FUNDAMENTOS DE LAS TECNICAS DE MUESTREO ESTADISTICO

Fecha	Duración	Tema	Profesor
Febrero 13	18 a 21 h	Introducción	M. en C. Edmundo Berumen Torr
Feb. 15, 20, 22	18 a 21 h	Muestreo Aleatorio Simple	Dr. Octavio A. Rascón Chávez
Feb. 27	18 a 19:30 h	Muestreo Aleatorio Simple	Dr. Octavio A. Rascón Chávez
Feb. 27	19:30 a 21 h	Muestreo Aleatorio Simple	M. en I. Augusto Villarreal Aranc
Marzo 1° y 6	18 a 21 h	Tamaño de la Muestra	M. en I. Augusto Villarreal Aranc
Marzo 8 y 13	18 a 21 h	Muestreo Aleatorio Simple para Razones	M. en C. Adela Abad de Servín
Marzo 15, 20 y 22	18 a 21 h	Muestreo Estratificado	M. en C. Adela Abad de Servín
Marzo 27 y 29	18 a 21 h	Muestreo por Conglomerados	M. en C. Adela Abad de Servín
Abril 3 y 5	18 a 21 h	Muestreo Sistemático	M. en C. Edmundo Berumen Torr

DIRECTORIO DE PROFESORES DEL CURSO: FUNDAMENTOS DE LAS TECNICAS DE MUESTREO ESTADISTICO 1979

1. M. EN C. ADELA ABAD DE SERVIN
   Profesora
   Escuela Nacional de Estudios Profesionales
   Acatlán , UNAM
   Estado de México
   Tel. 373.23.18

2. M. EN C. EDMUNDO BERUMEN TORRES
   Subdirector de Sistemas de Evaluación
   Secretaría de Educación Pública
   Añil No. 571-6°
   Col. Granjas México
   México, D.F.
   Tel.

3. M. EN I. AUGUSTO VILLARREAL ARANDA
   Investigador
   Instituto de Ingeniería
   U.N.A.M.
   México 20, D.F.

4. DR. OCTAVIO A. RASCON CHAVEZ
   Insurgentes Sur 1023 Edif. Detroit 6°Piso
   México, D.F.
   Tel. 598.04.89

EVALUACION DE LA ENSEÑANZA

CURSO: FUNDAMENTOS DE LAS TECNICAS DE MUESTREO ESTADISTICO FECHAS: Del 13 de febrero al 5 de abril, 1979.	DOMINIO DEL TEMA	EFICIENCIA EN EL USO DE AYUDAS AUDIOVISUALES	MANT. DEL INTERES (AMENIDAD, FACILIDAD DE EXPRESION, COMUNICACION CON LOS ASISTENTES)	PUNTUALIDAD
Introducción (Berumen Torres)				
Muestreo Aleatorio Simple (Rascón)				
Tamaño de la Muestra (Villarreal)				
Muestreo Aleatorio Simple para Razones (Abad)				
Muestreo Estratificado (Abad)				
Muestreo por Conglomerados (Abad)				
Muestreo Sistemático (Berumen)				
ESCALA DE EVALUACION DEL 1 AL 10 'edcs. 22, I, 78.				

EVALUACION DEL CURSO

	CONCEPTO	EVALUACION
1.	APLICACION INMEDIATA DE LOS CONCEPTOS EXPUESTOS	
2.	CLARIDAD CON QUE SE EXPUSIERON LOS TEMAS	
3.	GRADO DE ACTUALIZACION LOGRADO CON EL CURSO	
4.	CUMPLIMIENTO DE LOS OBJETIVOS DEL CURSO	
5.	CONTINUIDAD EN LOS TEMAS DEL CURSO	
6.	CALIDAD DE LAS NOTAS DEL CURSO	
7.	GRADO DE MOTIVACION LOGRADO CON EL CURSO	

ESCALA DE EVALUACION DE 1 A 10

1. ¿Qué le pareció el ambiente del Centro de Educación Continua?

   Muy agradable ☐ Agradable ☐ Desagradable ☐

2. Medio de comunicación por el que se enteró del curso:

   Periódico Excélsior ☐ Periódico Novedades ☐ Folleto del Curso ☐

   Cartel mensual ☐ Radio Universidad ☐ Comunicación carta, teléfono, verbal, etc. ☐

3. Medio de transporte utilizado para venir al Palacio de Minería:

   Automóvil particular ☐ Metro ☐ Otro medio ☐

4. ¿Qué cambios haría usted en el programa para tratar de perfeccionar el curso? ______________________________

   ____

   ____

5. ¿Recomendaría el curso a otras personas? Si ☐ No ☐

6. ¿Qué curso le gustaría que ofreciera el Centro de Educación Continua? ______________________________

   ____

   ____

7. ¿Qué servicios desearía que tuviese el CEC para los asistentes a cursos? ______________________________

   ____

   ____

8. Otras sugestiones: ______________________________

   ____

   ____

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

FUNDAMENTOS DE LAS TECNICAS DE MUESTREO ESTADISTICO

TEORIA DE PROBABILIDADES

DR. OCTAVIO A. RASCON CHAVEZ

FEBRERO, 1979.

TEORIA DE PROBABILIDADES

1982 DR. OCTAVIO A. RASCON

EXPERIMENTO. - PARA FINES DE ESTE CURSO, SE ENTENDERA POR EXPERIMENTO A TODO PROCESO DE OBSERVACION. ASI UN EXPERIMENTO PUEDE SER PLANEADO Y REALIZADO POR EL HOMBRE, O PUEDE SER EFECTUADO POR LA NATURALEZA EN CASO DE UN FENOMENO NATURAL. POR EJEMPLO, EL LANZAR UNA MONEDA O UN DADO Y OBSERVAR LA CARA QUE QUEDA HACIA ARRIBA, ES UN EXPERIMENTO PLANEADO Y REALIZADO POR EL HOMBRE. EL OBSERVAR LA CANTIDAD DE AGUA QUE LLUEVE ANUALMENTE EN UNA CIUDAD, ES UN EXPERIMENTO ASOCIADO A UN FENOMENO NATURAL.

AL RESULTADO DE UN EXPERIMENTO SE LE DENOMINA *DATO*. A UN GRUPO DE DATOS SE LE LLAMA *MUESTRA*.

PROBABILIDAD: ES UNA MEDIDA DE LA CERTIDUMBRE QUE SE LE ASOCIA A LA OCURRENCIA U OBSERVACION DE UN RESULTADO DETERMINADO, AL REALIZARSE EL EXPERIMENTO CORRESPONDIENTE.

LA *TEORIA DE PROBABILIDADES* ES UNA RAMA DE LAS MATEMATICAS APLICADAS QUE TRATA LO CONCERNIENTE A LA ASIGNACION Y MANEJO DE PROBABILIDADES.

ESTADISTICA: ES LA RAMA DE LAS MATEMATICAS QUE SE ENCARGA DE ENSEÑAR LAS REGLAS PARA COLECTAR, ORGANIZAR, PRESENTAR Y PROCESAR LOS DATOS OBTENIDOS AL REALIZAR VARIAS VECES EL EXPERIMENTO ASOCIADO A UN FENOMENO DE INTERES Y PARA INFERIR CONCLUSIONES ACERCA DE ESTE ULTIMO. PROPORCIONA, ADEMAS, LOS METODOS PARA EL DISEÑO DE EXPERIMENTOS Y PARA TOMAR DECISIONES CUANDO APARECEN SITUACIONES DE INCERTIDUMBRE.

ESTADISTICA

* *DESCRIPTIVA.*- TRATA LO CONCERNIENTE A LA OBTENCION, ORGANIZACION, PROCESAMIENTO Y PRESENTACION DE LOS DATOS.
* *INFERENCIAL.*- TRATA LO CONCERNIENTE A LOS METODOS PARA INFERIR CONCLUSIONES ACERCA DEL FENOMENO DEL CUAL PROVIENEN LOS DATOS

TEORIA DE PROBABILIDADES

AL LANZAR UNA MONEDA NO PODEMOS PREDECIR CON CERTEZA CUAL CARA QUEDARA HACIA ARRIBA. LO UNICO QUE SE PUEDE ASEGURAR, SI LA MONEDA NO ESTA CARGADA, ES QUE AMBAS CARAS TIENEN LA MISMA OPORTUNIDAD DE SALIR, ES DECIR, QUE LOS EVENTOS SIMPLES {CARA} Y {CRUZ} TIENEN LA MISMA *PROBABILIDAD* DE OCURRIR.

COMO YA SE DIJO, LA *PROBABILIDAD* DE QUE OCURRA UN EVENTO ES UNA MEDIDA DEL GRADO DE CONFIANZA QUE SE TIENE DE QUE ESTE OCURRA AL REALIZAR EL EXPERIMENTO CORRESPONDIENTE.

EXISTEN POR LO MENOS TRES MANERAS DE ASIGNARLE UNA PROBABILIDAD A UN EVENTO:

1. EN TERMINOS DE LOS RESULTADOS DE REPETIR VARIAS VECES UN EXPERIMENTO (METODO FRECUENCIAL).
2. APLICANDO LA DEFINICION CLASICA DE PROBABILIDADES.
3. CON BASE EN UN MODELO MATEMATICO (PROBABILISTICO) DEL FENOMENO DE QUE SE TRATE.

METODO FRECUENCIAL

SI N(A) ES EL NUMERO DE VECES QUE SE OBSERVA EL EVENTO A AL REALIZAR N VECES UN EXPERIMENTO, LA <u>FRECUENCIA RELATIVA</u> DE A, DEFINIDA COMO N(A)/N, SE CONSIDERA COMO ESTIMACION DE LA PROBABILIDAD DE A,

$$P(A) = \frac{N(A)}{N}$$

YA QUE, EN EL LIMITE, $P(A) = \lim_{N\to\infty} \frac{N(A)}{N}$

<u>EJEMPLO</u>

DE UNA URNA QUE CONTIENE BOLAS ROJAS, BLANCAS Y AZULES, SE SACO UNA BOLA, SE ANOTO SU COLOR Y SE REGRESO A LA URNA. SI ESTE EXPERIMENTO SE REPITE 20 VECES Y LOS RESULTADOS SON

b,b,a,r,r,r,a,b,r,a,b,b,a,r,b,r,r,a,r,a, DONDE

r = ROJA, b = BLANCA, a = AZUL

¿QUE PROBABILIDADES LE ASIGNARIA A LOS EVENTOS $B=\{b\}$, $A=\{a\}$, y $R=\{r\}$, DE ACUERDO CON EL METODO FRECUENCIAL?

