{5}\right)^2 \text{var}\left[X_3\right]$$

$$= \frac{1}{25}\,\sigma_X^2 + \frac{4}{25}\,\sigma_X^2 + \frac{4}{25}\,\sigma_X^2 = \frac{9}{25}\,\sigma_X^2$$

Como $\text{var}\left[\hat{\theta}_3\right] < \text{var}\left[\hat{\theta}_1\right] < \text{var}\left[\hat{\theta}_2\right]$, el estimador $\hat{\theta}_3$ es el más eficiente y por ser insesgado se puede considerar el mejor.

ESTIMADOR CONSISTENTE

Un estimador $\hat{\theta}_n$ obtenido de una muestra de tamaño n es consistente, si $\hat{\theta}_n$ converge en probabilidad a θ; es decir:

$$\boxed{\lim_{n \to \infty} P\{|\hat{\theta}_n - \theta| > \varepsilon\} = 0}$$

para cualquier constante positiva ε.

De acuerdo con la definición anterior y por la ley de los grandes números:

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

es un estimador consistente del parámetro μ_X, pues a medida que aumenta n, $\overline{X}$ tiende a μ_X.

También por la ley de los grandes números se puede asegurar que cualquier estimador insesgado es consistente. Sin embargo, existen algunos que son consistentes, aun cuando sean sesgados, como:

$$S_X^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

que a pesar de ser sesgado se puede demostrar que es consistente.

METODO DE LA MAXIMA VEROSIMILITUD

Fue desarrollado por R.A. Fisher. Puede demostrarse que con este método se obtienen estimadores eficientes y consistentes, sin embargo algunos de ellos son sesgados.

Para definir el método, considérese un conjunto de variables aleatorias independientes, que representan a los posibles valores de una muestra aleatoria de tamaño n, cada una con cierta función masa o densidad de probabilidad.

$$p_{X_i}(x_i) \text{ ó } f_{X_i}(x_i); \quad i = 1, 2, \ldots, n$$

A la función masa o densidad de probabilidad conjunta

$$\left.\begin{array}{l} P(x_1, x_2, x_3, \ldots, x_n) = P(x_1)\,P(x_2)\,\ldots\,P(x_n) \quad \text{ó} \\ \\ f(x_1, x_2, \ldots x_n) = f(x_1)\,f(x_2)\,\ldots\,f(x_n) \end{array}\right\} \ldots \quad (5)$$

se le llama función de verosimilitud.

Como los elementos de la muestra pertenecen a la misma población, las funciones de densidad o de probabilidad de las variables $x_1, x_2, \ldots, x_n$ son iguales y tienen los mismos parámetros $(\theta_1, \theta_2, \ldots)$.

El método de máxima verosimilitud consiste en encontrar el valor del parámetro o los parámetros $\theta_1, \theta_2, \ldots,$ de tal manera que la función de verosimilitud tome un valor máximo.

Para resaltar el parámetro o los parámetros que se desean encontrar, la función de verosimilitud se puede representar como:

$$P(x_1, x_2, \ldots, x_n/\theta_1, \theta_2, \ldots)$$

o en el caso continuo, como:

$$f(x_1, x_2, \ldots, x_x/\theta_1, \theta_2, \ldots)$$

Ejemplo VII.6

Encontrar el estimador de máxima verosimilitud para el parámetro θ de una variable aleatoria X cuya F.D.P. está dada por:

$$f_T(t) = \begin{cases} \theta e^{-\theta t} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

Solución

Como es una variable continua, de la expresión (5), la función de verosimilitud es:

$$f(t_1, t_2, \ldots, t_n/\theta) = \begin{cases} (\theta e^{-\theta t_1})(\theta e^{-\theta t_2}) \ldots (\theta e^{-\theta t_n}); & t_i \geq 0;\ i = 1, 2, \ldots n \\ 0 & ;\ t_i < 0 \end{cases}$$

analizando el intervalo en el que $t \geq 0$:

$$f(t_1, t_2, \ldots, t_n/\theta) = \theta^n e^{-\theta[t_1 + t_2 + \ldots + t_n]} \qquad \ldots \text{ (a)}$$

haciendo:

$$\bar{t} = \frac{1}{n} \sum_{i=1}^{n} t_i$$

entonces:

$$\sum_{i=1}^{n} t_i = n\,\bar{t}$$

sustituyendo en (a):

$$f(t_1, t_2, \ldots, t_n/\theta) = \theta^n e^{-\theta n \bar{t}}$$

derivando con respecto a θ para maximizar la función:

$$\frac{df(t_1, t_2, \ldots, t_n/\theta)}{d\theta} = \theta^n(-n\bar{t}e^{-\theta n\bar{t}}) + e^{-\theta n\bar{t}}\left[n\,\theta^{n-1}\right]$$

$$= n\,\theta^{n-1} e^{-\theta n \bar{t}} (1 - \theta\bar{t})$$

igualando a cero:

$$n\,\theta^{n-1} e^{-\theta n t} (1 - \theta\bar{t}) = 0$$

Esta ecuación se satisface si $\theta = 0$ o el factor $(1 - \theta\bar{t}) = 0$. En el primer caso se puede verificar que se obtiene un mínimo de la función de verosimilitud. Para $\theta = 1/\bar{t}$ se cumple que $(1 - \theta\bar{t}) = 0$ y se obtiene un máximo, por lo tanto $1/\bar{t}$ es un estimador de máxima verosimilitud.

La función de densidad en este ejemplo corresponde a una variable aleatoria con distribución exponencial, donde la tasa promedio de éxitos por unidad de tiempo $\lambda = \theta$, y t_i representa el tiempo entre la ocurrencia de dos éxitos consecutivos de un proceso de Poisson.

De manera semejante se puede demostrar que $\bar{X}$ es un estimador de máxima verosimilitud de la distribución normal y x/n lo es para la distribución binomial.

VII.2.2 ESTIMACION POR INTERVALOS DE CONFIANZA

En algunos casos no es suficiente estimar un parámetro de manera puntual, debido a que el estimador es una variable aleatoria y la probabilidad de que tome el mismo valor que el parámetro es prácticamente cero.

Para aumentar la probabilidad de que la estimación coincida con el parámetro deseado, se debe establecer un intervalo denominado intervalo de confianza que contenga, con cierta probabilidad, al parámetro deseado.

A los límites del intervalo de confianza se les denomina *límites de confianza* y a la probabilidad de que el parámetro de la población esté en el intervalo de confianza se le llama *nivel de confianza* de la estimación y, en lo sucesivo, se representará como $(1 - \alpha)$.

Para establecer los límites de confianza (LC) de un parámetro θ, se parte de una distribución muestral. En la siguiente figura se pueden observar los elementos que intervienen en la estimación muestral.

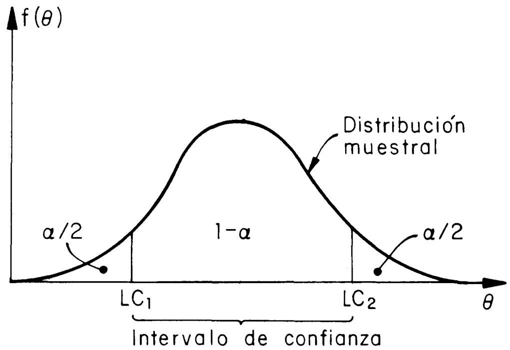


Figura VII.3

En ella se puede notar que el complemento del nivel de confianza α se distribuye en dos partes iguales a los lados del intervalo, con el objeto de centrarlo.

INTERVALO DE CONFIANZA PARA LA MEDIA

Como se vio al inicio del capítulo, el estadístico

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

es un buen estimador de la media poblacional μ_X, por este motivo se usará como base para hacer la estimación por intervalos. Si el tamaño de la muestra n es grande, la distribución de $\bar{X}$ tiende a la distribución normal con media

μ_X y desviación estándar $\sigma_X/\sqrt{n}$, en consecuencia la variable:

$$\boxed{Z = \frac{\overline{X} - \mu_X}{\sigma_X/\sqrt{n}}} \qquad \ldots (6)$$

tiende a la distribución normal estándar.

Para encontrar el intervalo de confianza, se deben fijar los límites LC_1 y LC_2 de tal manera que:

$$1 - \alpha = P(LC_1 \leq \mu_X \leq LC_2) \qquad \ldots (7)$$

de la expresión (6), esta probabilidad equivale a:

$$1 - \alpha = P(-z \leq \frac{\overline{X} - \mu_X}{\sigma_X/\sqrt{n}} \leq z) \qquad \ldots (8)$$

donde $-z$ y z son los límites de un intervalo que contiene al $100(1 - \alpha)$ porciento de los posibles valores de Z, como se muestra en la siguiente figura:

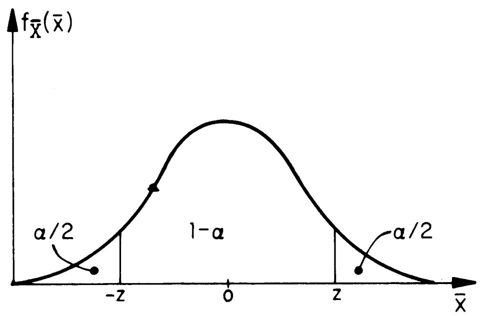


Figura VII.4

Tomando en cuenta que $\sigma_X/\sqrt{n}$ es siempre positivo se puede multiplicar la desigualdad de la expresión (8) por $\sigma_X/\sqrt{n}$, con lo cual:

$$1 - \alpha = P(z \frac{\sigma_X}{\sqrt{n}} \leq \overline{X} - \mu_X \leq z \frac{\sigma_X}{\sqrt{n}})$$

despejando μ_X:

$$1 - \alpha = P(\overline{X} - z \frac{\sigma_X}{\sqrt{n}} \leq \mu_X \leq x + z \frac{\sigma_X}{\sqrt{n}})$$

comparando esta expresión con (7), se tiene que:

$$LC_1 = \overline{X} - z \frac{\sigma_X}{\sqrt{n}}$$

$$LC_2 = \overline{X} + z \frac{\sigma_X}{\sqrt{n}}$$

En forma simplificada se puede representar el intervalo de μ_X como:

$$\boxed{\overline{X} \pm z \frac{\sigma_X}{\sqrt{n}}} \qquad \ldots (9)$$

En general no se conoce σ_X, sin embargo cuando la población es normal, se puede estimar puntualmente con la desviación estándar de la muestra. Como s_X es un estimador sesgado, es conveniente multiplicarlo por el factor $\sqrt{n}/\sqrt{n-1}$, con lo cual el intervalo de confianza de la media es:

$$\overline{X} \pm z \frac{s_X}{\sqrt{n-1}} \qquad \ldots (10)$$

Ejemplo VII.7

Un comerciante desea estimar el peso medio de los pavos que se crían en una granja, ya que piensa comprar un gran número de ellos.

a) ¿Cuál es el intervalo que contiene a la media de la población, con un nivel de confianza del 95%, si en una muestra representativa de 35 aves se encontró una media de 8 kg con una desviación estándar de 0.95 kg?

b) ¿De qué tamaño debe ser la muestra para que la media de la población pertenezca al intervalo $\overline{x} \pm 0.3$ kg, considerando el mismo nivel de confianza y que la desviación estándar es constante?

Solución

a) Como la población se puede considerar infinita, la muestra es grande y se desconoce σ_X, de la expresión (10):

$$\overline{X} - z \frac{s_X}{\sqrt{n-1}} \leq \mu_X \leq \overline{X} + z \frac{s_X}{\sqrt{n-1}}$$

Para buscar el valor de z, debe tomarse en cuenta que en cada extremo de la distribución normal, se encuentra el 2.5% del área, de manera que el intervalo central contenga al 95%, como se muestra en la figura siguiente:

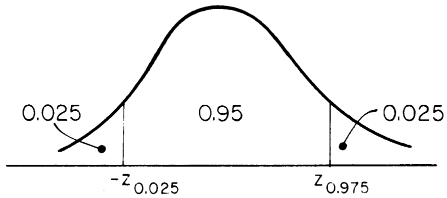


Figura VII.5

En la tabla de la distribución normal estándar se puede observar que para $z = 1.96$, $F_Z(1.96) = 0.975$, lo cual se representa en la figura VII.5 como $z_{0.975} = 1.96$.

Sustituyendo este valor en (10):

$$8 \pm \frac{1.96\ (0.95)}{\sqrt{35-1}}$$

haciendo operaciones:

$$\mu_X \in [7.681,\ 8.319]$$

Esto significa que con una probabilidad de 0.95, el peso medio de los pavos de la granja está entre 7.681 y 8.319 kg.

b) En este inciso se desea determinar el tamaño de la muestra, manteniendo constante el nivel de confianza y la desviación estándar.

El nuevo intervalo es ligeramente más pequeño, ya que:

$$\frac{1.96(0.95)}{\sqrt{35-1}} = 0.32$$

por lo cual, se puede esperar que la muestra sea mayor que 35.

Como el intervalo de confianza está centrado y la distribución muestral es simétrica se puede analizar un solo límite del intervalo.

Así, con el límite superior:

$$LC_2 = \overline{X} + z \frac{\sigma_X}{\sqrt{n}}$$

Comparando con el intervalo dado:

$$0.3 = z \frac{\sigma_X}{\sqrt{n}}$$

despejando n:

$$n = \left[\frac{z \ \sigma_X}{0.3}\right]^2$$

Buscando en la tabla de la distribución normal el valor de Z, tal que $F_Z(z) = 0.975$, se obtiene:

$$z = 1.96$$

Con lo cual:

$$n = \left[\frac{1.96(0.95)}{0.3}\right]^2 \doteq 39$$

Cuando la muestra es pequeña ($n < 30$) y la desviación estándar de la población es desconocida, no se debe usar la desviación estándar de la muestra como un estimador de σ_X, pues son muy diferentes.