EN ESTA MUESTRA SE TIENE QUE N(B)=6, N(A)=6, N(R)=8, N=20

POR LO QUE $P(B)=\frac{6}{20} = \frac{3}{10}$; $P(A) = \frac{6}{20} = \frac{3}{10}$; $P(R) = \frac{8}{20} = \frac{4}{10}$

NOTESE QUE LOS EVENTOS B, A Y R SON MUTUAMENTE EXCLUSIVOS, YA QUE SON EVENTOS SIMPLES, Y QUE

$$P(B) + P(A) + P(R) = \frac{3}{10} + \frac{3}{10} + \frac{4}{10} = 1 = P(S)$$

EN DONDE $S = \{r,b,a\}$

DEFINICION CLASICA DE PROBABILIDADES

SI N(A) ES EL NUMERO DE MANERAS IGUALMENTE PROBABLES EN QUE PUEDE OCURRIR EL EVENTO A Y N ES EL NUMERO TOTAL DE ELEMENTOS DEL ESPACIO DE EVENTOS CORRESPONDIENTE, ENTONCES LA PROBABILIDAD DE A ES

$$P(A) = \frac{N(A)}{N}$$

EJEMPLOS

A) SI EN UNA URNA SE TIENEN 5 BOLAS BLANCAS Y 15 NEGRAS, Y SE VA A SELECCIONAR UNA AL AZAR, ¿CUAL ES LA PROBABILIDAD DE QUE SEA BLANCA(A={BLANCA})?:

$$N= 5+15=20;\ N(A)=5 \Rightarrow P(A)=\frac{5}{20}= \frac{1}{4}$$

B) SI SE LANZAN DOS DADOS, ¿CUAL ES LA PROBABILIDAD DE QUE

1. SALGA UN 2 Y UN 5 (EVENTO B)?
2. LA SUMA SEA 7 (EVENTO A)

PARA EL INCISO 1 EL ESPACIO DE EVENTOS ES:

$$S = \left\{ \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2.1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{array} \right\}$$

SI EL DADO NO ESTA CARGADO, CADA PAREJA DE NUMEROS ES IGUALMENTE PROBABLE. EN TAL CASO, N=36 y N(B)=2 (APARECE (2,5) O (5,2))

$\Rightarrow$ P(B)=2/36=1/18.

PARA EL INCISO 2 EL ESPACIO DE EVENTOS ES

S_1={2,3,4,5,6,7,8,9,10,11,12}

PERO NO TODOS LOS ELEMENTOS (EVENTOS SIMPLES) SON IGUALMENTE PROBA-

BLES, YA QUE, POR EJEMPLO, EL 2 SOLO APARECERA SI SE OBSERVA LA PAREJA (1,1), EN CAMBIO EL 3 APARECERA SI OCURREN LAS PAREJAS (1,2) O (2.1), ES DECIR, EL 3 TIENE EL DOBLE DE PROBABILIDAD QUE EL 2. POR ESTO, PARA CALCULAR LA PROBABILIDAD DE QUE LA SUMA SEA 7 ES NECESARIO TRABAJAR CON EL ESPACIO S Y CONTAR LAS MANERAS POSIBLES DE QUE LA SUMA SEA 7, LO CUAL OCURRE SI SE OBSERVA CUALQUIERA DE LAS PAREJAS (6,1), (5,2), (4,3), (3,4), (2,5) o (1,6), ES DECIR, HAY 6 MANERAS IGUALMENTE PROBABLES DE QUE OCURRA EL EVENTO A. POR LO TAN

$$P(A) = \frac{N(A)}{N} = \frac{6}{36} = \frac{1}{6}$$

PROCEDIENDO DE ESTA MANERA SE PUEDEN CALCULAR LAS PROBABILIDADES DE QUE LA SUMA SEA 2,3,4,ETC. LOS RESULTADOS SON:

$$P(\{2\}) = \frac{1}{36};\ P(\{3\}) = \frac{2}{36};\ P(\{4\}) = \frac{3}{36};\ P(\{5\}) = \frac{4}{36};$$

$$P(\{6\}) = \frac{5}{36};\ P(\{7\}) = \frac{6}{36};\ P(\{8\}) = \frac{5}{36};\ P(\{9\}) = \frac{4}{36};$$

$$P(\{10\}) = \frac{3}{36}\ ;\ P(\{11\}) = \frac{2}{36}\ \text{y}\ P(\{12\}) = \frac{1}{36}$$

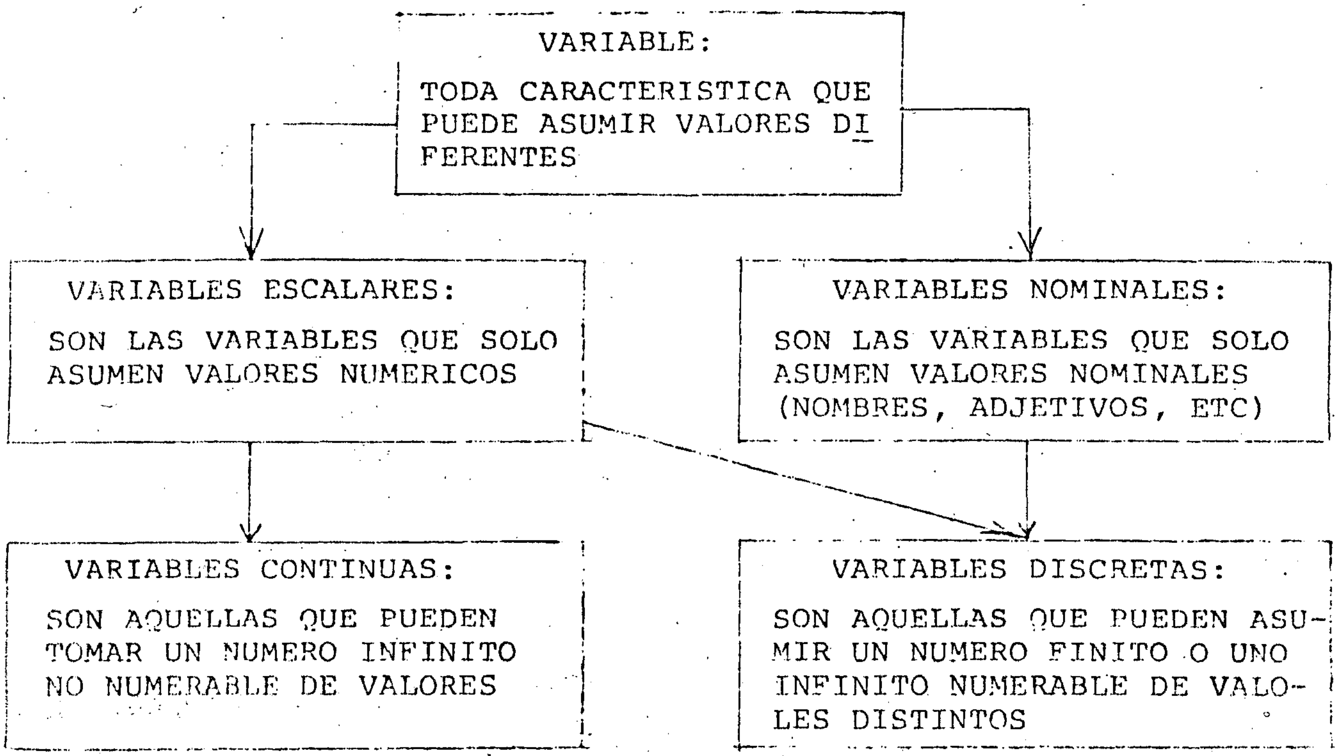
(OBSERVESE QUE $\sum_{i=2}^{12} P(\{i\})=1$

ASIGNACION DE PROBABILIDADES MEDIANTE UN MODELO MATEMATICO

MEDIANTE ESTE METODO LAS PROBABILIDADES SE ASIGNAN A PARTIR DE UN MODELO MATEMATICO QUE INVOLUCRE TODOS LOS FACTORES POSIBLES QUE INTERVIENEN EN LA ALEATORIEDAD DEL FENOMENO.

VARIABLES ALEATORIAS

CLASIFICACION DE VARIABLES



UNA *VARIABLE ALEATORIA* ES UNA VARIABLE TAL QUE NO PUEDE PREDECIRSE CON CERTEZA EL VALOR QUE ASUMIRA AL REALIZAR UN EXPERIMENTO. POR EJEMPLO, LA RESISTENCIA O CARGA DE FALLA DE UNAS VIGAS ES UNA VARIABLE ALEATORIA, YA QUE ANTES DE ROMPER UNA VIGA TOMADA AL AZAR NO SE PUEDE PRECISAR CUAL SERA SU RESISTENCIA. EN LA SIGUIENTE TABLA SE PRESENTAN LOS RESULTADOS EXPERIMENTALES CON 15 VIGAS DE CONCRETO REFORZADO, OBSERVANDOSE QUE ESTOS VARIAN DE UNAS A OTRAS DE MANERA ALEATORIA.