En este caso, si la población es normal, la estimación por intervalos se puede apoyar en la variable aleatoria T, formada por la relación:

$$T = \frac{\dfrac{\overline{X} - \mu_X}{\sigma_X/\sqrt{n}}}{\left[\dfrac{\dfrac{n\ S_X^2}{\sigma_X^2}}{n-1}\right]^{1/2}} \qquad \ldots (11)$$

Como se demostró al inicio del capítulo que:

$$\frac{\overline{X} - \mu_X}{\sigma_X/\sqrt{n}}$$

tiene distribución normal estándar y:

$$\frac{n\, S_X^2}{\sigma_X^2}$$

tiene distribución ji-cuadrada con n-1 grados de libertad, entonces la variable aleatoria T tiene distribución t de student, pues está formada por el cociente de una variable con distribución normal estándar, entre la raíz cuadrada de otra variable con distribución ji-cuadrada que a su vez está dividida entre sus grados de libertad.

Simplificando la expresión (11):

$$\boxed{T = \frac{\overline{X} - \mu_X}{S_X/\sqrt{n-1}}} \qquad \ldots (12)$$

Como los grados de libertad de T dependen del tamaño de la muestra, cuando ésta tiende a infinito, la distribución de T tiende a la normal estándar. Por lo tanto, se puede concluir que cuando la muestra es grande, también se puede usar la distribución t de student para hacer la estimación por intervalos de la media poblacional.

El procedimiento para determinar los límites de confianza a partir de la variable aleatoria T es muy similar al seguido anteriormente. De esa manera, los límites de confianza para la media de la muestra, cuando se desconoce σ_X y la población es normal, son:

$$\boxed{\overline{X} \pm t\, \frac{S_X}{\sqrt{n-1}}} \qquad \ldots (13)$$

Ejemplo VII.8

Con el objeto de determinar el ruido medio que provoca una fresadora en cierta empresa, se han registrado en forma aleatoria las siguientes mediciones en decibeles: 63, 65, 60, 63, 64, 62, 68, 65, 64, 64. Considerando que el ruido tiene una distribución normal, ¿cuál es el intervalo que contiene a la media de la población con un nivel de confianza del 90%?

Solución

Como se tienen únicamente 10 datos y la población es normal, de la expresión (13):

$$\bar{X} - t\,\frac{S_X}{\sqrt{n-1}} \leq \mu_X \leq \bar{X} + t\,\frac{S_X}{\sqrt{n-1}}$$

donde:

$$\bar{X} = \frac{63 + 65 + \ldots + 64}{10} = 63.8$$

y:

$$S_X = \sqrt{\frac{(63-63.8)^2 + (65-63.8)^2 + \ldots + (64-63.8)^2}{10}} = 1.99$$

Buscando en la tabla C del apéndice la distribución t de student con $\nu = 10 - 1 = 9$ grados de libertad, el valor t que corresponde a $F_T(t) = 0.95$ es 1.83. Véase la siguiente figura:

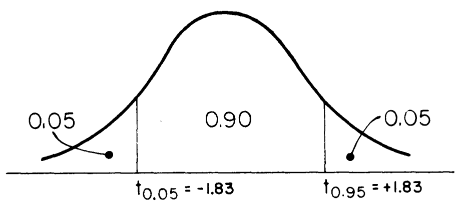


Figura VII.6

Con lo cual:

$$63.8 - \frac{1.83(1.99)}{\sqrt{10-1}} \leq \mu_X \leq 63.8 + \frac{1.83(1.99)}{\sqrt{10-1}}$$

$$62.486 \leq \mu_X \leq 65.114$$

Es decir, que el ruido medio que provoca la fresadora, se encuentra entre 62.486 y 65.114 decibles, con un nivel de confianza del 90%.

Cuando el muestreo se hace sin reemplazo y la población es finita se debe ajustar la desviación estándar de las expresiones (9), (10) y (13), multiplicando por el factor $\sqrt{\frac{N-n}{n-1}}$; donde N es el tamaño de la población y n es el tamaño de la muestra.

Como ya se demostró, cuando la población es normal, la variable aleatoria:

$$\chi^2 = \frac{n\ S_X^2}{\sigma_X^2}$$

donde S_X^2 es la variancia de una muestra, σ_X^2 la variancia de la población y n el tamaño de la muestra, tiene distribución ji-cuadrada con $n-1$ grados de libertad.

Utilizando esta variable como base de la estimación y a partir de la figura VII.7 se puede plantear la expresión:

$$1 - \alpha = P(\chi_1^2 \leq \frac{n\ S_X^2}{\sigma_X^2} \leq \chi_2^2)$$

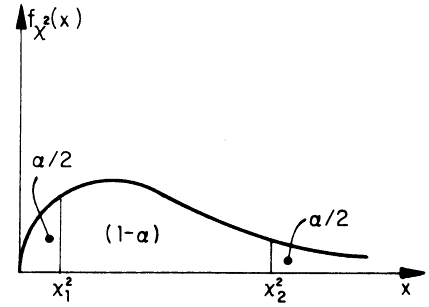


Figura VII.7

Como σ_X^2 siempre es positiva, se puede multiplicar la desigualdad por σ_X^2 sin que se altere, de esta manera:

$$1 - \alpha = P(\chi_1^2\ \sigma_X^2 \leq n\ S_X^2 \leq \chi_2^2\ \sigma_X^2)$$

lo cual equivale a:

$$1 - \alpha = P\left[(\chi_1^2\ \sigma_X^2 \leq n\ S_X^2) \cap (n\ S_X^2 \leq \chi_2^2\ \sigma_X^2)\right]$$

$$= P\left[(\sigma_X^2 \leq \frac{n\ S_X^2}{\chi_1^2}) \cap (\frac{n\ S_X^2}{\chi_2^2} \leq \sigma_X^2)\right]$$

con lo cual:

$$1 - \alpha = P\left[\frac{n\ S_X^2}{\chi_2^2} \leq \sigma_X^2 \leq \frac{n\ S_X^2}{\chi_1^2}\right]$$

Por lo tanto, el intervalo que contiene a la variancia de la población con un nivel de confianza $(1 - \alpha)$, es:

$$\boxed{\sigma_X^2 \in \left[\frac{n\ S_X^2}{\chi_2^2},\ \frac{n\ S_X^2}{\chi_1^2}\right]} \qquad \ldots (14)$$

Para establecer el intervalo de confianza de la desviación estándar, se utiliza la raíz cuadrada de los límites anteriores, esto es:

$$\boxed{\sigma_X \in \left[S_X\sqrt{\frac{n}{\chi_2^2}},\ S_X\sqrt{\frac{n}{\chi_2^2}}\right]} \qquad \ldots (15)$$

Ejemplo VII.9

Para estimar la desviación estándar que tienen las resistencias eléctricas producidas por cierta empresa, se han medido al azar 10 unidades, con los siguientes resultados:

173, 169, 170, 170, 173, 171, 167, 171, 168, 168

Suponiendo que la resistencia tiene una distribución aproximadamente normal, ¿cuál es el intervalo de confianza del 95% para la desviación estándar?

Solución

Calculando primero la media y la desviación estándar de la muestra:

$$\bar{x} = \frac{173 + 169 + \ldots + 168}{10} = 170$$

$$S_x = \sqrt{\frac{(173-170)^2 + (169-170)^2 + \ldots + (168-170)^2}{10}} = 1.95$$

con 10 - 1 = 9 grados de libertad y un nivel de confianza del 95%:

$$\chi^2_{0.025} = 2.7004$$

$$\chi^2_{0.975} = 19.0228$$

de la expresión (15):

$$1.95\sqrt{\frac{10}{19.022}} \leq \sigma_X \leq 1.95\sqrt{\frac{10}{2.7004}}$$

$$1.414 \leq \sigma_X \leq 3.752$$

Esto significa que con un nivel de confianza del 95%, la desviación estándar se encuentra entre 1.414 y 3.752.

INTERVALOS DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS

Frecuentemente se presenta la necesidad de comparar dos poblaciones semejantes e independientes, con el fin de seleccionar alguna de ellas. Por ejemplo, si existen dos proveedores de varilla en una empresa constructora, sería conveniente determinar cuál de los dos produce la varilla más resistente. Para ello, lo más adecuado sería seleccionar una muestra al azar de cada uno de los proveedores, probarlas y calcular la resistencia media de las dos muestras.

Al comparar estas medias muestrales se tendrá entonces una idea de cuál es la más resistente.

Para dos poblaciones (1 y 2) el estadístico $\overline{X}_2 - \overline{X}_1$ ó $\overline{X}_1 - \overline{X}_2$ se puede usar como un estimador puntual de la diferencia de las medias poblaciones $\mu_2 - \mu_1$ ó $\mu_1 - \mu_2$ respectivamente.

Como las medias muestrales $\overline{X}_1$ y $\overline{X}_2$ son variables aleatorias, su diferencia es también una variable aleatoria y para estimar la diferencia entre las medias poblacionales μ_1 y μ_2 en forma de intervalo de confianza, se debe conocer el comportamiento del estadístico $\overline{X}_2 - \overline{X}_1$ ó $\overline{X}_1 - \overline{X}_2$.

En el capítulo IV se encontró que la media y la variancia de una combinación lineal de dos variables aleatorias son:

$$E[aX + bY] = aE[X] + bE[Y] \qquad y$$

$$VAR[aX + bY] = a^2\,\sigma_X^2 + b^2\,\sigma_Y^2 - 2a\,b\,\sigma_{XY}$$

como en este caso $a = 1$, $b = -1$, $X = \overline{x}_2$, $Y = \overline{x}_1$ y $\sigma_{XY} = 0$, entonces:

$$E[\overline{X}_2 - \overline{X}_1] = E[\overline{X}_2] - E[\overline{X}_1]$$

$$VAR[\overline{X}_2 - \overline{X}_1] = \sigma^2_{\overline{x}_2} + (-1)^2\, \sigma^2_{\overline{x}_1}$$

además como:

$$E[\overline{X}_1] = \mu_1 \quad ; \quad E[\overline{X}_2] = \mu_2$$

$$\sigma^2_{\overline{X}_1} = \frac{\sigma_1^2}{n_1} \quad ; \quad \sigma^2_{\overline{X}_2} = \frac{\sigma_2^2}{n_2}$$

donde n_1 y n_2 representan el tamaño de la muestra correspondiente.

De esta manera:

$$E[\overline{X}_2 - \overline{X}_1] = \mu_2 - \mu_1 \qquad \ldots (16)$$

lo cual implica que $\overline{X}_2 - \overline{X}_1$ es un estimador insesgado de $\mu_2 - \mu_1$, y:

$$VAR[\overline{X}_2 - \overline{X}_1] = \frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1} \qquad \ldots (17)$$

En esta expresión puede observarse que a medida que aumenta el tamaño de la muestra, la variancia del estadístico tiende a cero, lo que indica que también es un estimador consistente de $\mu_2 - \mu_1$.

Como las distribuciones de $\overline{X}_1$ y $\overline{X}_2$ tienden a la distribución normal cuando n_1 y n_2 son suficientemente grandes (mayores que 30), se puede demostrar que la diferencia entre ellas también tiene distribución normal con media:

$$\mu_{\overline{X}_2 - \overline{X}_1} = \mu_2 - \mu_1$$

y desviación estándar:

$$\sigma_{\overline{X}_2 - \overline{X}_1} = \sqrt{\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}}$$

Entonces, se puede plantear la ecuación:

$$1 - \alpha = P\left[-z \leq \frac{(\overline{X}_2 - \overline{X}_1) - (\mu_2 - \mu_1)}{\sigma_{\overline{X}_2 - \overline{X}_1}} \leq z \right]$$

donde $\pm$ z son los límites de un intervalo que contiene al $(1 - \alpha)$ 100% de los posibles valores de la distribución normal estándar.

Despejando de la desigualdad $(\mu_2 - \mu_1)$:

$$1 - \alpha = P\left[\bar{X}_2 - \bar{X}_1 - z\ \sigma_{\bar{X}_2 - \bar{X}_1} \leq \mu_2 - \mu_1 \leq \bar{X}_2 - \bar{X}_1 + z\ \sigma_{\bar{X}_2 - \bar{X}_1}\right]$$

En forma simplificada se pueden expresar los límites del intervalo de confianza de la siguiente manera:

$$\boxed{\bar{X}_2 - \bar{X}_1 \pm z\sqrt{\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}}} \qquad \ldots (18)$$

Como generalmente no se conocen los valores de σ_1^2 y σ_2^2, se pueden estimar de manera puntual mediante las variancias de las muestras S_1^2 y S_2^2, siempre que las poblaciones sean normales. De esta forma los límites de confianza son:

$$\boxed{\bar{X}_2 - \bar{X}_1 \pm z\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \qquad \ldots (19)$$

Ejemplo VII.10

Para determinar en cuánto se ha incrementado la contaminación del aire en una zona industrial de la ciudad de México, se han tomado algunas observaciones durante el presente año y se piensa compararlas con la información de hace diez años. Los datos son los siguientes:

i	Muestra 1 Hace 10 años	Muestra 2 Actualmente
Número de observaciones	37	45
Media $\bar{X}_i$ (Partes por millón)	42	44
Variancia S_i^2	103	98

Con un nivel de confianza del 99%, ¿cuál es el intervalo que contiene a la diferencia entre los niveles promedio de contaminación $(\mu_2 - \mu_1)$?