TABLA 2. PRUEBAS DE VIGAS DE CONCRETO REFORZADO

Número de la viga	Carga de agrietamiento, en kg	Carga de falla, en kg
1	4 700	4 700
2	3 840	4 220
3	3 270	4 360
4	2 310	4 680
5	2 950	4 270
6	4 810	4 810
7	2 720	4 590
8	2 720	4 490
9	4 310	4 310
10	2 950	4 630
11	4 220	4 220
12	2 720	4 340
13	2 720	4 340
14	2 630	4 770
15	2 950	4 630

A TODO EXPERIMENTO SE LE PUEDE ASOCIAR AL MENOS UNA VARIABLE ALEATORIA, DEPENDIENDO ESTA DEL PROBLEMA QUE SE TENGA PLANTEADO. POR EJEMPLO, EN EL CASO DE LA RESISTENCIA DE LAS VIGAS DE VARIABLE ALEATORIA PUEDE SER DIRECTAMENTE DICHA RESISTENCIA, EN CUYO CASO SU ESPACIO DE EVENTOS SERIA

$$S_1 = \{X: 0<X<\infty\}$$

LA VARIABLE TAMBIEN PUDO HABER SIDO UNA CUYO ESPACIO DE EVENTOS FUERA

$$S_2 = \{\text{EXITO, FRACASO}\}$$

EN DONDE EL EXITO OCURRIRIA SI LA VIGA RESISTIERA MAS DE CIERTA CANTIDAD, POR EJEMPLO 4600 KG, Y EL FRACASO OCURRIRIA SI RESISTIERA MENOS, ES DECIR:

EXITO: SI $X \geq 4600$ KG

FRACASO: SI $X < 4600$ KG

LEYES DE PROBABILIDADES

EL COMPORTAMIENTO DE UNA VARIABLE ALEATORIA SE DESCRIBE MEDIANTE SU *LEY DE PROBABILIDADES*, LA CUAL PUEDE ESPECIFICARSE DE DIFERENTES FORMAS. LA MANERA MAS COMUN DE HACERLO ES MEDIANTE SU *DISTRIBUCION O DENSIDAD DE PROBABILIDADES.*

A FIN DE EVITAR CONFUSION, SE EMPLEARA UNA LETRA MAYUSCULA PARA DENOTAR UNA VARIABLE ALEATORIA, Y LA MINUSCULA CORRESPONDIENTE PARA LOS VALORES QUE PUEDE ASUMIR. SI LA VARIABLE ALEATORIA X ES DISCRETA Y PUEDE ASUMIR LOS VALORES x_i, SU DENSIDAD DE PROBABILIDADES, $f_X(x)$ SERA EL CONJUNTO DE LAS PROBABILIDADES

$$P_X(x_i) = P(X = x_i)$$

LA CUAL SE LEE "PROBABILIDAD DE QUE $X = x_i$". ESTO ES

$$f_X(x) = \{P_X(x_i)\}$$

PARA QUE UNA DENSIDAD DE PROBABILIDADES SATISFAGA LOS TRES AXIOMAS DE LA TEORIA DE PROBABILIDADES, SE DEBEN CUMPLIR LOS SIGUIENTES REQUISITOS

A) $0 \leq P_X(x_i) \leq 1$ PARA TODA x_i

B) $\sum_{i=1}^{n} P_X(x_i) = 1$, DONDE n ES EL NUMERO TOTAL DE VALORES QUE PUEDE ASUMIR X

C) $P(x_m \leq X \leq x_r) = \sum_{i=m}^{i=r} P_X(x_i) \quad ; \; m \leq r$

LA DISTRIBUCION DE PROBABILIDADES ACUMULADAS O FUNCION DE DISTRIBUCION

OTRA FORMA DE ESPECIFICAR LA LEY DE PROBABILIDADES DE UNA VARIABLE ALEATORIA ES MEDIANTE LA DISTRIBUCION DE PROBABILIDADES ACUMULADAS, $F_X(x_i)$, QUE SE DEFINE COMO EL CONJUNTO DE LAS SUMAS PARCIALES DE LAS PROBABILIDADES, $P_X(x_i)$, CORRESPONDIENTES A TODOS LOS VALORES DE X MENORES O IGUALES QUE x_i. POR LO TANTO, ESTA FUNCION DA LAS PROBABILIDADES DE QUE LA VARIABLE ALEATORIA TOME VALORES MENORES O IGUALES QUE x_m PARA CUALQUIER m, ES DECIR

$$F_X(x) = \{F_X(x_m)\} \quad ; \quad m = 1,2,\ldots,n$$

EN DONDE

$$F_X(x_m) = \sum_{i=1}^{i=m} P_X(x_i) = P(X \leq x_m)$$

EJEMPLO

SEA X LA VARIABLE ALEATORIA DISCRETA "NUMERO TOTAL DE CARROS QUE SE DETIENEN EN UNA ESQUINA DEBIDO A LA LUZ ROJA DE UN SEMAFORO". SI LAS PROBABILIDADES ASOCIADAS A CADA VALOR, DETERMINADAS POR EL METODO FRECUENCIAL, SON

$$P_X(x) = \begin{cases} 0.1 & \text{SI} \quad x = 0 \\ 0.2 & \text{SI} \quad x = 1 \\ 0.3 & \text{SI} \quad x = 2 \\ 0.2 & \text{SI} \quad x = 3 \\ 0.1 & \text{SI} \quad x = 4 \\ 0.1 & \text{SI} \quad x = 5 \\ 0 & \text{SI} \quad x \geq 6 \end{cases}$$

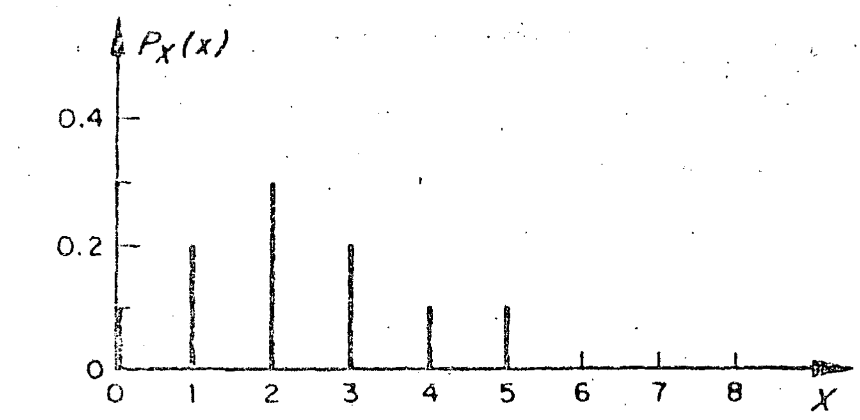
LA DISTRIBUCION DE PROBABILIDADES Y LA DE PROBABILIDADES ACUMULADAS CORRESPONDIENTES SERAN

x	$f_X(x)$	$F_X(x)$
<0	0	0
0	0.1	0.1
1	0.2	0.3
2	0.3	0.6
3	0.2	0.8
4	0.1	0.9
5	0.1	1.0
≥ 6	0	1.0

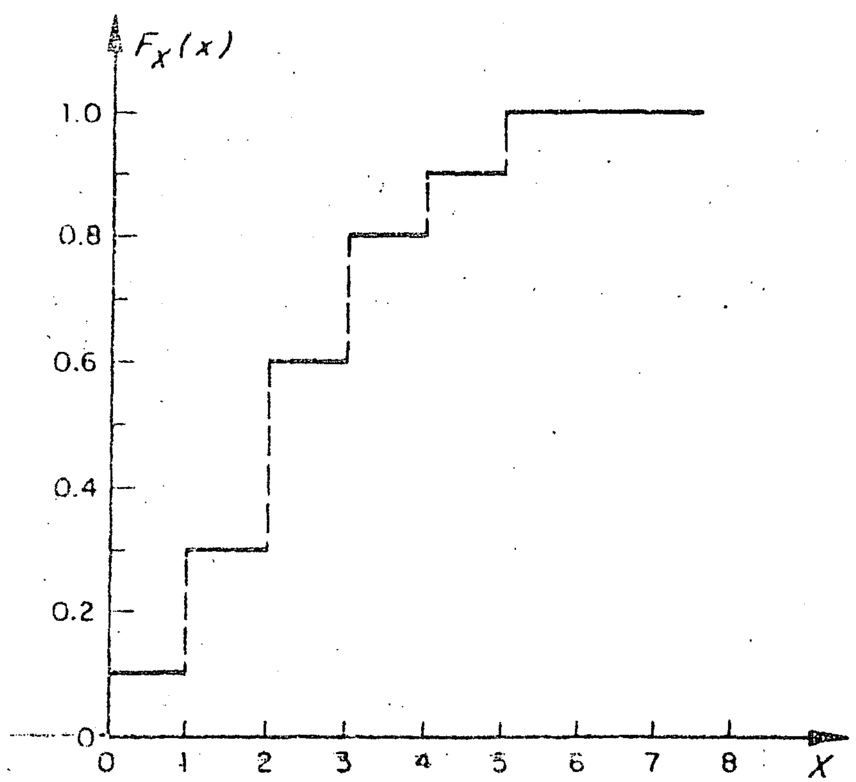
O SEA

$$F_X(x) = \begin{cases} 0, & \text{SI } x < 0 \\ 0.1, & \text{SI } 0 \leq x < 1 \\ 0.3, & \text{SI } 1 \leq x < 2 \\ 0.6, & \text{SI } 2 \leq x < 3 \\ 0.8, & \text{SI } 3 \leq x < 4 \\ 0.9, & \text{SI } 4 \leq x < 5 \\ 1.0, & \text{SI } 5 \leq x \end{cases}$$

LAS GRAFICAS DE ESTAS DISTRIBUCIONES SE PRESENTAN EN LA FIGURA DE LA SIGUIENTE HOJA.



a) Distribución de probabilidades



b) Función de distribución

Ley de probabilidades del ejemplo del tráfico

EJEMPLO

SEA LA VARIABLE ALEATORIA X DEFINIDA POR LA SUMA DE LOS DOS NUMEROS QUE QUEDEN HACIA ARRIBA AL LANZAR DOS DADOS. EN ESTE CASO EL ESPACIO DE EVENTOS ES

$$S = \{2,3,4,5,6,7,8,9,10,11,12\}$$

Y LA DENSIDAD DE PROBABILIDADES ES

$$f_X(x) = \{\frac{1}{36}, \frac{2}{36}, \frac{3}{36}, \frac{4}{36}, \frac{5}{36}, \frac{6}{36}, \frac{5}{36}, \frac{4}{36}, \frac{3}{36}, \frac{2}{36}, \frac{1}{36}\}$$

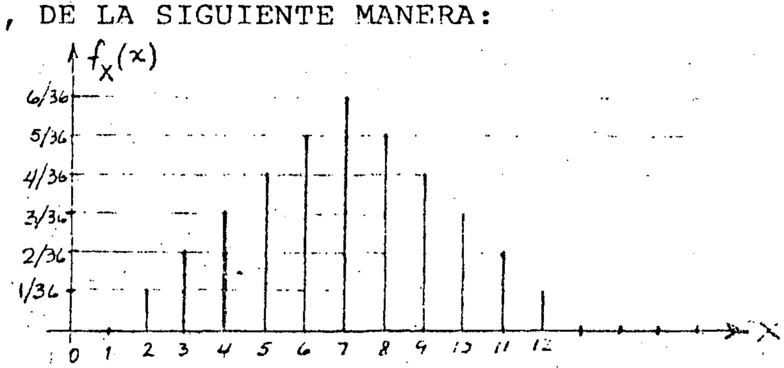
EN ESTE CASO $X_1=2$, $X_2=3,\ldots,X_{11}=12$

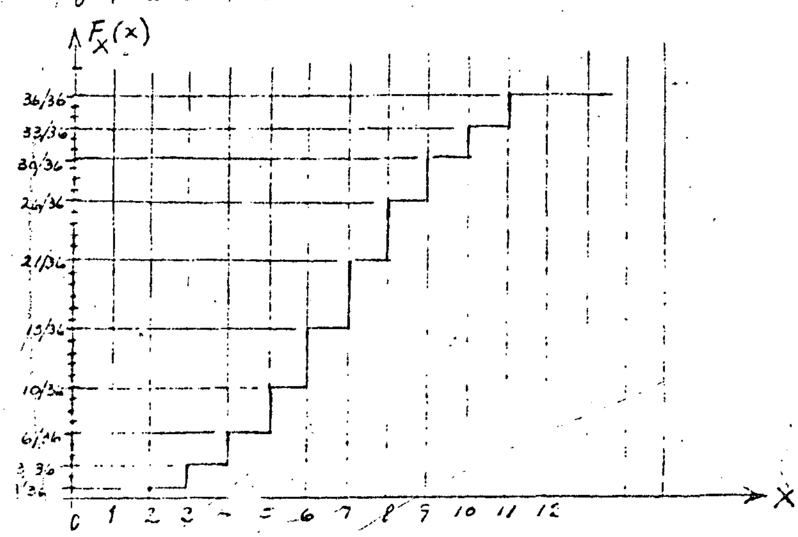
$$Y \quad f_X(2) = \frac{1}{36},\ f_X(3) = \frac{2}{36},\ldots,f_X(12) = \frac{1}{36}$$

ESTAS PROBABILIDADES FUERON CALCULADAS EN UN EJEMPLO PREVIO SOBRE PROBABILIDADES DE EVENTOS .

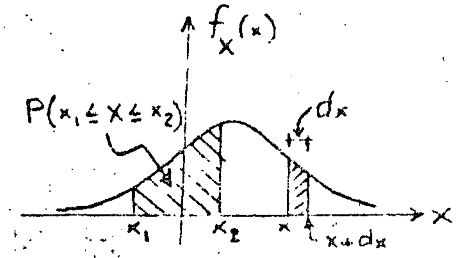
CON ESTAS PROBABILIDADES SE PUEDE OBTENER LA FUNCIÓN DE DISTRIBUCION O DE PROBABILIDADES ACUMULADAS, DE LA SIGUIENTE MANERA:

x	$f_X(x)$	$F_X(x)$
<2	0	0
2	1/36	1/36
3	2/36	3/36
4	3/36	6/36
5	4/36	10/36
6	5/36	15/36
7	6/36	21/36
8	5/36	26/36
9	4/36	30/36
10	3/36	33/36
11	2/36	35/36
12	1/36	36/36=1
>12	0	1
	$\Sigma=1$	





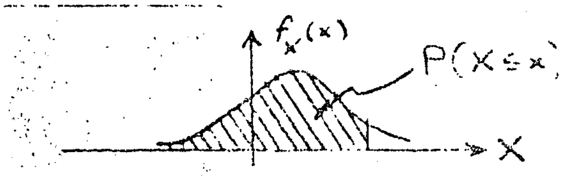
EN EL CASO DE UNA VARIABLE ALEATORIA CONTINUA, X, LA PROBABILIDAD DE QUE ESTA TOME UN VALOR COMPRENDIDO ENTRE x Y x + dx ESTA DADA POR $f_X(x)dx$, DONDE $f_X(x)$ ES LA DENSIDAD DE PROBABILIDADES DE X. POR LO TANTO, LA PROBABILIDAD DE QUE X ASUMA VALORES COMPRENDIDOS EN EL INTERVALO $x_1 \leq X \leq x_2$ ES



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

LA INTERPRETACION GRAFICA DE ESTA PROBABILIDAD ES QUE CORRESPONDE AL AREA BAJO LA CURVA DE $f_X(x)$ COMPRENDIDA ENTRE x_1 Y x_2.

PUESTO QUE $F_X(x) = P(X \leq x) = P(-\infty \leq X \leq x)$, Y EN VIRTUD DE LA ECUACION ANTERIOR SE TIENE QUE LA FUNCION DE DISTRIBUCION ES:



$$F_X(x) = \int_{-\infty}^{x} f_X(U)\, dU$$

DONDE U ES SOLO UNA VARIABLE MUDA DE INTEGRACION. EL VALOR DE ESTA INTEGRAL ES IGUAL AL AREA BAJO LA CURVA DE $F_X(x)$ A LA IZQUIERDA DE x. DE ESTA ECUACION SE CONCLUYE QUE

$$\frac{dF_X(x)}{dx} = \frac{d}{dx}\left(\int_{-\infty}^{x} f_X(U)\, dU\right) = f_X(x)$$

ALGUNAS PROPIEDADES DE $F_X(x)$ SON:

$$0 \leq F_X(x) \leq 1$$

$$F_X(-\infty) = 0$$

$$F_X(\infty) = 1$$

$$F_X(x + \varepsilon) \geq F_X(x), \text{ SI } \varepsilon \geq 0$$

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

PARA SATISFACER LOS AXIOMAS DE LA TEORIA DE PROBABILIDADES SE NECESITA QUE

$$f_X(x) \geq 0 \text{ PARA TODA } x$$

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

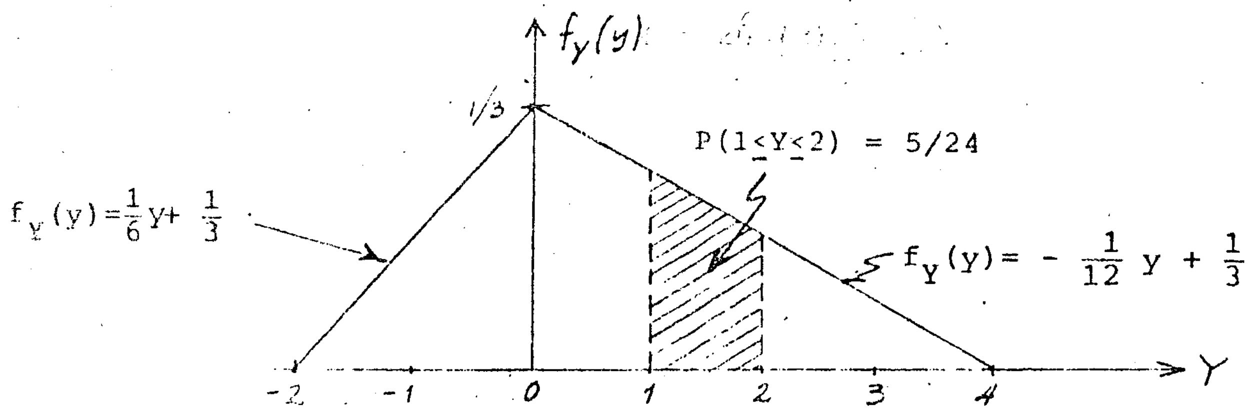
EJEMPLO

SEA UNA VARIABLE ALEATORIA CONTINUA CUYA DENSIDAD DE PROBABILIDADES ES DE FORMA TRIANGULAR DADA POR LAS SIGUIENTES ECUACIONES:

$$f_Y(y) = \frac{1}{6}\, y + \frac{1}{3}\,,\ \text{SI } -2 \le Y \le 0$$

$$f_Y(y) = -\frac{1}{12}\, y + \frac{1}{3}\,,\ \text{SI } 0 \le Y \le 4$$

$$f_Y(y) = 0 \qquad \text{SI } Y \le 2 \ \text{ O } \ Y \ge 4$$



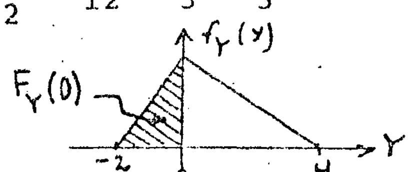
LA DISTRIBUCION DE PROBABILIDADES ACUMULADAS ES, ENTONCES:

SI $-2 \le Y \le 0$

$$F_Y(y) = \int_{-\infty}^{y} f_Y(U)\,dU = \int_{-2}^{y} \left(\frac{1}{6}\,U + \frac{1}{3}\right) dU$$

$$= \left[\frac{U^2}{12} + \frac{U}{3}\right]_{-2}^{y} = \frac{y^2}{12} + \frac{y}{3} + \frac{1}{3}$$

SI $0 \le Y \le 4$

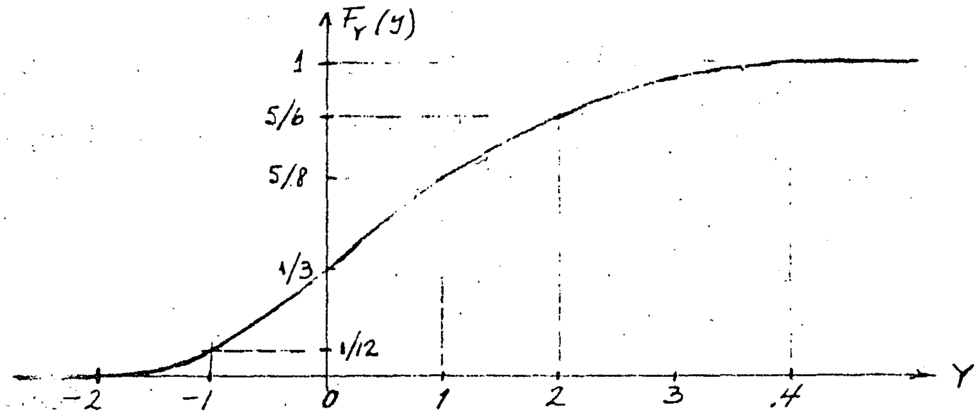


$$F_Y(y) = F_Y(0) + \int_0^{y} \left(-\frac{1}{12}\,U + \frac{1}{3}\right) dU = \frac{1}{3} + \left[\frac{-U^2}{24} + \frac{U}{3}\right]_0^{y} =$$

$$= \frac{1}{3} - \frac{y^2}{24} + \frac{y}{3} \qquad \text{SI } 0 \le y \le 4$$

$F_Y(y) = 0 \qquad$ SI $\quad y \leq -2$

$F_Y(y) = 1 \qquad$ SI $\quad y \geq 4$



SI SE DESEA CALCULAR LA PROBABILIDAD DE QUE AL REALIZAR UNA VEZ EL EXPERIMENTO QUE INVOLUCRA A DICHA VARIABLE, EL VALOR QUE SE OBSERVE CAIGA EN EL INTERVALO $1 \leq Y \leq 2$, ENTONCES

$$P[1 \leq Y \leq 2] = \int_1^2 \left(-\frac{1}{12}y + \frac{1}{3}\right) dy = \left[-\frac{y^2}{24} + \frac{y}{3}\right]_1^2 = \frac{5}{24}$$