Solución

$$\sigma^2_{\overline{X}_2 - \overline{X}_1} = \frac{103}{37} + \frac{98}{45} = 4.962$$

entonces:

$$\sigma_{\overline{X}_2 - \overline{X}_1} = \sqrt{4.962} = 2.227$$

de la tabla de la distribución normal estándar:

$$z = 2.58$$

con lo cual, de la expresión (19):

$$(44 - 42) \pm 2.58(2.227) = 2 \pm 5.746$$

es decir:

$$\mu_2 - \mu_1 \in [-3.746,\ 7.746]$$

lo cual indica que la diferencia puede estar entre -3.746 y 7.746 con un nivel de confianza del 99%. El signo negativo del límite inferior significa que la contaminación se pudo haber reducido, aun cuando la diferencia entre las medias muestrales indique lo contrario.

Cuando las muestras son pequeñas y las poblaciones normales, la estimación del intervalo de confianza para la diferencia de medias poblacionales, se puede basar en la variable aleatoria T formada como:

$$T = \frac{\dfrac{(\overline{X}_2 - \overline{X}_1) - (\mu_2 - \mu_1)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}}{\dfrac{\sqrt{\dfrac{n_1 S_1^2}{\sigma_1^2} + \dfrac{n_2 S_2^2}{\sigma_2^2}}}{\sqrt{n_1 + n_2 - 2}}} \quad \ldots (20)$$

donde la diferencia $\overline{X}_2 - \overline{X}_1$ tiene distribución normal con media $\mu_2 - \mu_1$ y desviación estándar $\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}$, y la combinación:

$$\frac{n_1 S_1^2}{\sigma_1^2} + \frac{n_2 S_2^2}{\sigma_2^2}$$

tiene distribución ji-cuadrada con $\nu = n_1 + n_2 - 2$ grados de libertad. Por todo lo anterior, la combinación de la expresión (20) tiene distribución t de student con $\nu = n_1 + n_2 - 2$ grados de libertad.

Bajo la consideración de que las variancias de las dos poblaciones son estadísticamente iguales (homogéneas), la expresión (20) se puede simplificar, con lo cuál:

$$\boxed{T = \frac{(\overline{X}_2 - \overline{X}_1) - (\mu_2 - \mu_1)}{\sqrt{n_1 S_1^2 + n_2 S_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}} \quad \ldots (21)$$

Mediante un procedimiento similar al seguido en la deducción de la expresión (18), se obtiene que los límites del intervalo de confianza de la diferencia de dos medias $\mu_1 - \mu_2$, cuando existe homogeneidad de variancias, es:

$$\boxed{\overline{X}_1 - \overline{X}_2 \pm t \sqrt{\frac{(n_1 S_1^2 + n_2 S_2^2)(n_1 + n_2)}{n_1 n_2 (n_1 + n_2 - 2)}}} \quad \ldots (22)$$

donde $\pm\, t$ son los límites de un intervalo central en la distribución t de student que contiene al $(1 - \alpha)$ 100% de los posibles valores de T, con $\nu = n_1 + n_2 - 2$ grados de libertad.

Ejemplo VII.11

Con el objeto de evaluar en qué medida se ha reducido la cantidad de desperdicio, que en promedio produce cierta máquina durante un día de trabajo, se han registrado las cantidades del desperdicio generado durante algunos días, antes y después de haberla ajustado. Los resultados fueron los siguientes:

	Desperdicio diario en kg
Antes del ajuste	85, 104, 79, 123, 108
Después del ajuste	81, 96, 103, 70, 75, 80, 107

Considerando que la cantidad de desperdicio tiene distribución normal y que su desviación estándar no cambió con el ajuste, ¿cuál es la estimación por intervalo de la disminución del desperdicio medio, debida al ajuste, con un nivel de confianza del 95%?

Solución

Calculando la media y la variancia de las dos muestras:

$$\bar{X}_1 = 99.8 \qquad S_1^2 = 254.96 \qquad n_1 = 5$$

$$\bar{X}_2 = 87.42 \qquad S_2^2 = 179.10 \qquad n_2 = 7$$

Como $n_1 = 5$ y $n_2 = 7$, de la tabla de la distribución t de student con $\nu = 5 + 7 - 2 = 10$ grados de libertad, se obtiene que:

$$t_{0.975} = 2.228$$

con lo cual, de la expresión (22):

$$99.8 - 87.42 \pm 2.228 \sqrt{\frac{[5(254.96) + 7(179.10)](5+7)}{5(7)(5+7-2)}}$$

haciendo operaciones, se tiene que:

$$-8.36 \leq \mu_1 - \mu_2 \leq 33.12$$

Es decir, que la reducción media está entre -8.36 y 33.12 kg. El signo negativo del límite inferior, implica que en lugar de reducción puede haber un incremento.

VII.3 PRUEBA DE HIPOTESIS ESTADISTICA

De acuerdo con el método científico, para aceptar un enunciado es necesario plantear hipótesis, experimentar y comprobarlas con los resultados del experimento.

Cuando se trata de un experimento aleatorio no es fácil corroborar la hipótesis, debido a que se presentan diferentes resultados en cada ensayo y, por este motivo, se hace necesario un procedimiento especial para los fenómenos aleatorios. Por ejemplo, si un técnico de la Compañía de Luz asegura que, en promedio, el voltaje que reciben los usuarios en cierta zona de la ciudad es de 117 volts, resulta prácticamente imposible verificar con certeza esta afirmación.

La *prueba de hipótesis estadística* es otra forma de inferencia que se utiliza para aceptar o rechazar hipótesis sobre un fenómeno aleatorio, planteadas en términos de un parámetro o una combinación de dos o más parámetros; también se usa para investigar si una distribución empírica se ajusta a un determinado modelo probabilístico.

Por la naturaleza aleatoria de los fenómenos, la prueba de hipótesis estadística no se puede considerar como una prueba definitiva, ya que en ella se establece un criterio o regla de decisión para aceptar o rechazar una hipótesis y, como se verá posteriormente, la regla de decisión está asociada con ciertos riesgos.

PROCEDIMIENTO GENERAL

En primer lugar, para realizar una prueba de hipótesis estadística se debe plantear el enunciado en cuestión, en términos de algún o algunos parámetros poblacionales como la media, la variancia, etcétera.

Generalmente se habla de dos tipos de hipótesis relacionadas con una prueba: *la hipótesis nula* (H_0) y *las hipótesis alternas* (H_1, H_2, ...).

La hipótesis nula es la que se toma como referencia en la prueba y, por conveniencia, se plantea como una igualdad de la forma $\theta = \theta_0$, donde θ_0 es un valor supuesto del parámetro θ.

Las hipótesis alternas son aquéllas que difieren de H_0 y pueden tener las siguientes formas:

$$\theta \neq \theta_0$$

$$\theta > \theta_0$$

$$\theta < \theta_0$$

Es importante señalar, que aceptar como verdadera una hipótesis alterna equivale a rechazar la hipótesis nula correspondiente y cuando se acepta H_0, se rechaza como verdadera H_1 (considerando que es la única alternativa). Lo más importante de una prueba estadística es el diseño de una *regla de decisión*, la cual señala en qué casos se debe aceptar o rechazar la hipótesis nula en función de lo que indica un estadístico, obtenido de una muestra representativa.

En el ejemplo del voltaje promedio (μ_X), el enunciado ya está expresado en términos de la media poblacional, por lo cual la hipótesis nula es:

$$H_0: \mu_X = 117 \text{ volts.}$$

La hipótesis alterna debe plantearse de acuerdo con los resultados que se desean obtener de la prueba, en este ejemplo se trata de encontrar evidencias que muestren que el voltaje promedio es diferente de 117, sin importar si es mayor o menor, por lo cual la hipótesis alterna que le corresponde es:

$$H_1: \mu_X \neq 117 \text{ volts.}$$

Cuando el interés está centrado en uno de los extremos de la variable aleatoria, como en la vida útil de una máquina o las utilidades de una empresa, en donde solamente importa un aumento, entonces la hipótesis alterna sería de la forma:

$$H_1: \theta > \theta_0$$

Por lo anterior, cuando la hipótesis alterna es de la forma $\theta < \theta_0$ ó $\theta > \theta_0$ a la prueba de hipótesis estadística se le denomina *unilateral o de una cola* y si es de la forma $\theta \neq \theta_0$, se le llama *bilateral o de dos colas*.

Para establecer la regla de decisión se toma como base una distribución muestral; en el ejemplo la distribución muestral de la media ($\bar{X}$) sería la más adecuada. Posteriormente se fija una región de aceptación y una de rechazo en el dominio del estadístico, es decir que si para una muestra, el estadístico toma un valor en la región de aceptación, entonces se considera que la hipótesis nula es verdadera; en caso contrario, se piensa que es falsa, con lo cual se ha establecido la *regla de decisión*.

Considerando el mismo ejemplo, supóngase en principio, que se fija arbitrariamente como región de aceptación el intervalo [116, 118] volts, y como región de rechazo, su complemento; esto significa que si en una muestra se obtiene una media $\bar{x}$ entre 116 y 118 volts, se aceptará que en promedio el voltaje de la población es de 117, en caso contrario se rechazará.

En la figura VII.8 se muestra la región de aceptación, de rechazo y la distribución muestral de $\bar{x}$.

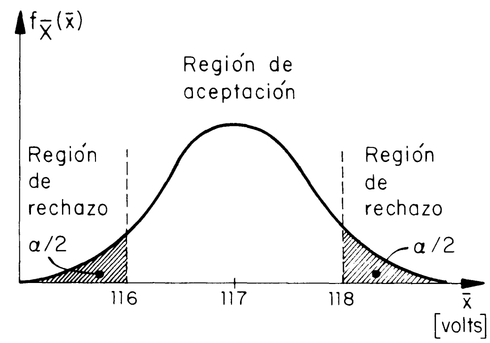


Figura VII.8

A partir de esta figura se puede establecer formalmente la regla de decisión como sigue:

Si $116 \leq \bar{x} \leq 118$, se acepta $H_0(\mu_X = 117)$; en caso contrario se rechaza. En el ejemplo se rechaza que $\mu = 117$ vots.

ERRORES EN LAS DECISIONES ESTADISTICAS

Como $\bar{X}$ es una variable aleatoria, al seguir la regla de decisión se corre el riesgo de rechazar la hipótesis nula, cuando $\bar{X}$ tome un valor en la región de rechazo y realmente corresponda a una población con media $\mu_X = 117$, esto significa que se está rechazando una hipótesis nula que es verdadera.

A este error se le conoce como de *tipo I* y a la probabilidad de cometerlo se le llama *nivel de significación* y se representa mediante α.

En la figura VII.8 se muestra el nivel de significación de la regla de decisión correspondiente. Nótese que está dividido en dos partes iguales. En este ejemplo, los límites del intervalo de aceptación (116 y 118), llamados puntos críticos, se establecieron de manera arbitraria. En general estos valores se calculan a partir de la distribución muestral del estadístico y de acuerdo con un nivel de significación determinado.

En la figura VII.8 se puede observar que si el nivel de significación α aumenta, la región de aceptación de H_0 se hace más pequeña, con lo cual la prueba es más estricta para aceptar la hipótesis nula; por esta razón, una prueba de hipótesis estadística es más significativa o más estricta para aceptar H_0, cuando el nivel de significación sea mayor.

Es común que se utilicen niveles de significación de 0.1, 0.05 ó 0.01; sin embargo, puede usarse cualquier otro valor, considerando que α es el riesgo de cometer el error tipo I.

Otro error que puede cometerse al decidir sobre la veracidad de una hipótesis nula, basándose en la regla de decisión, es el denominado *error tipo II*, que consiste en aceptar una hipótesis nula que es falsa.

En general no es posible determinar con precisión en qué casos se comete este tipo de error, debido a que no se llega a conocer el verdadero valor del parámetro poblacional. Lo único que se puede hacer es suponer algunos valores del parámetro y calcular la probabilidad o riesgo de cometer el error de tipo II. Esta probabilidad se representa mediante la letra β.

En el ejemplo del voltaje, si el verdadero valor de la media poblacional fuera de 118.5 en lugar de 117 como se consideró para establecer la regla de decisión, el área bajo la gráfica de la distribución muestral cuya media es $\mu = 118.5$ y que se encuentra en la región de aceptación, (véase la siguiente figura), es la probabilidad de cometer el error de tipo II, es decir β.

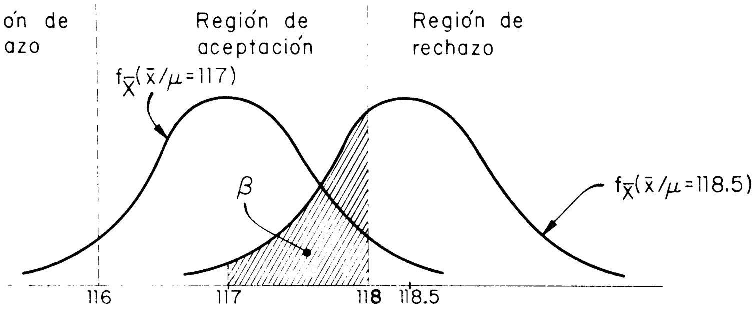


Figura VII.9

Para establecer una regla de decisión, lo más deseable es que la probabilidad de cometer los errores de tipo I y II sea mínima. No obstante, en la figura VII.9 se puede observar que si aumenta el intervalo de aceptación, se reduce α, pero aumenta β.

Asimismo, puede verse que si el valor del parámetro está cercano al valor supuesto en la hipótesis nula, aumentará la probabilidad β, aunque eso no se puede controlar.

La única manera efectiva de reducir los valores de α y β es aumentando el tamaño de la muestra, pues de esa manera se reduce la dispersión de la distribución muestral.

En resumen, para efectuar una prueba de hipótesis se debe seguir el siguiente procedimiento:

1. Establecer las hipótesis nula y alterna.
2. Seleccionar el nivel de significación que será utilizado.
3. Calcular los puntos críticos del estadístico de prueba.
4. Establecer la regla de decisión.
5. Determinar, de acuerdo con la regla, si se acepta o se rechaza la hipótesis nula.

Este procedimiento puede aplicarse en cualquier prueba de hipótesis y a pesar de que se ilustró en un caso bilateral, es muy semejante la aplicación en una prueba estadística unilateral.

Al inicio de este capítulo se estudió la distribución muestral del estadístico

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

que es la media de una muestra de n elementos.

Dependiendo del tipo de población, del tamaño de la muestra y de otras características, se vio cuál es la distribución muestral más apropiada para realizar una inferencia estadística sobre la media. Así por ejemplo, cuando la población es infinita y la muestra es grande, la variable aleatoria

$$z = \frac{\bar{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}}$$

tiene distribución normal estándar.

A partir de esta expresión, se pueden conocer los puntos críticos $\bar{x}_1$ y $\bar{x}_2$ de una prueba estadística bilateral mediante un simple despeje, con lo cual:

$$\bar{x}_1 = \mu_X + z\ \sigma_X/\sqrt{n}$$

$$\bar{x}_2 = \mu_X - z\ \sigma_X/\sqrt{n}$$

donde:

- μ_X es el valor del parámetro supuesto en la hipótesis nula,
- $\pm z$ son los valores de la variable z que delimitan un intervalo central con el $(1 - \alpha)$ 100% de los posibles valores. (Véase la gráfica (a) de la figura VII.10),
- σ_X es la desviación estándar de la población que puede estimarse mediante la desviación estándar de la muestra, con el sesgo corregido ($\hat{S}_X$) cuando no se conoce.

Si la prueba es unilateral, sólo se necesita un valor crítico del estadístico como se muestra en las gráficas (b) y (c) de la figura VII.10. En estos casos también se obtiene el valor crítico despejando $\bar{x}_c$ de la expresión (6), con la diferencia de que el punto crítico en z corresponde a una sola área en la región de rechazo (α) y en el caso bilateral corresponde a la mitad del área $\alpha/2$.

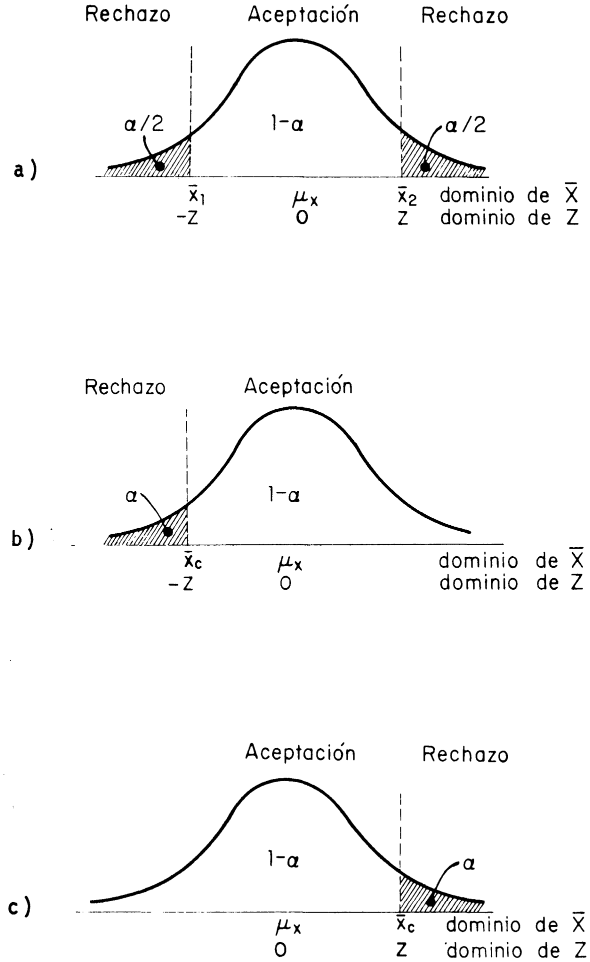


Figura VII.10

Si la población es normal y la muestra pequeña se usa la variable T en lugar de Z.

Algunos autores plantean la regla de decisión en términos de las variables Z o T en lugar del estadístico $\bar{X}$; sin embargo, como puede verse en las figuras VII.10 la regla sería equivalente a la que se obtiene cuando se refiere a $\bar{X}$.

Ejemplo VII.12

Al experimentar con la duración de una muestra de 50 cinescopios, se encontró una media de 1570 horas con una desviación estándar de 120 horas. Probar estadísticamente la hipótesis de que la media poblacional es $\mu = 1600$ horas, con un nivel de significación de 0.05, contra la hipótesis de que es menor, suponiendo que la duración de los cinescopios es normal.

Solución

Del enunciado:

$$H_0: \mu = 1600$$

$$H_1: \mu < 1600$$

Como se trata de una prueba estadística unilateral, considerando que la población es infinita y la muestra es grande, se usará la variable aleatoria

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}}$$

que tiene distribución normal estándar. Para encontrar el valor crítico $\bar{x}_c$, de acuerdo con la siguiente figura, el valor z_c que corresponde a $F_Z(z_c) = 0.05$, es $z_c = -1.64$.

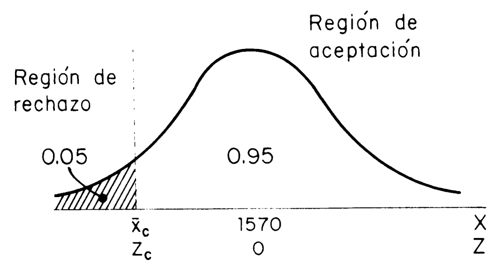


Figura VII.11

Como se desconoce σ_X, se puede estimar puntualmente mediante:

$$s_X \frac{\sqrt{n}}{\sqrt{n-1}} ,$$

de esta manera:

$$z_c = \frac{\bar{x}_c - \mu_X}{\frac{S_X \frac{\sqrt{n}}{\sqrt{n-1}}}{\sqrt{n}}} = \frac{\bar{x}_c - \mu_X}{S_X/\sqrt{n-1}}$$

despejando el punto crítico:

$$\bar{x}_c = \mu_X + z_c \frac{S_X}{\sqrt{n-1}}$$

sustituyendo valores:

$$\bar{x}_c = 1600 + (-1.64)\frac{(120)}{\sqrt{50-1}} \doteq 1572 \text{ horas}$$

con lo cual la regla de decisión es como sigue:

Si $\bar{x} < 1572$ horas se rechaza H_0; en caso contrario se acepta.

Como en la muestra $\bar{x} = 1570 < 1572$, se rechaza H_0 y se concluye que la duración media de los cinescopios ha bajado, con un nivel de significación de 5%.

Ejemplo VII.13

El tiempo promedio que tarda un obrero calificado en realizar cierta tarea es de 45 minutos de acuerdo con los registros de la empresa.

A partir de nuevos estudios sobre tiempos de trabajo, se desea comprobar estadísticamente si se ha modificado el tiempo promedio de esa actividad.

a) ¿Cuál debe ser la regla de decisión para realizar la prueba estadística, considerando que $\alpha = 0.05$, $n = 10$ y $S_X = 8$ minutos y suponiendo que la población es normal?

b) ¿Cuál es la probabilidad de cometer un error tipo II, si el verdadero tiempo medio fuera de 40 minutos?

Solución

a) Como no se sabe si el tiempo medio ha disminuido o aumentado, la prueba será bilateral, de manera que:

$$H_0: \mu_X = 45$$

$$H_1: \mu_X \neq 45$$

Como se trata de muestras pequeñas y la población es normal, se usará la variable:

$$T = \frac{\overline{X} - \mu_X}{S_X/\sqrt{n-1}} \qquad \ldots (23)$$

la cual tiene distribución t de student con $\nu = n-1$ grados de libertad, como se vio al inicio del capítulo.

Buscando en la tabla de la distribución t de student el valor de t_c que corresponde a $F_T(t_c) = 0.975$, con $\nu = 10-1 = 9$ se obtiene que $t_c = 2.26$ y por la simetría de la distribución $F_T(-2.26) = 0.025$.

Al evaluar la expresión (23) en los dos puntos críticos, se tiene:

$$\pm\, t_c = \frac{\overline{x}_c - \mu_X}{S_X/\sqrt{n-1}}$$

despejando $\overline{x}_c$:

$$\overline{x}_c = \mu_X \pm t_c \frac{S_X}{\sqrt{n}}$$

sustituyendo valores:

$$\overline{x}_c = 45 \pm \frac{(2.26)(8)}{\sqrt{10-1}} = 45 \pm 6.03$$

con lo cual la regla de decisión es como sigue:

Si $38.97 \leq \overline{x} \leq 51.03$ se acepta H_0, en caso contrario se rechaza.

b) Si el verdadero tiempo promedio fuera de 40 minutos, entonces la probabilidad de cometer el error tipo II (β) sería el área bajo la gráfica de la distribución muestral cuya media es 40, que se encuentra en la región de aceptación, como se muestra en la siguiente figura:

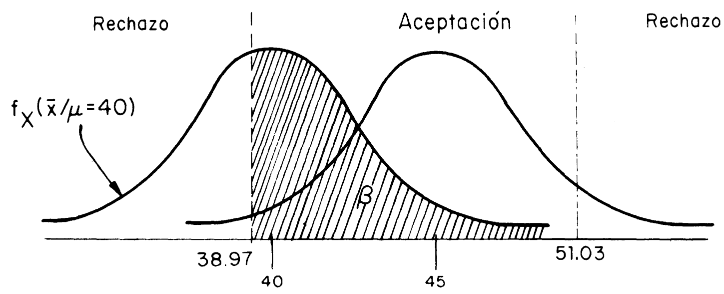


Figura VII.12

Para calcular la probabilidad β, se obtienen primero los valores de T que corresponden a 38.97 y 51.03, con $\mu = 40$, esto es:

$$t_1 = \frac{38.97 - 40}{8/\sqrt{10 - 1}} \qquad y \qquad t_2 = \frac{51.03 - 40}{8/\sqrt{10 - 1}}$$

$$= - 0.386 \qquad\qquad = 4.14$$

entonces:

$$\beta = P[(38.97 \leq \bar{x} \leq 51.03) | \mu = 40] = (-0.386 \leq T \leq 4.14)$$

como $P(T \geq 4.14)$ es prácticamente cero, como se puede verificar en la tabla:

$$\beta \doteq P(T \geq - 0.386)$$

además, por la simetría de la función:

$$\beta = P(T \geq - 0.386) = P(T \leq 0.386)$$

Como en la tabla de la distribución t de student con 9 grados de libertad, sólo aparecen $F_T(0.261) = 0.6$ y $F_T(0.543) = 0.7$, mediante una interpolación lineal:

$$\beta = F_T(0.386) = 0.644$$

PRUEBA DE HIPOTESIS SOBRE LA VARIANCIA Y LA DESVIACION ESTANDAR

Como en la estimación por intervalos, cuando la población es normal, la prueba de hipótesis sobre la variancia o la desviación estándar se apoya en la variable aleatoria:

$$\chi^2 = \frac{n\ S_X^2}{\sigma_X^2} \qquad \ldots (24)$$

que tiene distribución ji-cuadrada con n-1 grados de libertad.

Ejemplo VII.14

La máquina empacadora de cierto producto tenía algunas fallas que provocaban variaciones en el llenado de los paquetes. Para disminuirlas se ajustó la máquina y se desea probar estadísticamente si la desviación estándar del peso del producto depositado, que era de 25 gramos, se ha reducido en forma significativa. Para ello se tomó una muestra de 20 paquetes y se obtuvo una desviación estándar de 18 gramos.

a) Con un nivel de significación de 0.01, ¿cuál sería la decisión estadística?

b) ¿Con qué nivel de significación cambiaría la conclusión?

Solución

Como el interés está centrado en la disminución de la desviación estándar, las hipótesis estadísticas son:

$$H_0: \sigma_X = 25$$
$$H_1: \sigma_X < 25$$

Nótese que en este ejemplo se desea aceptar la hipótesis alterna, lo cual implica rechazar la hipótesis nula.

Buscando en la tabla de la distribución ji-cuadrada el valor de x_c, tal que $F_{\chi^2}(x_c) = 0.01$, con $\nu = 20 - 1 = 19$ grados de libertad (véase figura VII.13), se obtiene que $x_c = 7.63$.

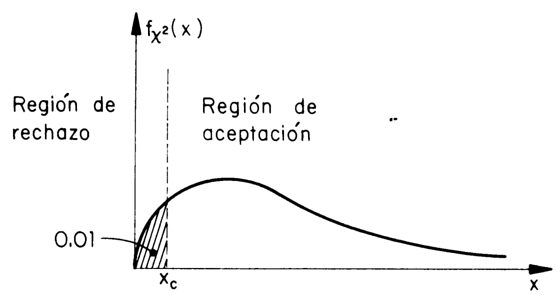


Figura VII.13

De la expresión (24):

$$s_X \text{ (crítico)} = \sqrt{\frac{x_c \, \sigma_X^2}{n}}$$

$$= \sqrt{\frac{(7.63)(25)^2}{20}}$$

$$= 15.44$$

de manera que la regla de decisión es como sigue:

Si $s_X \geq 15.44$ se acepta H_0, en caso contrario se rechaza.