O

$$P[1 \leq Y \leq 2] = F_Y(2) - F_Y(1) = \frac{5}{6} - \frac{5}{8} = \frac{5}{24}$$

EJEMPLO

UN INGENIERO ESTA INTERESADO EN DISEÑAR UNA TORRE QUE RESISTA LAS CARGAS DEBIDAS AL VIENTO. DE UNA SERIE DE OBSERVACIONES DE LA MAXIMA VELOCIDAD ANUAL DEL VIENTO CERCA DEL SITIO DE INTERES, SE ENCUENTRA QUE EL HISTOGRAMA PUEDE AJUSTARSE RAZONABLEMENTE, DESDE UN PUNTO DE VISTA ESTADISTICO, MEDIANTE UNA *DISTRIBUCION DE PROBABILIDADES EXPONENCIAL* DE LA FORMA

$$f_X(x) = Ke^{-\lambda x} \; ; \; x \geqslant 0$$

DONDE X ES LA MAXIMA VELOCIDAD DEL VIENTO, λ ES UNA CONSTANTE Y K ES OTRA CONSTANTE TAL QUE OBLIGA A QUE EL AREA BAJO LA CURVA DE $f_X(X)$ SEA IGUAL A UNO. POR TANTO,

$$\int_0^\infty Ke^{-\lambda x}\, dx = \frac{-K}{\lambda}\,[e^{-\lambda x}]_0^\infty = \frac{K}{\lambda} = 1$$

DE DONDE

$$K = \lambda$$

POR TANTO

$$f_X(x) = \lambda e^{-\lambda x} \; ; \; x \geqslant 0$$

LA FUNCION DE DISTRIBUCION SERA

$$F_X(x) = \int_0^x f_X(u)\, du = \int_0^x \lambda e^{-\lambda u}\, du = [-e^{-\lambda u}]_0^x = 1 - e^{-\lambda x} \; ; \; x \geqslant 0$$

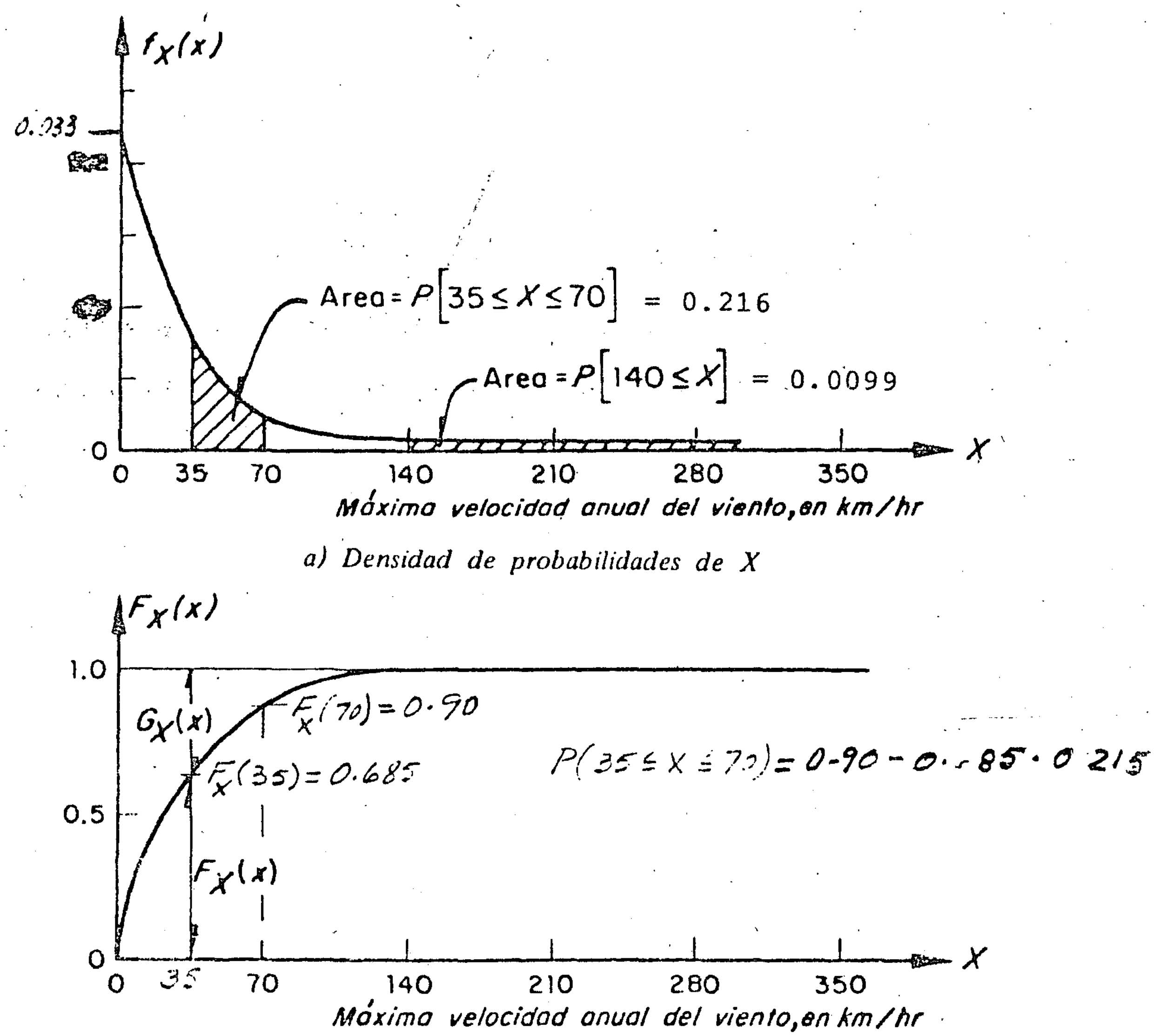
EL VALOR DE λ SE PUEDE TOMAR, POR EJEMPLO, DE MANERA QUE $F_X(x)$ SE AJUSTE PARA QUE COINCIDA CON UN VALOR EMPIRICO. ASI, SI LA FRECUENCIA RELATIVA DEL EVENTO $A = \{X \leq 70 \text{ KM/H}\}$ ES 0.9, ENTONCES

$$P(0 \leq X \leq 70) = F_X(70) = 0.9$$

DE DONDE

$$0.9 = 1 - e^{-70\lambda}$$

POR LO CUAL $\lambda = 0.033$.



b) Función de distribución de X

Ley de probabilidades correspondiente al ejemplo de la máxima velocidad anual del viento

SI SE DESEA CALCULAR, POR EJEMPLO, LA PROBABILIDAD DE QUE LA VELOCIDAD MAXIMA DEL VIENTO EN UN AÑO DADO ESTE ENTRE 35 Y 70 KM/H, SE TENDRA:

$$P(35 \leq X \leq 70) = \int_{35}^{70} 0.033e^{-0.033x}dx = \left[-e^{-0.033x}\right]_{35}^{70} =$$

$$= -e^{-0.033x70} - (-e^{-0.033x35}) = -e^{-2.31} + e^{-1.155} =$$

$$= -0.099 + 0.315 = 0.216$$

EN TERMINOS DE $F_X(x)$ ESTA PROBABILIDAD QUEDA DADA POR

$$P(35 \leq X \leq 70) = F_X(70) - F_X(35) = 0.90 - (1 - e^{-1.155}) = 0.90 - 0.685$$

$$= 0.215$$

FUNCION DE DISTRIBUCION COMPLEMENTARIA

EL COMPLEMENTO, $G_X(x)$, DE LA DISTRIBUCION DE PROBABILIDADES ACUMULADAS SE UTILIZA CUANDO LAS DECISIONES SE TOMAN CON BASE EN PROBABILIDADES DE QUE SE EXCEDA UN VALOR DADO DE LA VARIABLE. LA *FUNCION DE DISTRIBUCION COMPLEMENTARIA* SE DEFINE COMO

$$G_X(x) = P(X>x) = 1 - F_X(x)$$

EJEMPLO

PARA EL PROBLEMA ANTERIOR DE LA VELOCIDAD MAXIMA ANUAL DEL VIENTO, CALCULEMOS LA PROBABILIDAD DE QUE ESTA SEA MAYOR DE 140 KM/H:

$$G_X(140) = P(X \geq 140) = \int_{140}^{\infty} 0.033e^{-0.033x}dx = 0.0099$$

O, ALTERNATIVAMENTE

$$P(X \geq 140) = 1-F_X(140) = G_X(140) = 1-(1-e^{-0.033x140}) = e^{-4.62} = 0.0099$$

ESPERANZAS

LA ESPERANZA DE UNA FUNCION $g(X)$, DE UNA VARIABLE ALEATORIA DISCRETA, X, ES, POR DEFINICION

$$E(g(X)) = \sum_{i=1}^{i=n} g(x_i)P_X(x_i)$$

O PARA UNA VARIABLE CONTINUA

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f_X(x)\,dx$$

EJEMPLOS

1. SI $g(X)$ = CONSTANTE = c

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

2. SI $g(X) = cx$

$$E[cx] = c\overbrace{\int_{-\infty}^{\infty} x f_X(x)\,dx}^{E[X]} = cE[X]$$

3. SI $g(X) = a + bx$

$$E[a+bx] = a\int_{-\infty}^{\infty} f_X(x)\,dx + b\int_{-\infty}^{\infty} x f_X(x)\,dx = a + bE[X]$$