Como en este caso:

$$s_X = 18 > 15.44$$

entonces se acepta H_0 con $\alpha = 0.01$, lo cual significa que con el 1% de nivel de significancia no se acepta que hubo reducción en la desviación estándar.

b) La conclusión cambia si s_X (crítico) ≥ 18, con lo cual el valor crítico de la variable χ^2 es:

$$\chi_c^2 \leq \frac{(20)(18)^2}{(25)^2} = 10.37$$

buscando en la tabla, con $\nu = 19$:

$$\alpha = F_{\chi^2}(10.37) \doteq 0.05$$

por lo tanto, la decisión cambiará si $\alpha \geq 0.05$

PRUEBA DE HIPOTESIS SOBRE DIFERENCIA DE MEDIAS

Este tipo de prueba es de gran utilidad en la comparación de poblaciones, pues consiste en investigar si con la información obtenida en una muestra de cada población, se puede inferir que las medias poblacionales son diferentes; es decir, si existe superioridad por parte de alguna de ellas.

Dependiendo de las características de las poblaciones y de las muestras se pueden usar las variables aleatorias Z o T, definidas de acuerdo con las expresiones (18), (19) y (21), como base para realizar la prueba de hipótesis estadística.

Ejemplo VII.15

Con el objeto de comparar dos estrategias de venta (1 y 2) para un producto, se han registrado las ventas diarias en dos almacenes semejantes, cada uno con diferente estrategia.

El almacén que utilizó la estrategia 1 durante 30 días vendió en promedio \$ 8,000 diarios, con una desviación estándar de \$ 1,000. El otro almacén vendió en promedio \$ 7,500 diarios, durante 35 días de observación, con una desviación estándar de \$ 950.

a) ¿Se puede pensar que las ventas promedio son estadísticamente iguales en los dos almacenes, con un nivel de significación de 0.05?

b) ¿Cuál es el máximo nivel de significación con el que se puede aceptar, de acuerdo con las muestras que $\mu_1 - \mu_2 = \$\ 800$, contra la hipótesis de que $\mu_1 - \mu_2 < \$\ 800$?

Solución

a) Considerando que la diferencia puede ser a favor de cualquiera de las dos estrategias, las hipótesis estadísticas son:

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0$$

Como las muestras son suficientemente grandes, se usará la variable aleatoria:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

que tiene distribución normal estándar y donde σ_1^2 y σ_2^2, pueden estimarse mediante las variancias de las muestras, considerando que las poblaciones son normales.

De esta manera, los valores críticos de la diferencia $(\overline{X}_1 - \overline{X}_2)$ son:

$$(\overline{X}_1 - \overline{X}_2)_c = (\mu_1 - \mu_2) \pm z \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Los valores de z que corresponden a $F_Z(0.025)$ y $F_Z(0.975)$ son, de acuerdo con la tabla de la distribución normal estándar, $\pm$ 1.96, con lo cual:

$$(\bar{x}_1 - \bar{x}_2)_c = 0 \pm 1.96 \sqrt{\frac{(1000)^2}{30} + \frac{(950)^2}{35}}$$

$$= \pm 476.56$$

entonces la regla de decisión es como sigue:

Si $-476.56 \leq (\bar{x}_1 - \bar{x}_2) \leq 476.56$ se acepta H_0, en caso contrario se rechaza.

Como $(\bar{x}_1 - \bar{x}_2) = (8000 - 7500) > 476.56$, se rechaza H_0, lo cual indica que con un nivel de significación de 0.05 se puede pensar que sí existe diferencia significativa entre las dos estrategias.

b) En este caso las hipótesis son:

H_0: $\mu_1 - \mu_2 = 800$

H_1: $\mu_1 - \mu_2 < 800$

y la regla de decisión tendrá la forma siguiente:

Si $(\bar{x}_1 - \bar{x}_2) \geq (\bar{x}_1 - \bar{x}_2)_c$ se acepta H_0, en caso contrario se rechaza.

Con los datos de las muestras, para que se acepte H_0 $(\bar{x}_1 - \bar{x}_2)_c \leq 500$, entonces:

$$z_c = \frac{500 - 800}{\sqrt{\frac{(1000)^2}{30} + \frac{(950)^2}{35}}} = -1.23$$

Buscando en la tabla de la distribución normal estándar el valor $F_Z(-1.23) = \alpha$, como se muestra en la figura VII.14, se obtiene que $\alpha = 0.1093$.

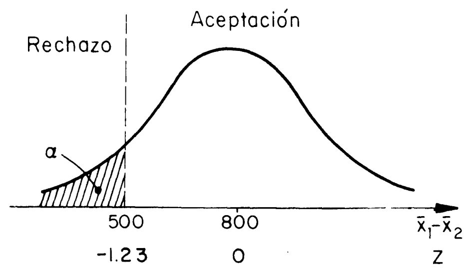


Figura VII.14

Por lo tanto, el nivel de significación debe ser cuando mucho de 10.93% para aceptar la hipótesis de que $\mu_1 - \mu_2 = 800$, cuando la diferencia entre las medias muestrales es $\bar{x}_1 - \bar{x}_2 = 500$.

CAPITULO VIII REGRESION Y CORRELACION

INTRODUCCION

En este capítulo se presentan dos métodos para estimar el valor esperado de una variable aleatoria dado que otra variable, que puede o no ser aleatoria, toma un valor determinado. Por ejemplo: se puede estimar la cantidad esperada de maíz que será cosechada en una parcela, en función del volumen de abono o agua que se le aplica, incluso se podría generalizar el método para estimar la cantidad esperada de maíz en función de dos o más variables.

También se definen algunos estimadores puntuales para la relación entre dos variables aleatorias, así como una estimación por intervalos de confianza para una de las variables aleatorias, dado que la otra toma un valor determinado.

VIII.1 REGRESION

En general una *curva de regresión* es una función que relaciona cada valor particular x de una variable independiente X, con el correspondiente valor esperado de la variable dependiente Y.

Esta función se encuentra obteniendo la esperanza de Y con la condición de que $X = x$, es decir, utilizando la distribución condicional $f_{Y/X}(y, x)$ si las variables son continuas, o $p_{Y/X}(y, x)$ si las variables son discretas, de esta forma:

$$\boxed{\mu_{Y/X} = E[Y/X = x] = \int_{-\infty}^{\infty} y \; f_{Y/X}(y, x) \; dy} \quad \ldots (1)$$

o bien:

$$\boxed{\mu_{Y/X} = E[Y/X = x] = \sum_{\forall y} y \; p_{Y/X}(y, x)}$$

Si en estas expresiones se considera como variable independiente a X, se tiene la curva de regresión de Y dado x, en caso contrario se denomina curva de regresión de X dado y.

A las variables aleatorias se les llama dependientes o independientes por el papel que tienen en la función, y no porque una dependa en el sentido probabilístico de la otra.

Cuando las variables aleatorias son independientes en el sentido probabilístico, la distribución condicional de Y dado x es igual a la distribución marginal de Y, con lo cual, en el caso continuo:

$$\mu_{Y/X} = \int_{-\infty}^{\infty} y \; f_{Y/X}(y, x) \; dy$$

$$= \int_{-\infty}^{\infty} y \; f_Y(y) \; dy$$

$$= \mu_Y$$

Por lo tanto, se puede observar que si dos variables aleatorias X, Y son independientes, la curva de regresión de Y dado x es una línea recta de pendiente cero, en la que todos los puntos tienen la misma ordenada μ_Y.

Para obtener la curva de regresión de acuerdo con las expresiones (1) es necesario establecer la distribución de probabilidad conjunta, la cual es a menudo difícil de conocer.

Existen algunos métodos para estimar la curva de regresión a partir de una muestra representativa de la población. Entre ellos se encuentran el método gráfico y el método de los mínimos cuadrados, los cuales se presentan a continuación.

METODO GRAFICO

Se basa en un diagrama denominado *de dispersión*, en el cual se encuentran graficados los n elementos o puntos (x_i, y_i) con $i = 1,2, \ldots, n$, obtenidos de una muestra representativa.

En la figura VIII.1 se encuentran cuatro diagramas de dispersión. En (a) y (b) se puede considerar que la curva de regresión de Y dado x es una línea recta, en el primer caso con pendiente positiva y en el segundo con pendiente negativa.

En el diagrama (c) puede suponerse una parábola como curva de regresión y, por último, en (d) se puede considerar una recta horizontal, pues no se percibe ninguna relación entre las variables y para cualquier valor particular x, la esperanza de Y es constante e igual a $\bar{y}$.

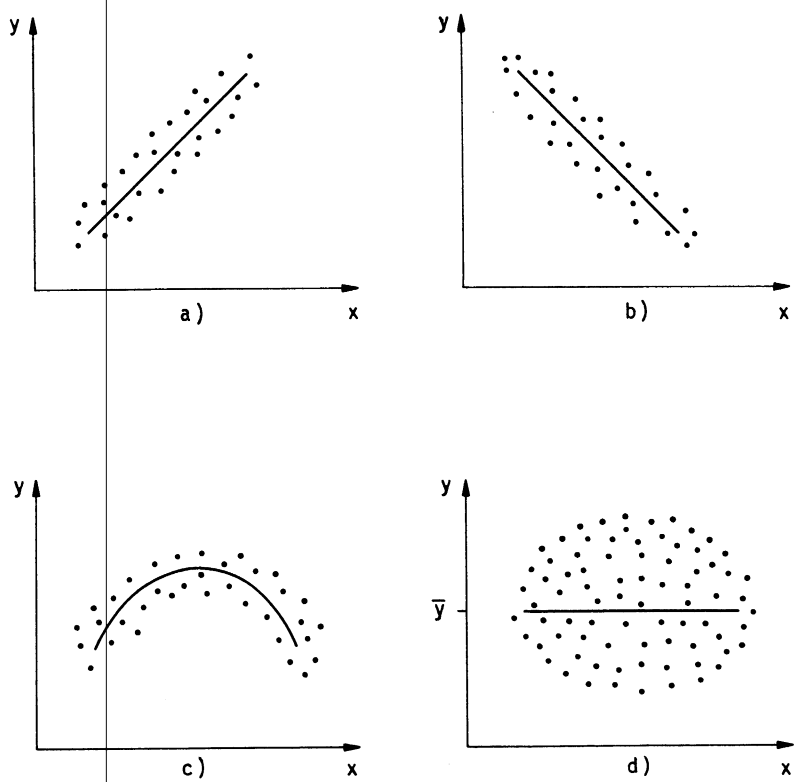


Figura VIII.1

Este método para estimar la curva de regresión es subjetivo y, por tal razón, cambia de una persona a otra; sin embargo, puede usarse para determinar qué tipo de curva (recta, exponencial, logarítmica, etc.) es la más apropiada para estimar la curva de regresión.

METODO DE LOS MINIMOS CUADRADOS

Consiste en estimar la curva de regresión de Y dado x mediante una función $\hat{y} = f(x)$, de tal manera que para una muestra de tamaño n se minimice la función D^2, tal que:

$$D^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad \ldots (2)$$

donde y_i son los valores muestrales de Y.

En la figura VIII.2 se puede observar que las diferencias $y_i - \hat{y}_i$ son las distancias entre el punto (x_i, y_i) y la ordenada de la curva de regresión que corresponde a x_i.

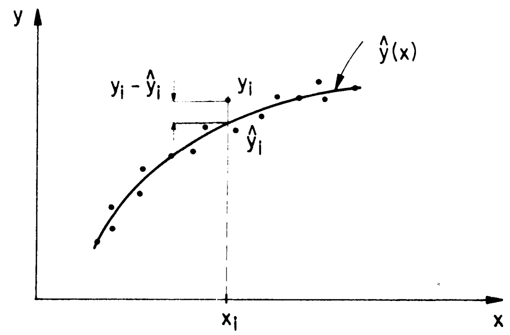


Figura VIII.2

El método de mínimos cuadrados se usa para aproximar funciones algebraicas a la curva de regresión, sin embargo, en estos apuntes se tratará únicamente el ajuste de una línea recta de la forma:

$$\boxed{\hat{y}_{(x)} = a + b\,x} \qquad \ldots (3)$$

que es la más utilizada, pues se puede demostrar que si X y Y tienen distribución normal o aproximadamente normal, la curva de regresión de Y dado x es una línea recta de la forma:

$$\mu_{Y/x} = \alpha + \beta\,x \qquad \ldots (4)$$

donde α y β son constantes. También se puede demostrar que los coeficientes a y b, obtenidos con el método de mínimos cuadrados, son buenos estimadores de α y β.

La única limitación del método de mínimos cuadrados consiste en que sólo puede aplicarse cuando la dispersión en la variable dependiente Y, con respecto a la media condicional $\mu_{Y/x}$, permanece constante o aproximadamente constante para cualquier valor de X.

En muchos casos no se cumple esta premisa y a medida que aumenta el valor de X, aumenta también la dispersión de la variable Y. Por ejemplo, si Y representa el peso de un individuo y X la estatura, es razonable que la variación en el peso de las personas que miden 1.80 m sea mayor que en aquéllas que miden 1.40 m.

Para encontrar los coeficientes a y b que minimizan la expresión (2), se valúa (3) en $x = x_i$ y se sustituye en (2), con lo cual:

$$D^2 = \sum_{i=1}^{n} \left[y_i - (a + b\,x_i)\right]^2$$

derivando parcialmente con respecto a los parámetros a y b:

$$\frac{\partial D^2}{\partial a} = -2 \sum_{i=1}^{n} (y_i - a - b\,x_i)$$

$$\frac{\partial D^2}{\partial b} = -2 \sum_{i=1}^{n} x_i(y_i - a - b\,x_i)$$

igualando a cero y simplificando:

$$\sum_{i=1}^{n} y_i - na - b \sum_{i=1}^{n} x_i = 0$$

$$\sum_{i=1}^{n} x_i\,y_i - a \sum_{i=1}^{n} x_i - b \sum_{i=1}^{n} x_i^2 = 0$$

Como todas las sumatorias son de 1 a n, se pueden omitir los subíndices, con lo cual:

$$na + b\,\Sigma\,x = \Sigma\,y$$

$$a\,\Sigma\,x + b\,\Sigma\,x^2 = \Sigma\,x\,y$$

que es un sistema de dos ecuaciones con dos incógnitas a y b.