4. SI $g(X) = g_1(X) + g_2(X)$

$$E[g_1(X) + g_2(X)] = \int_{-\infty}^{\infty} g_1(x) f_X(x)\,dx + \int_{-\infty}^{\infty} g_2(x) f_X(x)\,dx$$

$$= E[g_1(X)] + E[g_2(X)]$$

EJEMPLO

SI X ES UNA VARIABLE ALEATORIA CON DENSIDAD DE PROBABILIDADES EXPONENCIAL, CALCULAR LA ESPERANZA DE LA FUNCION

$$g(X) = X^2$$

EN ESTE CASO SE TIENE QUE

$$f_X(x) = \lambda e^{-\lambda x} \quad \text{SI } 0 \le X \le \infty \text{ , Y} \quad f_X(x) = 0 \quad \text{SI } X<0$$

POR LO QUE

$$E(X^2) = E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx = \lambda \int_{-\infty}^{\infty} x^2 e^{-\lambda x} dx$$

$$= \lambda \left[\frac{-x^2 e^{-\lambda x}}{\lambda}\right]_0^{\infty} + \frac{2\lambda}{\lambda} \int_0^{\infty} x e^{-\lambda x} dx = \frac{-2}{\lambda^2} \left[e^{-\lambda x}(1+\lambda x)\right]_0^{\infty} = \frac{2}{\lambda^2}$$

EN GENERAL, A LA ESPERANZA DE X^2 SE LE DENOMINA *VALOR MEDIO CUADRATICO.*

Ejemplo. Construcción de la carpeta de una carretera.

Un contratista construirá la carpeta de una carretera en tramos de 50 m; el gobierno aceptará o rechazará cada tramo de acuerdo con una prueba de control de calidad. El contratista tiene la opción de pedir el concreto a una de dos plantas premezcladoras; la planta A cobra 140 pesos/m^3 y la B 160 pesos/m^3, pero, el control de calidad que se lleva en la planta B es mejor, lo cual hace más probable que un tramo dado pase favorablemente la prueba de aceptación. Tomando en cuenta que en cada tramo se usan 100 m^3 de concreto y que la probabilidad de que el proveniente de la planta A no pase la prueba de control es 0.10, y la de B es 0.05, el constructor deberá decidirse por cuál planta usar. El árbol de decisiones de este problema es el mostrado en la fig 6.4, donde $P(\theta_1)$ y $P(\theta_2)$ son las probabilidades de que ocurran θ_1 y θ_2, respectivamente. La utilidad $U_1 = u(a_1, \theta_1)$ es la que corresponde a utilizar la planta A y que la carpeta pase la prueba de control de calidad; en este caso la utilidad (negativa) es el costo del concreto ($14,000.00) más la colocación (supongamos $100,000.00), por lo cual $U_1 =$ -114,000.00. $U_2 = u(a_1, \theta_2)$ es la que corresponde a usar la planta A y que la carpeta no pase la prueba de calidad; en este caso el constructor deberá demoler y reconstruir el tramo con los siguientes costos:

Carpeta demolida	Pérdida de prestigio	$ 5,000.00
	Mano de obra de demolición	15,000.00
	Concreto	14,000.00
	Mano de obra de colocación	100,000.00

Reconstrucción	Mano de obra	\$ 100,000.00
	Concreto	14,000.00
	T O T A L	\$ 248,000.00

De manera similar se obtienen U_3 y U_4, cuyos valores resultan ser U_3 = - \$116,000.00 y U_4 = - \$252,000.00.

Si la decisión se tomara sin considerar las probabilidades de aceptar la carpeta, el constructor se decidiría por la planta A, ya que la pérdida (utilidad negativa) sería menor. Si sí se toman en cuenta y adoptamos como *criterio de decisión* el escoger la planta que conduzca a una *esperanza de pérdida* menor se tendrá (recuerde que la esperanza de la variable aleatoria X, $E[X]$, es $E[X] = \sum_{i=1}^{m} P[X_i]X_i$, donde las X_i son los valores que puede asumir X, y $P[X_i]$ son las probabilidades correspondientes):

Para la planta A:

$$E[U] = 0.90 \times (-114{,}000) + 0.10 \times (-248{,}000) = -\$127{,}400.$$

Para la planta B:

$$E[U] = 0.95 \times (-116{,}000) + 0.05 \times (-252{,}000) = -\$122{,}800.$$

Comparando ambas cifras se concluye que la decisión de comprar el concreto de la planta B conduce a una pérdida esperada menor que la de la planta A, es decir, se escoge la planta B aunque el precio unitario del concreto sea mayor.

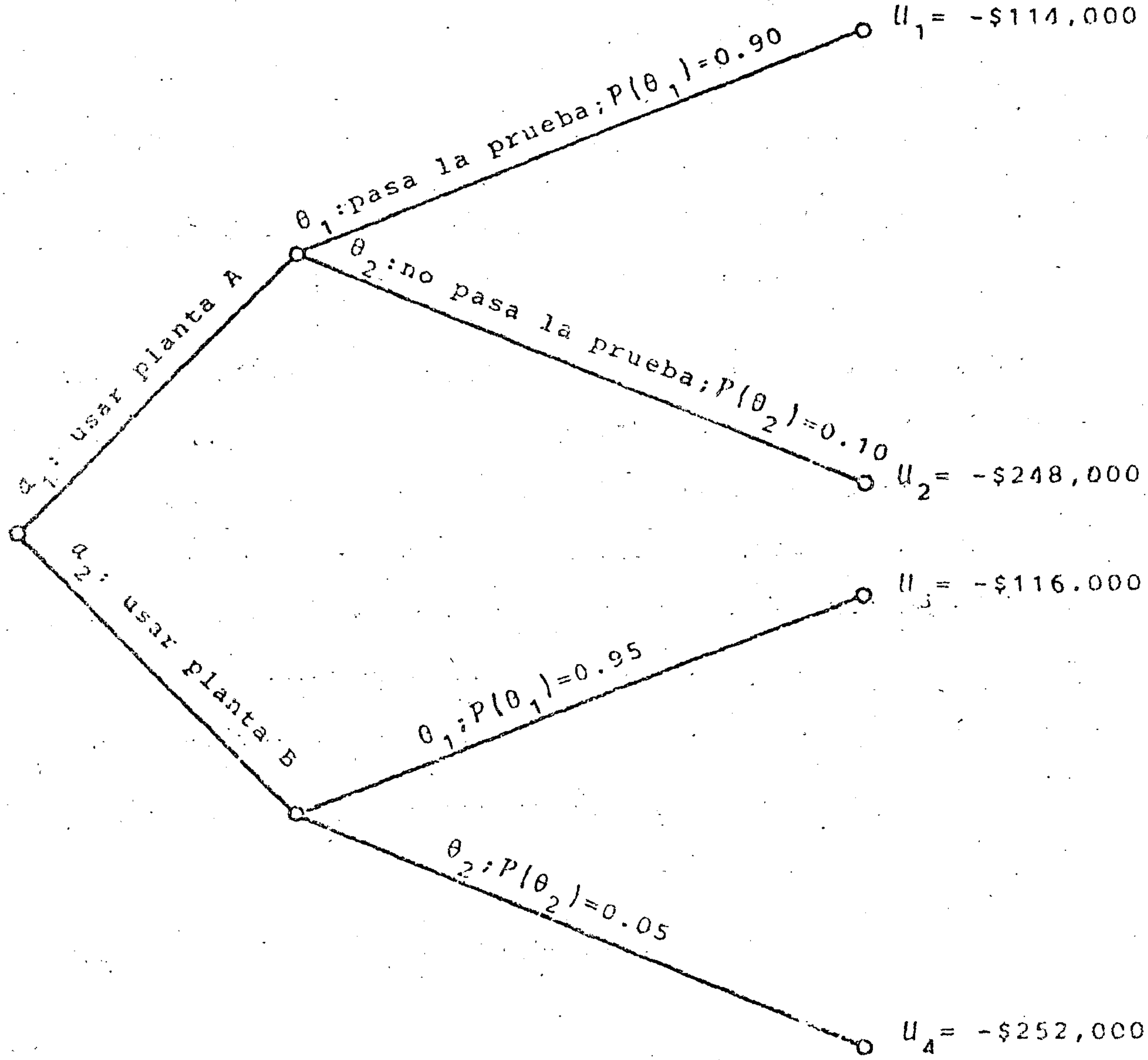


Fig 6.4 Arbol de decisiones del ejemplo 6.2

MEDIDAS DE TENDENCIA CENTRAL

LA *MEDIA O ESPERANZA*, $E[X]$, DE UNA VARIABLE ALEATORIA, X, SE CALCULA CON LAS ECUACIONES ANTERIORES PARA EL CASO EN QUE $g(X)=X$. DE ESTA MANERA, SI LA VARIABLE ES DISCRETA, SU ESPERANZA QUEDA DADA POR

$$E(X) = \sum_{i=1}^{i=n} x_i \, P_X \, (x_i)$$

DONDE n ES EL TOTAL DE VALORES QUE X PUEDE ASUMIR.

PARA EL CASO DE UNA VARIABLE ALEATORIA CONTINUA, LA MEDIA ES

$$m_X = E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx$$

OTRAS MEDIDAS USUALES DE TENDENCIA CENTRAL DE UNA VARIABLE ALEATORIA SON LA *MEDIANA* Y EL *MODO*. LA PRIMERA SE DEFINE COMO EL VALOR DE LA VARIABLE AL CUAL CORRESPONDE UNA PROBABILIDAD ACUMULADA DE 50%. Y LA SEGUNDA, COMO EL VALOR DE LA VARIABLE AL CUAL CORRESPONDE LA MAYOR PROBABILIDAD.