Resolviendo el sistema, se tiene que:

$$\boxed{\begin{aligned} a &= \frac{\Sigma\, y\, \Sigma\, x^2 - \Sigma\, x\, \Sigma\, x\, y}{n\, \Sigma\, x^2 - (\Sigma\, x)^2} \\ b &= \frac{n\, \Sigma x\, y - \Sigma\, x \Sigma\, y}{n\, \Sigma x^2 - (\Sigma\, x)^2} \end{aligned}} \qquad \ldots (5)$$

Es importante notar que el método de mínimos cuadrados se basa en las distancias verticales entre los puntos de la muestra y las ordenadas correspondientes de la recta y no toma en cuenta las distancias horizontales. Por esta razón, la recta de regresión de Y dado x es, en general, diferente a la recta de X dado y, que se obtiene al considerar a X como variable dependiente y a Y como independiente.

Ejemplo VIII.1

El número de ingenieros egresados de las escuelas de educación superior y el número de ingenieros recibidos entre 1975 y 1981, fue como sigue:

AÑO	1975	1976	1977	1978	1979	1980	1981
Egresados (miles)	8.81	9.29	9.77	10.25	10.74	11.30	11.78
Titulados (miles)	3.78	3.99	4.20	4.41	4.62	4.87	5.01

Mediante el método de mínimos cuadrados:

a) ¿Cuál es la estimación de la recta de regresión del número de ingenieros titulados, dado el número de egresados?

b) Si en un año egresan 20,000 estudiantes, ¿cuántos de ellos pueden esperarse que obtengan el título?

Solución

a) Como se busca el número de ingenieros titulados, en función de los egresados, la variable dependiente Y es el número de titulados y la independiente X es el número de egresados.

Para calcular Σx, Σy, Σx^2 así como Σxy, es conveniente construir una tabla como la siguiente:

	X	Y	X^2	XY
	8.81	3.78	77.616	33.302
	9.24	3.99	85.378	36.868
	9.77	4.20	95.453	41.034
	10.25	4.41	105.063	45.203
	10.74	4.62	115.348	49.619
	11.30	4.87	127.690	55.03
	11.78	5.01	138.768	59.018
Σ	71.89	30.88	745.316	320.074

Sustituyendo en las expresiones (5):

$$a = \frac{30.88(745.316) - 71.89(320.074)}{7(745.316) - (71.89)^2}$$

$$= 0.1068$$

$$b = \frac{7(320.074) - (71.89)(30.88)}{7(745.316) - (71.89)^2}$$

$$= 0.4191$$

con lo cual la estimación de la recta de regresión es, de acuerdo con la expresión (3):

$$\hat{y}(x) = 0.1068 + 0.4191\ x$$

En la siguiente figura se puede observar cómo se ajusta la recta a los puntos de la muestra.

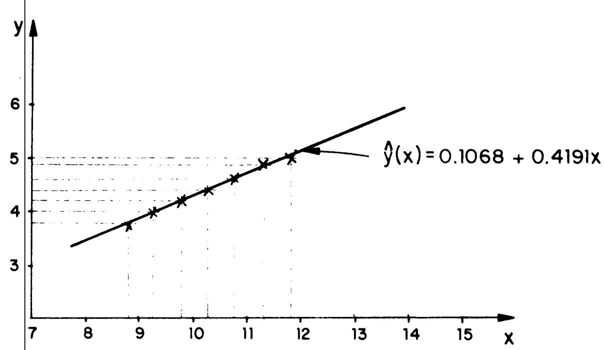


Figura VIII.3

b) Para $x = 20$

$$\hat{y}(20) = 0.1068 + 0.4191\,(20) \doteq 8.489$$

Esto significa que si egresan 20,000 estudiantes, el número esperado de titulados es 8,489.

VIII.2 CORRELACION MUESTRAL

En probabilidad y estadística el concepto de *correlación* indica la relación que existe entre dos o más variables aleatorias. Generalmente se habla de correlación cuando se desea hacer predicciones de una variable aleatoria, dado que otra toma un valor determinado.

Si las variables aleatorias no son independientes y se puede predecir con certeza el valor que tomará una de ellas, mediante una función del valor que tomará la otra, se dice que la correlación es perfecta.

La función referida es la curva de regresión, pues ella relaciona cada valor de la variable independiente X con el valor esperado de Y; y si la correlación es perfecta, para cada x existe un solo valor de Y que es $\mu_{Y/X}$, de manera que todos los posibles puntos (x, y) se encuentran en la curva de regresión de Y dado x.

En la mayoría de los casos, no todos los puntos (x, y) del espacio muestral se encuentran en la curva de regresión y, entre mayor sea la dispersión, con respecto a la curva de regresión, se dice que existe menos correlación entre las variables. Ahora bien, si las variables son independientes, no existe correlación entre ellas.

En la gráfica (a) de la figura VIII.4 se muestra un diagrama de dispersión obtenido de una muestra que corresponde a dos variables con una correlación perfecta; en la (b) se puede considerar que existe correlación, aunque ésta no es perfecta; finalmente, en la gráfica (c) se muestra un diagrama de dispersión que corresponde a dos variables independientes y, por consiguiente, no existe correlación.

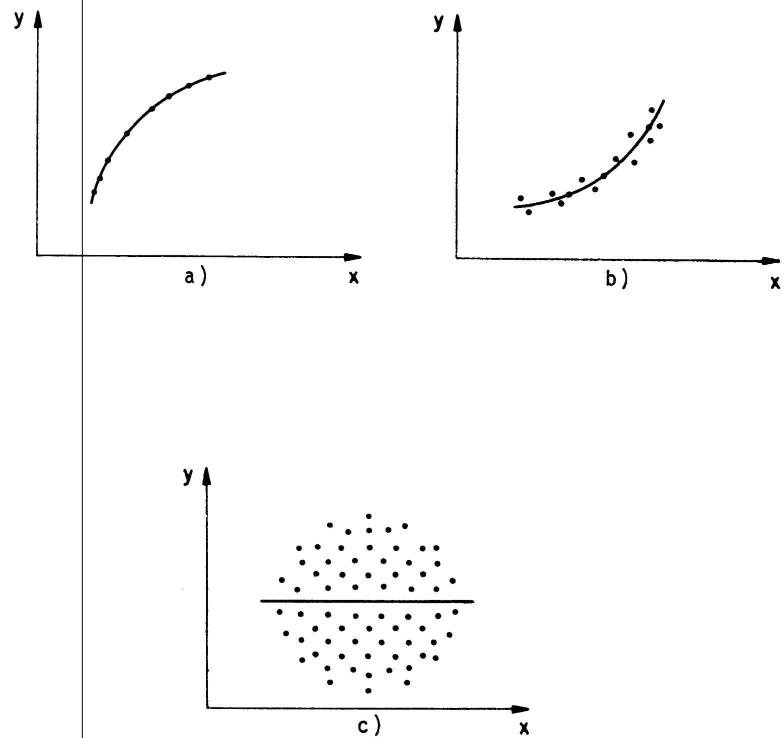


Figura VIII.4

En el capítulo IV se definieron como medidas de correlación la covariancia y el coeficiente de correlación, los cuales solamente son válidos cuando la curva de regresión es una línea recta.

Debido a que la correlación está directamente relacionada con la dispersión de los datos con respecto a la curva de regresión, se define la desviación estándar condicional de Y dado x como:

$$\sigma_{Y/x} = \sqrt{E\left[(y(x) - \mu_{Y/x})^2\right]} \qquad \ldots (6)$$

Esta medida de dispersión puede dar una idea de la correlación entre las variables, pues a mayor correlación le corresponde una menor desviación estándar condicional y viceversa.

En el ejemplo VIII.1 se hizo la predicción del número de egresados que obtendrán el título de ingeniero en función de los que terminaron sus estudios. Esta predicción será más precisa si existe una buena correlación entre las dos variables. De manera subjetiva se puede observar en la figura VIII.3 que la correlación es buena puesto que los puntos de la muestra se encuentran muy cercanos al valor correspondiente en la recta de regresión estimada.

Como en la mayoría de los casos únicamente se conoce una muestra de la población, se deben estimar las medidas de correlación de la población por medio de estadísticos muestrales.

Un buen estimador puntual de la desviación estándar condicional de Y dado x (llamado error estándar) es:

$$\boxed{s_{Y/x} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}} \qquad \ldots (7)$$

donde $\hat{y}_i$ es la estimación muestral de la ordenada de la curva de regresión en $x = x_i$, $i = 1, 2, 3, \ldots, n$.

Cuando las variables tienen una relación lineal, y_i se obtiene mediante la recta de mínimos cuadrados y, para simplificar el cálculo del error estándar de Y dado x, se puede sustituir $a + bx_i$ en lugar de y_i, con lo cual:

$$s_{Y/x} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - a - bx_i)^2}$$

desarrollando el cuadrado y simplificando se obtiene:

$$\boxed{s_{Y/x} = \sqrt{\frac{\Sigma y^2 - a\,\Sigma y - b\,\Sigma x y}{n}}} \qquad \ldots (8)$$

A pesar de que esta expresión parece más complicada que (7), resulta más práctica, pues no requiere de todos los valores $\hat{y}_i$ y se basa en varios de los elementos ya calculados para la estimación de la recta de regresión. Sin embargo, no debe olvidarse que se ha limitado a variables cuya regresión se supone lineal.

La covariancia de una muestra de dos variables aleatorias conjuntas se define como:

$$s_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \qquad \ldots (9)$$

y es también un buen estimador de la covariancia σ_{XY}.

El coeficiente de correlación de una muestra se define, de la misma forma que en el capítulo IV, como el cociente de la covariancia entre el producto de las desviaciones estándar de cada variable, pero en este caso muestrales, de manera que:

$$r = \frac{S_{XY}}{S_X\, S_Y} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}\ \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

simplificando la expresión y omitiendo los subíndices:

$$\boxed{r = \frac{n\,\Sigma xy - \Sigma x\ \Sigma y}{\sqrt{\left[n\ \Sigma x^2 - (\Sigma x)^2\right]\left[n\ \Sigma y^2 - (\Sigma y)^2\right]}}} \quad \ldots (10)$$

El coeficiente de correlación es la medida de correlación lineal más utilizada pues, a diferencia de la covariancia y el error estándar, es adimensional y, como se vio en el capítulo IV, varía entre menos uno y uno, indicando mayor correlación lineal cuando, en valor absoluto, se aproxima a la unidad.

Ejemplo VIII.2

A partir de los datos del ejemplo VIII.1, estimar:

a) El error estándar de Y dado x.

b) El coeficiente de correlación.

Solución

a) Para aplicar la expresión (8), es necesario calcular primero:

$$\sum_{i=1}^{n} y_i^2 = (3.78)^2 + (3.99)^2 + \ldots + (5.01)^2 = 137.46$$

con lo cual, de (8):

$$S_{Y/x} = \sqrt{\frac{137.46 - 0.1068(30.88) - 0.4191(320.074)}{7}}$$

$$= 0.052 \text{ (miles de alumnos)}$$

Esto puede considerarse como un promedio (aproximado) de las desviaciones entre la recta $\hat{y}(x_i)$ y las ordenadas de los puntos muestrales y_i.

b) De la expresión (10):

$$r = \frac{7(320.074) - (71.89)(30.88)}{\sqrt{\left[7(745.316) - (71.89)^2\right]\left[7(137.46) - (30.88)^2\right]}} = 0.998$$

Este resultado confirma que existe una buena correlación lineal entre las dos variables, como se había observado en la figura VIII.3.

VIII.3 INTERVALOS DE CONFIANZA PARA $\mu_{Y/x}$

La recta de mínimos cuadrados proporciona una estimación puntual $\hat{y}(x)$ de la esperanza condicional $\mu_{Y/x}$; sin embargo con este tipo de estimación no se conocen los riesgos inherentes a la misma, para ello es más adecuado estimar $\mu_{Y/x}$ mediante un intervalo de confianza.

Los límites del intervalo de confianza se obtienen, como en los casos vistos en el capítulo anterior, a partir de la distribución muestral de un estadístico o de una función del estadístico.

En este caso, si las variables aleatorias X, Y tienen distribución normal, se puede demostrar que para cualquier x el estadístico $\hat{y}(x)$ también tiene distribución normal con media $\mu_{Y/x}$.

Por otra parte, considerando que la variancia condicional de Y dado x es la misma para cualquier valor de x, entonces la variancia condicional del estadístico $\hat{y}(x)$ también es la misma para cualquier valor de x, como se muestra en la siguiente figura:

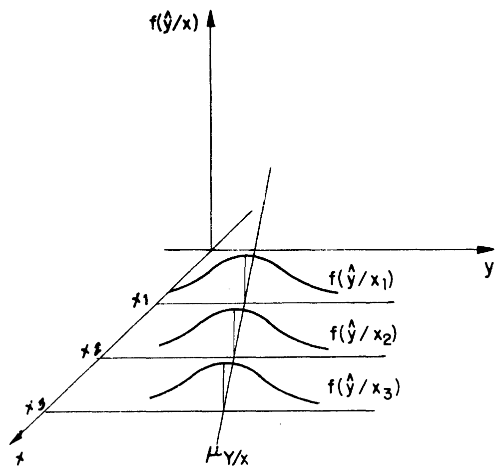


Figura VIII.5

Si se conociera la variancia de $\hat{y}$ para cada x, se podría partir de la variable:

$$Z = \frac{\hat{y} - \mu_{\hat{y}/x}}{\sigma_{\hat{y}/x}}$$

para obtener el intervalo de confianza de $\mu_{Y/x}$. Sin embargo, para tener una buena estimación de $\sigma_{\hat{y}/x}$ se requieren muestras enormes y es preferible tomar como base a la variable:

$$T = \frac{\hat{y} - \mu_{Y/x}}{\frac{S_{Y/x}}{\sqrt{n-2}} \sqrt{1 + \frac{(x - \bar{x})^2}{S_x^2}}} \quad ; \quad n > 2$$

que tiene distribución t de student con $\nu = n - 2$ grados de libertad, la cual puede usarse con muestras grandes o pequeñas. Siguiendo el mismo procedimiento que en los intervalos de confianza para la media, se obtiene que los límites de confianza para $\mu_{Y/x}$ son:

$$\boxed{\hat{y} \pm t \frac{S_{Y/x}}{\sqrt{n-2}} \sqrt{1 + \frac{(x + \bar{x})^2}{S_X^2}}} \qquad \ldots (11)$$

donde $\pm t$ son los valores críticos de la distribución t de student que corresponden a un nivel de confianza $(1 - \alpha)$ con $\nu = n - 2$.

VIII.4 ESTIMACION POR INTERVALOS DE CONFIANZA PARA Y DADO x

El estadístico $\hat{y}$ que se obtiene con el método de los mínimos cuadrados se utiliza más como un estimador puntual del valor que toma Y cuando $X = x$, que como estimador de $\mu_{Y/x}$; es decir, en la *práctica* lo que se busca es predecir el valor y que corresponde a una determinada x. La expresión (11) proporciona una estimación por intervalos de la media condicional $\mu_{Y/x}$ y no de la variable Y directamente, por lo cual estos intervalos tienen poca importancia práctica.

Para establecer el intervalo de confianza que contiene al valor de Y dado x, se podría usar la variable:

$$Z = \frac{Y - \mu_{Y/x}}{\sigma_{Y/x}}$$

pues se ha considerado que Y tiene distribución normal, sin embargo, también en este caso la muestra debe ser grande para que la estimación de $\sigma_{Y/x}$ por medio de $S_{Y/x}$ sea aceptable.

Por esta razón, para establecer los límites de confianza, se utiliza la variable:

$$T = \frac{\hat{y} - Y}{\frac{S_{Y/x}}{\sqrt{n-2}} \sqrt{n + 1 + \frac{(x + \bar{x})^2}{S_X^2}}}$$

que tiene distribución t de student con $\nu = n - 2$ grados de libertad. Con lo cual los límites de confianza para Y dado x son:

$$\boxed{\hat{y} \pm t \frac{S_{Y/x}}{\sqrt{n-2}} \sqrt{n + 1 + \frac{(x + \bar{x})^2}{S_X^2}}} \qquad \ldots (12)$$

donde $\pm$ t son los valores críticos de la distribución t de student, con $\nu = n - 2$, que corresponden a un nivel de confianza $(1 - \alpha)$ 100%.

BANDAS DE CONFIANZA

Si se obtienen los límites de confianza para algunos valores de X y se trazan en un diagrama de dispersión, por un lado, la curva que pasa por todos los límites superiores y, por otro, la línea que une los límites inferiores, como se muestra en la figura VIII.6, se tendrá una región entre las dos líneas que contiene al $(1 - \alpha)$ 100% de los posibles valores de Y. A esta región se le denomina *banda de confianza de* Y.

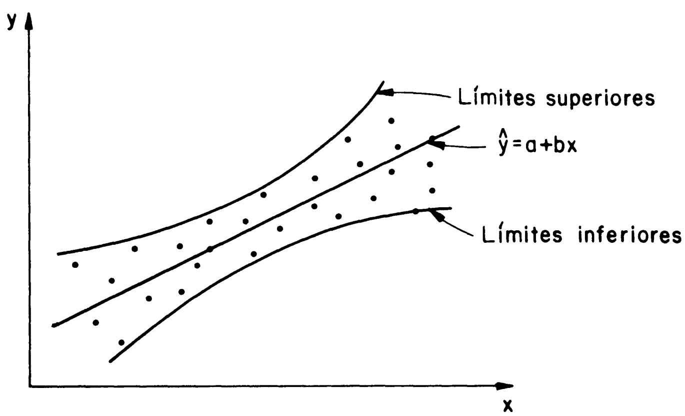


Figura VIII.6

De manera semejante se puede encontrar la banda de confianza para $\mu_{Y/x}$.

Ejemplo VIII.3

Con lòs datos del ejemplo VIII.1 y los resultados del ejemplo VIII.2:

a) Encontrar un intervalo de confianza del 95% para el número esperado de personas tituladas, cuando han egresado 12,000 personas.

b) Graficar la banda de confianza del 99% para Y.

Solución

a) Para poder aplicar la expresión (12) es necesario calcular primero:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{7} (71.89) = 10.27$$

$$S_x = \sqrt{\frac{1}{n} \Sigma x^2 - \bar{x}^2} = \sqrt{\frac{1}{7} (745.316) - (10.27)^2}$$

$$= 1$$

De la tabla de la distribución t de student con $\nu = 7 - 2 = 5$ y $F_T(t) = 0.975$, se encuentra que $t = 2.57$.

Del ejemplo VIII.1:

$$\hat{y}(12) = 0.1068 + 0.4191(12)$$

$$= 5.136$$

finalmente, del ejemplo VIII.2 $S_{Y/x} = 0.052$, con lo cual, los límites del intervalo son:

$$5.136 \pm \frac{2.57(0.052)}{\sqrt{7-2}} \sqrt{1 + \frac{(12 - 10.27)^2}{1}}$$

haciendo operaciones se obtiene que:

$$[5.017 \le y(12) \le 5.255]$$

Esto significa que para 12,000 estudiantes egresados la media condicional de los alumnos titulados está entre 5,017 y 5,255 con un nivel de confianza del 95%.

b) Mediante la expresión (12) se encontrarán los límites de confianza para Y, dado que x = 8,9,10,11 y 12.

Calculando los valores correspondientes de $\hat{y}$ con la recta de mínimos cuadrados:

$$\hat{y}\ (8) = 0.1068 + 0.4191\ (8) = 3.4596$$

$$\hat{y}\ (9) = 3.8787$$

$$\hat{y}(10) = 4.2978$$

$$\hat{y}(11) = 4.7169$$

$$\hat{y}(12) = 5.1360$$

El valor de la variable t de student con $\nu = 5$ que corresponde a $F_T(t) = 0.995$ es 4.03, con lo cual, los límites de confianza para x = 8 son:

$$3.4596 \pm \frac{4.03(0.052)}{\sqrt{7-2}} \sqrt{7 + 1 + \frac{(8 - 10.27)^2}{1}}$$

efectuando las operaciones:

$$[3.1197 \leq y(8) \leq 3.7995]$$

Repitiendo el procedimiento para los demás puntos:

$$[3.5881 \leq y(9) \leq 4.1692]$$

$$[4.0315 \leq y(10) \leq 4.5641]$$

$$[4.4431 \leq y(11) \leq 4.9907]$$

$$[4.8253 \leq y(12) \leq 5.4467]$$

Al trazar estos límites en el diagrama de dispersión unidos mediante dos curvas, como se muestra en la figura VIII.7, se obtiene la banda de confianza del 99% para Y.

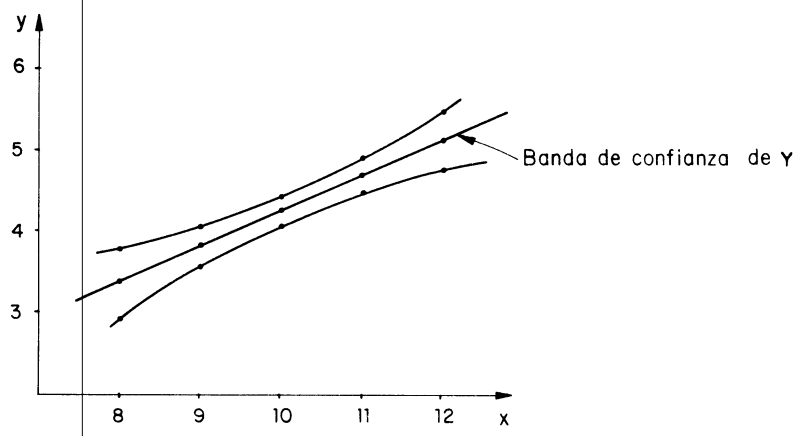
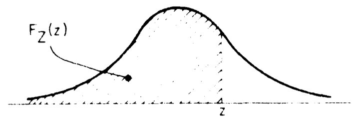


Figura VIII.7

APENDICE

TABLA A

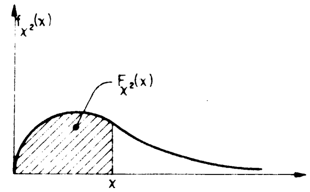
Función de distribución acumulada de la distribución normal estándar



$F_Z(z)$ / Z	0	1	2	3	4	5	6	7	8	9
- 3.4	0.00034	0.00033	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
- 3.3	0.00049	0.00047	0.00045	0.00044	0.00042	0.00041	0.00039	0.00038	0.00036	0.00035
- 3.2	0.00069	0.00067	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
- 3.1	0.00097	0.00094	0.00091	0.00088	0.00085	0.00082	0.00079	0.00077	0.00074	0.00071
- 3.0	0.00135	0.00131	0.00127	0.00123	0.00119	0.00115	0.00111	0.00107	0.00104	0.00100
- 2.9	0.00187	0.00181	0.00176	0.00170	0.00165	0.00159	0.00154	0.00149	0.00145	0.00140
- 2.8	0.00256	0.00248	0.00241	0.00233	0.00226	0.00219	0.00212	0.00206	0.00199	0.00193
- 2.7	0.00348	0.00337	0.00363	0.00317	0.00308	0.00299	0.00290	0.00281	0.00273	0.00264
- 2.6	0.00467	0.00454	0.00441	0.00428	0.00415	0.00403	0.00392	0.00380	0.00369	0.00358
- 2.5	0.00622	0.00605	0.00588	0.00571	0.00555	0.00540	0.00524	0.00509	0.00495	0.00481
- 2.4	0.00821	0.00799	0.00777	0.00756	0.00735	0.00715	0.00696	0.00677	0.00658	0.00640
- 2.3	0.01074	0.01045	0.01018	0.00991	0.00965	0.00940	0.00915	0.00890	0.00867	0.00843
- 2.2	0.01391	0.01356	0.01322	0.01288	0.01256	0.01224	0.01192	0.01161	0.01131	0.01102
- 2.1	0.01787	0.01744	0.01701	0.01660	0.01619	0.01579	0.01540	0.01501	0.01464	0.01427
- 2.0	0.02276	0.02222	0.02170	0.02119	0.02068	0.02019	0.01971	0.01924	0.01877	0.01832
- 1.9	0.02872	0.02807	0.02744	0.02681	0.02620	0.02560	0.02501	0.02443	0.02386	0.02330
- 1.8	0.03593	0.03515	0.03438	0.03363	0.03289	0.03216	0.03145	0.03075	0.03006	0.02939
- 1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03921	0.03837	0.03754	0.03673
- 1.6	0.05480	0.05370	0.05261	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551
- 1.5	0.06680	0.06552	0.06425	0.06300	0.06178	0.06057	0.05938	0.05820	0.05705	0.05591
- 1.4	0.08075	0.07926	0.07780	0.07635	0.07493	0.07352	0.07214	0.07077	0.06943	0.06811
- 1.3	0.09679	0.09509	0.09341	0.09175	0.09011	0.08850	0.08691	0.08534	0.08379	0.08226
- 1.2	0.11506	0.11313	0.11122	0.10934	0.10748	0.10564	0.10382	0.10203	0.10026	0.09852
- 1.1	0.13566	0.13349	0.13135	0.12923	0.12713	0.12506	0.12301	0.12099	0.11899	0.11701
- 1.0	0.15865	0.15624	0.15386	0.15150	0.14916	0.14685	0.14456	0.14230	0.14006	0.13785
- 0.9	0.18406	0.18141	0.17878	0.17618	0.17361	0.17105	0.16852	0.16602	0.16354	0.16108
- 0.8	0.21186	0.20897	0.20611	0.20327	0.20046	0.19766	0.19489	0.19215	0.18943	0.18673
- 0.7	0.24197	0.23886	0.23577	0.23270	0.22966	0.22663	0.22363	0.22065	0.21770	0.21477
- 0.6	0.27426	0.27094	0.26764	0.26436	0.26110	0.25786	0.25464	0.25144	0.24826	0.24511
- 0.5	0.30855	0.30504	0.30154	0.29807	0.29461	0.29117	0.28775	0.28435	0.28097	0.27761
- 0.4	0.34459	0.34091	0.33725	0.33361	0.32998	0.32637	0.32277	0.31919	0.31563	0.31208
- 0.3	0.38209	0.37828	0.37449	0.37071	0.36693	0.36318	0.35943	0.35570	0.35198	0.34828
- 0.2	0.42074	0.41683	0.41293	0.40904	0.40516	0.40129	0.39743	0.39358	0.38974	0.38591
- 0.1	0.46016	0.45620	0.45223	0.44827	0.44432	0.44037	0.43643	0.43250	0.42857	0.42465
- 0.0	0.50000	0.49601	0.49202	0.48803	0.48404	0.48005	0.47607	0.47209	0.46811	0.46413
+ 0.0	0.50000	0.50399	0.50798	0.51197	0.51596	0.51995	0.52393	0.52791	0.53189	0.53587
+ 0.1	0.53984	0.54380	0.54777	0.55173	0.55568	0.55963	0.56357	0.56750	0.57143	0.57535
+ 0.2	0.57926	0.58317	0.58707	0.59096	0.59484	0.59871	0.60257	0.60642	0.61026	0.61409
+ 0.3	0.61791	0.62172	0.62551	0.62929	0.63307	0.63682	0.64057	0.64430	0.64802	0.65172
+ 0.4	0.65541	0.65909	0.66275	0.66639	0.67002	0.67363	0.67723	0.68081	0.68437	0.68792
+ 0.5	0.69145	0.69496	0.69846	0.70193	0.70539	0.70883	0.71225	0.71565	0.71903	0.72239
+ 0.6	0.72574	0.72906	0.73236	0.73564	0.73890	0.74214	0.74536	0.74856	0.75174	0.75489
+ 0.7	0.75803	0.76114	0.76423	0.76730	0.77034	0.77337	0.77637	0.77935	0.78230	0.78523
+ 0.8	0.78814	0.79103	0.79389	0.79673	0.79954	0.80234	0.80511	0.80785	0.81057	0.81327
+ 0.9	0.81594	0.81859	0.82122	0.82382	0.82639	0.82895	0.83148	0.83398	0.83646	0.83892
+ 1.0	0.84135	0.84376	0.84614	0.84850	0.85084	0.85315	0.85544	0.85770	0.85994	0.86215
+ 1.1	0.86434	0.86651	0.86865	0.87077	0.87287	0.87494	0.87699	0.87901	0.88101	0.88299
+ 1.2	0.88494	0.88687	0.88878	0.89066	0.89252	0.89436	0.89618	0.89797	0.89974	0.90148
+ 1.3	0.90321	0.90491	0.90659	0.90825	0.90989	0.91150	0.91309	0.91466	0.91621	0.91774
+ 1.4	0.91925	0.92074	0.92220	0.92365	0.92507	0.92648	0.92786	0.92923	0.93057	0.93189
+ 1.5	0.93320	0.93448	0.93575	0.93700	0.93822	0.93943	0.94062	0.94180	0.94295	0.94409
+ 1.6	0.94520	0.94630	0.94739	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
+ 1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96079	0.96163	0.96246	0.96327
+ 1.8	0.96407	0.96485	0.96562	0.96637	0.96711	0.96784	0.96855	0.96925	0.96994	0.97061
+ 1.9	0.97128	0.97193	0.97256	0.97319	0.97380	0.97440	0.97499	0.97557	0.97614	0.97670
+ 2.0	0.97724	0.97778	0.97830	0.97881	0.97932	0.97981	0.98929	0.98076	0.98123	0.98168
+ 2.1	0.98213	0.98256	0.98299	0.98340	0.98381	0.98421	0.98460	0.98499	0.98536	0.98573
+ 2.2	0.98609	0.98644	0.98678	0.98712	0.98744	0.98776	0.98808	0.98839	0.98869	0.98898
+ 2.3	0.98926	0.98955	0.98982	0.99009	0.99035	0.99060	0.99085	0.99110	0.99133	0.99157
+ 2.4	0.99179	0.99201	0.99223	0.99244	0.99265	0.99285	0.99304	0.99323	0.99342	0.99360
+ 2.5	0.99378	0.99395	0.99412	0.99429	0.99445	0.99460	0.99476	0.99491	0.99505	0.99519
+ 2.6	0.99533	0.99546	0.99559	0.99572	0.99585	0.99597	0.99608	0.99620	0.99631	0.99642
+ 2.7	0.99652	0.99663	0.99673	0.99683	0.99692	0.99701	0.99710	0.99719	0.99727	0.99736
+ 2.8	0.99744	0.99752	0.99759	0.99767	0.99774	0.99781	0.99788	0.99794	0.99801	0.99807
+ 2.9	0.99813	0.99819	0.99824	0.99830	0.99835	0.99841	0.99846	0.99851	0.99855	0.99860
+ 3.0	0.99865	0.99869	0.99873	0.99877	0.99881	0.99885	0.99889	0.99893	0.99896	0.99900
+ 3.1	0.99903	0.99906	0.99909	0.99912	0.99915	0.99918	0.99921	0.99923	0.99926	0.99929
+ 3.2	0.99931	0.99933	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
+ 3.3	0.99951	0.99953	0.99955	0.99956	0.99958	0.99959	0.99961	0.99962	0.99964	0.99965
+ 3.4	0.99966	0.99967	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976