EJEMPLO

SI LA DENSIDAD DE PROBABILIDADES DE LA VARIABLE ALEATORIA X CORRESPONDE A LOS ERRORES EN UNA NIVELACIÓN, ES LA DE LA SEGUNDA COLUMNA DE LA SIGUIENTE TABLA, LA MEDIA DE DICHA VARIABLE RESULTA SER 4 167 LA MEDIANA 4000 Y EL MODO 4000 MICRAS. LOS CÁLCULOS CORRESPONDIENTES SE LOCALIZAN EN LA TERCERA COLUMNA.

x_i, EN MICRAS	$P_X(x_i)$	$x_i P_X(x_i)$, EN MICRAS	$F_X(x_i)$
0	6/60	0	6/60
1 000	2/60	2 000/60	8/60
2 000	4/60	8 000/60	12/60
3 000	8/60	24 000/60	20/60
4 000	13/60	52 000/60	33/60≥0.5
5 000	12/60	60 000/60	45/60
6 000	7/60	42 000/60	52/60
7 000	4/60	28 000/60	56/60
8 000	2/60	16 000/60	58/60
9 000	2/60	18 000/60	60/60
TOTAL: $E[X]$ = 250 000/60=4 167 MICRAS			

EJEMPLO

CALCULAR LA ESPERANZA DE UNA VARIABLE ALEATORIA CUYA DENSIDAD DE PROBABILIDADES ES TRIANGULAR DADA POR

$$f_Y(y) = \frac{1}{6}y + \frac{1}{3} \qquad \text{SI } -2 \leq y \leq 0$$

$$f_Y(y) = \frac{-1}{12}y + \frac{1}{3} \qquad \text{SI } 0 \leq y \leq 4$$

$$f_Y(y) = 0 \qquad \text{SI } y \leq -2 \text{ O } y \geq 4$$

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)\,dy = \int_{-2}^{0} y\left(\frac{y}{6} + \frac{1}{3}\right)dy + \int_{0}^{4} y\left(\frac{-y}{12} + \frac{1}{3}\right)dy$$

$$= \left[\frac{y^3}{18} + \frac{y^2}{6}\right]_{-2}^{0} + \left[\frac{-y^3}{36} + \frac{y^2}{6}\right]_{0}^{4} = \frac{2}{3}$$

EJEMPLO

CALCULAR LA ESPERANZA DE UNA VARIABLE ALEATORIA CON DENSIDAD DE PROBABILIDADES EXPONENCIAL

$$f_x(x) = \lambda e^{-\lambda x}$$

$$E(X) = \int_{-\infty}^{\infty} x f_x(x)\,dx = \lambda \int_{0}^{\infty} x e^{-\lambda x} dx = \lambda \left[\frac{-e^{-\lambda x}}{\lambda^2} (1+\lambda x) \right]_0^{\infty} = \frac{1}{\lambda}$$

MEDIDAS DE DISPERSION

UNA MEDIDA MUY COMUN DE LA DISPERSION O VARIABILIDAD DE LOS VALORES QUE PUEDE ASUMIR UNA VARIABLE ALEATORIA ES LA *VARIANCIA*, LA CUAL SE DENOTA COMO $\sigma^2(X)$ O VAR (X), LA CUAL SE DEFINE COMO LA ESPERANZA DE LA FUNCION $g(X) = [X-E(X)]^2$. ASI, PARA UNA VARIABLE ALEATORIA DISCRETA

$$\sigma^2(X) = VAR(X) = \sum_{i=1}^{i=n} (x_i - E(X))^2 \, P_X(x_i)$$

Y PARA UNA CONTINUA

$$\sigma^2(X) = VAR(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f_X(x)\,dx$$

DESARROLLANDO EL INTEGRANDO DE ESTA ULTIMA ECUACION:

$$\sigma^2(X) = \int_{-\infty}^{\infty} (x^2 - 2xE(X) + E^2(X)) f_X(x)\,dx$$

$$= \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - 2E(X)\int_{-\infty}^{\infty} x f_X(x)\,dx + E^2(X)\int_{-\infty}^{\infty} f_X(x)\,dx = E[X^2] - E^2[X]$$

ES DECIR, LA VARIANCIA SE PUEDE CALCULAR COMO LA DIFERENCIA DEL VALOR MEDIO CUADRATICO Y EL CUADRADO DE LA MEDIA DE X.

OTRAS MEDIDAS DE DISPERSION DE LA VARIABLE ALEATORIA X SON LA *DESVIACION ESTANDAR*, $\sigma(X)$, LA CUAL ES IGUAL A LA RAIZ CUADRADA DE LA VARIANCIA, Y EL *COEFICIENTE DE VARIACION* QUE SE DEFINE COMO

$$\nu(X) = \sigma(X)/E(X) \quad , \text{ SI } E(X) \neq 0$$

EJEMPLO

EN LA SIGUIENTE TABLA SE CALCULA LA VARIANCIA DE LA VARIABLE ALEATORIA CUYA DENSIDAD DE PROBABILIDADES SE PRESENTO EN EL EJEMPLO ANTERIOR ($E(x) = 4167$ micras).

$x_i - E(X)$ EN MICRAS	$(x_i - E(x))^2$	$P_X(x_i)$	$(x_i - E(X))^2 P_X(x_i)$, EN MICRAS
-4 167		6/60	1 740 000
-3 167		2/60	333 000
-2 167		4/60	313 000
-1 167		8/60	181 000
- 167		13/60	6 000
833		12/60	139 000
1 833		7/60	390 000
2 833		4/60	531 000
3 833		2/60	487 000
4 833		2/60	687 000

TOTAL: 4 798 000 MICRAS^2 $= \sigma^2(X)$

LA DESVIACION ESTANDAR Y EL COEFICIENTE DE VARIACION DE ESTA VARIABLE ALEATORIA SON, RESPECTIVAMENTE,

$$\sigma(X) = \sqrt{4\ 798\ 000} = 2\ 200 \text{ MICRAS, Y } \nu(X) = \sigma(X)/E(X) = \frac{2\ 200}{4\ 167} = 0.528$$

EJEMPLO

SI X ES UNA VARIABLE ALEATORIA CON DISTRIBUCION DE PROBABILIDADES EXPONENCIAL, CALCULAR SU VARIANCIA, DESVIACION ESTANDAR Y COEFICIENTE DE VARIACION:

$$\sigma^2(X) = E(X-E[X])^2 = \int_{-\infty}^{\infty}(x-E[X])^2\lambda e^{-\lambda x}dx = \lambda \int_0^{\infty}(x^2-2xE[X]+E^2[X])e^{-\lambda x}dx$$

$$= \lambda \int_0^{\infty}x^2 e^{-\lambda x}dx - 2E[X]\lambda\int_0^{\infty}xe^{-\lambda x}dx + E^2[X]\int_0^{\infty}\lambda e^{-\lambda x}dx$$

$$= \frac{2}{\lambda^2} - 2\,\frac{1}{\lambda}\,\frac{1}{\lambda} + \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

YA QUE $E(X) = 1/\lambda$ Y $E[X^2] = 2/\lambda^2$.

USANDO LA FORMULA $\sigma^2(X) = E[X^2] - E^2[X]$, Y TOMANDO EN CUENTA QUE $E[X^2] = 2/\lambda^2$ SE OBTIENE:

$$\sigma^2(X) = 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2$$

EN CONSECUENCIA, LA DESVIACION ESTANDAR ES

$$\sigma(X) = \sqrt{1/\lambda^2} = 1/\lambda$$

Y EL COEFICIENTE DE VARIACION

$$\nu(X) = \sigma(X)/E(X) = \frac{\frac{1}{\lambda}}{\frac{1}{\lambda}} = 1$$

EJEMPLO

SEA Y UNA VARIABLE ALEATORIA CON DENSIDAD DE PROBABILIDADES TRIANGULAR DADA POR

$$f_Y(y) = \frac{1}{6}y + \frac{1}{3} \qquad \text{SI} \quad -2 \leq y \leq 0$$

$$f_Y(y) = \frac{-1}{12}y + \frac{1}{3} \qquad \text{SI} \quad 0 \leq y \leq 4$$

$$f_Y(y) = 0 \qquad \text{SI} \quad y \leq -2 \text{ O } y \geq 4$$

CALCULAR LA VARIANCIA, LA DESVIACION ESTANDAR Y EL COEFICIENTE DE VARIACION.

CALCULAREMOS PRIMERO EL VALOR MEDIO CUADRATICO PARA LUEGO APLICAR LA ECUACION $\sigma^2(Y) = E(Y^2) - E^2(Y)$

$$E[Y^2] = \int_{-2}^{0} y^2\left(\frac{1}{6}y + \frac{1}{3}\right)dy + \int_{0}^{4} y^2\left(-\frac{y}{12} + \frac{1}{3}\right)dy = \left[\frac{y^4}{24} + \frac{y^3}{9}\right]_{-2}^{0} + \left[\frac{-y^4}{48} + \frac{y^3}{9}\right]_{0}^{4} = 2$$

$$\sigma^2(Y) = 2-(2/3)^2 = 14/9$$

$$\sigma(Y) = 1.25 \quad \left(\sqrt{14/9}\right)$$

$$\nu(Y) = 1.25/(2/3) = 1.88$$

DISTRIBUCIONES PARTICULARES

VARIABLES ALEATORIAS DISCRETAS

DISTRIBUCION BINOMIAL O DE BERNOULLI

LA DISTRIBUCION BINOMIAL O DE BERNOULLI SE EMPLEA COMO DENSIDAD DE PROBABILIDADES DE VARIABLES ALEATORIAS DISCRETAS ASOCIADOS A EXPERIMENTOS EN LOS QUE SOLO HAY (O SOLO IMPORTAN) DOS RESULTADOS POSIBLES, UNO DE LOS CUALES USUALMENTE SE DENOMINA "EXITO" Y, EL OTRO, "FRACASO". $(S = \{\text{éxito}, \text{fracaso}\})$

SEAN p= PROBABILIDAD DE OBSERVAR "EXITO" AL REALIZAR UNA VEZ EL EXPERIMENTO

q= PROBABILIDAD DE "FRACASO" = $1-p$

X= VARIABLE ALEATORIA "NUMERO DE EXITOS OBSERVADOS AL REPETIR n VECES EL EXPERIMENTO "CON REEMPLAZO"