TABLA B

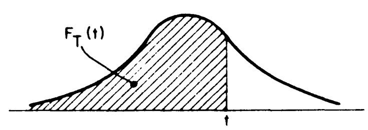
Función de distribución acumulada de la distribución χ^2 (ji-cuadrada)



$F_{\chi^2}(x)$ / ν	0.005	0.01	0.025	0.05	0.10	0.5	0.75	0.90	0.95	0.975	0.99	0.995
1	0.393E-04	0.157E-03	0.982E-03	0.393E-02	0.158E-01	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	2.37	4.11	6.25	7.82	9.35	11.35	12.84
4	0.21	0.30	0.48	0.71	1.06	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	10.22	13.36	15.51	17.53	20.09	21.95
9	1.73	2.09	2.70	3.33	4.17	8.34	11.39	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	12.55	15.99	18.31	20.48	23.71	25.19
11	2.60	3.05	3.82	4.58	5.58	10.34	13.70	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	14.85	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	17.12	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.57	8.67	10.09	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	17.34	21.60	25.99	28.87	31.53	34.81	37.16
19	6.85	7.63	8.91	10.12	11.65	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	23.83	28.41	31.41	34.17	37.57	40.00
21	8.04	8.90	10.28	11.59	13.24	20.3	24.9	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.3	26.0	30.81	33.93	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.3	27.1	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.3	28.2	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.3	29.3	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.3	30.4	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.3	31.5	36.74	40.11	43.19	46.96	49.65
28	12.46	13.56	15.31	16.93	18.94	27.3	32.6	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.3	33.7	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.3	34.8	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.3	45.6	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.3	56.3	63.17	67.51	71.42	76.16	79.49
60	35.53	37.48	40.48	43.19	46.46	59.3	67.0	74.40	79.08	83.30	88.38	91.95
70	43.27	45.44	48.76	51.74	55.33	69.3	77.6	85.53	90.53	95.02	100.43	104.22
80	51.17	53.54	57.15	60.39	64.28	79.3	88.1	96.58	101.88	106.63	112.33	116.32
90	59.19	61.75	65.65	69.13	73.29	89.3	98.6	107.57	113.15	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	99.3	109.1	118.50	124.34	129.56	135.81	140.17

TABLA C

Función de distribución acumulada de la distribución t de student



$F_T(t)$ / ν	.55	.60	.70	.75	.80	.90	.95	.975	.99	.995
1	.158	.325	.727	1.000	1.376	3.08	6.31	12.71	31.82	63.66
2	.142	.289	.617	.816	1.061	1.89	2.92	4.30	6.96	9.92
3	.137	.277	.584	.765	.978	1.64	2.35	3.18	4.54	5.84
4	.134	.271	.569	.741	.941	1.53	2.13	2.78	3.75	4.60
5	.132	.267	.559	.727	.920	1.48	2.02	2.57	3.36	4.03
6	.131	.265	.553	.718	.906	1.44	1.94	2.45	3.14	3.71
7	.130	.263	.549	.711	.896	1.42	1.90	2.36	3.00	3.50
8	.130	.262	.546	.706	.889	1.40	1.86	2.31	2.90	3.36
9	.129	.261	.543	.703	.883	1.38	1.83	2.26	2.82	3.25
10	.129	.260	.542	.700	.879	1.37	1.81	2.23	2.76	3.17
11	.129	.260	.540	.697	.876	1.36	1.80	2.20	2.72	3.11
12	.128	.259	.539	.695	.873	1.36	1.78	2.18	2.68	3.06
13	.128	.259	.538	.694	.870	1.35	1.77	2.16	2.65	3.01
14	.128	.258	.537	.692	.868	1.34	1.76	2.14	2.62	2.98
15	.128	.258	.536	.691	.866	1.34	1.75	2.13	2.60	2.95
16	.128	.258	.535	.690	.865	1.34	1.75	2.12	2.58	2.92
17	.128	.257	.534	.689	.863	1.33	1.74	2.11	2.57	2.90
18	.127	.257	.534	.688	.862	1.33	1.73	2.10	2.55	2.88
19	.127	.257	.533	.688	.861	1.33	1.73	2.09	2.54	2.86
20	.127	.257	.533	.687	.860	1.32	1.72	2.09	2.53	2.84
21	.127	.257	.532	.686	.859	1.32	1.72	2.08	2.52	2.83
22	.127	.256	.532	.686	.858	1.32	1.72	2.07	2.51	2.82
23	.127	.256	.532	.685	.858	1.32	1.71	2.07	2.50	2.81
24	.127	.256	.531	.685	.857	1.32	1.71	2.06	2.49	2.80
25	.127	.256	.531	.684	.856	1.32	1.71	2.06	1.48	2.79
26	.127	.256	.531	.684	.856	1.32	1.71	2.06	2.48	2.78
27	.127	.256	.531	.684	.855	1.31	1.70	2.05	2.47	2.77
28	.127	.256	.530	.683	.855	1.31	1.70	2.05	2.47	2.76
29	.127	.256	.530	.683	.854	1.31	1.70	2.04	2.46	2.76
30	.127	.256	.530	.683	.854	1.31	1.70	2.04	2.46	2.75
40	.126	.255	.529	.681	.851	1.30	1.68	2.02	2.42	2.70
60	.126	.254	.527	.679	.848	1.30	1.67	2.00	2.39	2.66
120	.126	.254	.526	.677	.845	1.29	1.66	1.98	2.36	2.62
∞	.126	.253	.524	.674	.842	1.28	1.645	1.96	2.33	2.58

TABLA D

Dígitos aleatorios

5970	7277	3938	3429	2404	9134	641	2359	5813	6601
8780	9877	702	2997	35	9058	8116	5594	7056	2499
2744	7916	3069	8965	4875	8585	3419	6752	6831	2467
7970	8680	9952	9005	8039	7454	9201	6502	2098	1663
3029	6400	9899	7967	8325	2189	5720	6287	4225	9377
6057	6470	5941	848	809	7823	3185	8790	5076	8247
3342	7160	7937	3261	2943	7109	1600	5902	607	980
5421	8846	2414	9402	8143	7248	8108	4192	1961	4570
672	718	9914	7519	6889	4102	1836	7876	4059	2016
6408	5005	9835	2883	5311	3911	6233	9285	1695	1544
7473	8690	2777	8255	1290	2517	8923	6833	3131	3321
2333	2733	386	1854	8049	2081	9039	489	1187	3677
6670	8785	2695	2827	5245	5300	2075	4742	1840	6823
4480	3412	2969	4219	7564	7127	2224	9071	314	4069
6659	9809	51	3438	817	3989	2224	5914	4675	544
7102	7019	8204	228	5530	508	704	6137	6923	9414
6299	6657	4452	2134	8041	3715	7023	2003	1590	3601
2422	7122	3429	7473	9091	4776	3616	7641	1830	5001
9923	5321	9719	5804	9920	8205	5840	7010	3012	1204
5629	5885	7703	5892	5859	8581	9314	9377	3818	9712
2583	2936	3146	5147	4876	1384	288	3277	1484	951
7079	9615	3661	9291	6673	2330	1193	3281	7185	8498
9936	651	7260	5868	2396	6323	8435	9072	1623	4884
6140	1432	9097	5927	5457	2687	5206	5556	232	85
3751	9211	2953	9488	5313	1204	5042	8955	4959	632
3751	1073	4821	2265	4810	7584	2913	1027	9109	2577
7384	9409	4008	9886	5027	6433	3817	790	9689	1960
1001	4880	1725	9609	7619	1723	8878	3732	7249	2784
8407	6886	7180	9540	4659	4759	8934	8538	3230	2476
699	1539	1173	5356	5987	9650	5636	1303	4804	4862
3613	7126	7190	4519	4047	6275	37	7253	7219	633
2369	7081	9995	7998	6237	754	8689	9964	3643	5311
4012	455	1727	9485	4749	3416	7237	8082	2509	4725
7258	9880	5491	2118	3912	4269	1510	9536	2356	7973
9834	7044	6452	2696	5041	5964	6116	3266	4175	7893
3328	5654	430	3397	8773	4271	538	8432	1250	9034
5527	1908	9992	1002	4919	6042	2724	1140	6507	3120
264	2464	8048	3607	9799	4682	184	8658	7352	2025
4758	9906	8944	4847	1095	3119	8580	1134	8017	8234
4465	5716	7053	5613	188	4267	7824	815	9406	2820

B I B L I O G R A F I A

KENNEDY, J.B. y NEVILLE, A.M. *Estadística para ciencias e ingeniería.* 2a. edición. Editorial Harla. México, 1982.

LARSON, HAROLD J. *Introducción a la teoría de probabilidades e inferencia estadística.* Editorial Limusa. México, 1978.

MILLER, I. y FREUND, J.E. *Probability and statistics for engineers.* 3rd. edition. Prentice-Hall. New Jersey, 1985.

MOOD, A.M., GRAYBILL, F.A., BOES, D.C. *Introduction to the theory of statistics.* 3rd. edition. Mc Graw-Hill. Tokyo, 1974.

OBREGON SANIN, I. *Teoría de la probabilidad.* Editorial Limusa. México, 1977.

SCHEAFFER, R.L. y Mc. CLAVE, J.T. *Statistics for engineers.* Duxbury Press. Boston, 1982.

WALPOLE, R.E. y MYERS, R.H. *Probabilidad y estadística para ingenieros.* Editorial Interamericana. México, 1983.

Estos apuntes se terminaron de imprimir en el mes de junio de 1985, en los Talleres de la Facultad de Ingeniería de la U.N.A.M. La edición consta de 3,500 ejemplares.