LA DISTRIBUCION DE PROBABILIDADES BINOMIAL ES

$$f(x) = \frac{n!}{x!\,(n-x)!}\, p^x q^{n-x} \quad : \quad x = 0, 1, \ldots, n$$

SE PUEDE DEMOSTRAR QUE LOS PARAMETROS DE ESTA DISTRIBUCION SON

$$E(X) = np \;, \quad \sigma^2(X) = npq$$

SI n=2, ENTONCES X PUEDE ASUMIR LOS VALORES 0, 1 y 2, ES DECIR $S=\{0,1,2\}$. EL ESPACIO DE EVENTOS DEL EXPERIMENTO ES

$$S_1 = \{\underbrace{(\text{FRACASO, FRACASO})}_{X=0}, \underbrace{(\text{EXITO, FRACASO})}_{X=1}, \underbrace{(\text{FRACASO,EXITO})}_{X=1}, \underbrace{(\text{EXITO, EXITO})}_{X=2}\}$$

$$f_X(x) = \{P(0), P(1), P(2)\}$$

OBSERVESE QUE $x=0$ OCURRE DE UNA MANERA, $x=1$, DE DOS, Y $x=2$, DE UNA. ESTOS RESULTADOS SE PUEDEN OBTENER PERMUTANDO DOS GRUPOS, UNO CON x Y EL OTRO CON $n-x$ ELEMENTOS:

$$x=0: \quad {}_2P_{0,2} = \frac{2!}{0!\times 2!} = 1 \ ; \ P(\{0\}) = q \times q = q^2 = p^0q^2$$

$$x=1: \quad {}_2P_{1,1} = \frac{2!}{1!\times 1!} = 2 \ ; \ P(\{1\}) = 2pq$$

$$x=2: \quad {}_2P_{2,0} = \frac{2!}{2!\times 0!} = 1 \ ; \ P(\{2\}) = p\times p = p^2 = p^2q^0$$

$$\sum_{i=0}^{2} P(\{i\}) = q^2+2pq+p^2 = (p+q)^2 = 1$$

(OBSERVESE QUE LOS ELEMENTOS DE S_1 NO SON IGUALMENTE PROBABLES, A MENOS QUE $p = q = 1/2$.)

SI $n = 3$, $S = \{0,1,2,3\}$, e = EXITO Y f = FRACASO, ENTONCES

$$S_1 = \{(f,f,f),\ (e,f,f),\ (f,e,f),\ (f,f,e),\ (e,e,f),\ (e,f,e),\ (f,e,e),\ (e,e,e)\}$$

$$x = 0:\ {}_3P_{0,3} = \frac{3!}{0!\times 3!} = 1\ ;\ P(\{0\}) = 1\ p^0 q^3 = q^3$$

$$x = 1:\ {}_3P_{1,2} = \frac{3!}{1!\times 2!} = 3\ ;\ P(\{1\}) = 3\ pq^2$$

$$x = 2:\ {}_3P_{2,1} = \frac{3!}{2!\times 1!} = 3\ ;\ P(\{2\}) = 3p^2q$$

$$x = 3:\ {}_3P_{3,0} = \frac{3!}{3!\times 0!} = 1\ ;\ P(\{3\}) = 1p^3q^0 = p^3$$

$$\sum_{i=0}^{3} P(\{i\}) = (p+q)^3 = 1$$

PASANDO AL CASO GENERAL DE CUALQUIER VALOR DE n, LA PROBABILIDAD DE QUE OCURRAN x EXITOS Y $n-x$ FRACASOS *EN UN ORDEN DETERMINADO* ES

$$P(X=x) = p^x q^{n-x}$$

EN VIRTUD DE LA LEY GENERAL DE MULTIPLICACION.

UN ORDEN POSIBLE SERIA, POR EJEMPLO,

$$\underbrace{\text{EXITO, EXITO,} \ldots \text{, EXITO}}_{x},\ \underbrace{\text{FRACASO,} \ldots \text{, FRACASO}}_{n-x}$$

AHORA BIEN, LOS x EXITOS PUEDEN OCURRIR EN ${}_nP_{x,n-x}$ ORDENES DISTINTOS, CADA UNO CON PROBABILIDAD $p^x q^{n-x}$. POR LO TANTO, EN VIRTUD DE LA LEY GENERAL DE ADICION, LA DISTRIBUCION DE PROBABILIDADES DE X, RESULTA SER

$$f_X(x) = \frac{n!}{x!\,(n-x)!}\, p^x q^{n-x}\ ;\ x = 0,1,\ldots,n$$

LA CUAL SE CONOCE CON EL NOMBRE DE *BINOMIAL O DE BERNOULLI*.

LA ESPERANZA DE LA DISTRIBUCION DE BERNOULLI ES

$$E(X) = \sum_{x=0}^{n} x \frac{n!}{x!\,(n-x)!} p^x q^{n-x} = \sum_{x=1}^{n} x \frac{n!}{x!\,(n-x)!} p^x q^{n-x}$$

$$= np \underbrace{\sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!\,(n-x)!} p^{x-1} q^{n-x}}_{(p+q)^{n-1}} = np(p+q)^{n-1} = np$$

LA VARIANCIA DE LA DISTRIBUCION BINOMIAL ES

$$\sigma^2(X) = E\left[(X-E(X))^2\right] = E\left[(X-np)^2\right]$$

PERO $E\left[(X-np)^2\right] = E(X^2-2npX + n^2p^2) = E\left[X + X(X-1) - 2npX + n^2p^2\right]$

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

$$= (1-2np)np + n^2p^2 + \sum_{x=0}^{n} \frac{n!}{x!\,(n-x)!} p^x q^{n-x} x(x-1)$$

$$= np-n^2p^2 + \sum_{x=2}^{n} n(n-1) \frac{(n-2)!}{(x-2)!\,(n-x)!} p^2 p^{x-2} q^{n-x}$$

$$= np-n^2p^2 + n(n-1)p^2 \underbrace{\sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!\,(n-x)!} p^{x-2} q^{n-x}}_{(p+q)^{n-2}}$$

$$= np-n^2p^2 + n(n-1)p^2(p+q)^{n-2} = np-np^2 = np(1-p) = npq$$

EN RESUMEN, PARA LA DISTRIBUCION BINOMIAL,

$$E(X) = np \quad ; \quad \sigma^2(X) = npq \; ; \; \sigma(X) = \sqrt{npq}$$

TABLA 1

FUNCION DE DISTRIBUCION BINOMIAL

$$F_X(x) = \sum_{k=0}^{x} \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

n	x	p 0.10	0.20	0.30	0.40	0.50
2	0	.8100	.6400	.4900	.3600	.2500
	1	.9900	.9600	.9100	.8400	.7500
3	0	.7290	.5120	.3430	2160	.1250
	1	.9720	.8960	.7840	.6480	.5000
	2	.9990	.9920	.9730	.9360	.8750
4	0	.6561	.4096	.2401	.1296	.0625
	1	.9477	.8192	.6517	.4752	.3125
	2	.9963	.9728	9163	.8208	.6875
	3	.9999	.9984	.9919	.9744	.9375
5	0	.5905	.3277	.1681	.0778	.0312
	1	.9185	.7373	.5282	.3370	.1875
	2	.9914	.9421	.8369	.6826	.5000
	3	.9995	.9933	.9692	.9130	.8125
	4	1.0000	.9997	.9976	.9898	.9688
6	0	.5314	.2621	.1176	.0467	.0156
	1	.8857	.6554	.4202	.2333	.1094

TABLA 1 (continuación)

FUNCION ~~DE DISTRIBUCION~~ BINOMIAL

$$F_X(x) = \sum_{k=0}^{x} \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

n	x	p 0.10	0.20	0.30	0.40	0.50
6	2	.9842	.9011	.7443	.5443	.3438
	3	.9987	.9830	.9295	.8208	.6562
	4	.9999	.9984	.9891	.9590	.8906
	5	1.0000	.9999	.9993	.9959	.9844
7	0	.4783	.2097	.0824	.0280	.0078
	1	.8503	.5767	.3294	.1586	.0625
	2	.9743	.8520	.6471	.4199	.2266
	3	.9973	.9667	.8740	.7102	.5000
	4	.9998	.9953	.9712	.9037	.7734
	5	1.0000	.9996	.9962	.9812	.9375
	6	1.0000	1.0000	.9998	.9984	.9922
8	0	.4305	.1678	.0576	.0168	.0039
	1	.8131	.5033	.2553	.1064	.0352
	2	.9619	.7969	.5518	.3154	.1445
	3	.9950	.9437	.8059	.5941	.3633
	4	.9996	.9896	.9420	.8263	.6367
	5	1.0000	.9988	.9887	.9502	.8555
	6	1.0000	.9999	.9987	.9915	.9648
	7	1.0000	1.0000	.9999	.9993	.9961
9	0	.3874	.1342	.0404	.0101	.0020
	1	.7748	.4362	.1960	.0705	.0195
	2	.9470	.7382	.4628	.2318	.0898
	3	.9917	.9144	.7297	.4826	.2539
	4	.9991	.9804	.9012	.7334	.5000
	5	.9999	.9969	.9747	.9006	.7461
	6	1.0000	.9997	.9957	.9750	.9102
	7	1.0000	1.0000	.9996	.9962	.9805
	8	1.0000	1.0000	1.0000	.9997	.9980
10	0	.3487	.1074	.0282	.0060	.0010
	1	.7361	.3758	.1493	.0464	.0107
	2	.9298	.6778	.3828	.1673	.0547
	3	.9872	.8791	.6496	.3823	.1719
	4	.9984	.9672	.8497	.6331	.3770

TABLA 1 (continuación)

FUNCION DE DISTRIBUCION BINOMIAL

$$F_X(x) = \sum_{k=0}^{x} \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

n	x	p 0.10	0.20	0.30	0.40	0.50
13	8	1.0000	.9998	.9960	.9679	.8666
	9	1.0000	1.0000	.9993	.9922	.9539
	10	1.0000	1.0000	.9999	.9987	.9888
	11	1.0000	1.0000	1.0000	.9999	.9983
	12	1.0000	1.0000	1.0000	1.0000	.9999
14	0	.2288	.0440	.0068	.0008	.0001
	1	.5846	.1979	.0475	.0081	.0009
	2	.8416	.4481	.1608	.0398	.0065
	3	.9559	.6982	.3552	.1243	.0287
	4	.9908	.8702	.5842	.2793	.0898
	5	.9985	.9561	.7805	.4859	.2120
	6	.9998	.9884	.9067	.6925	.3953
	7	1.0000	.9976	.9685	.8499	.6047
	8	1.0000	.9996	.9917	.9417	.7880
	9	1.0000	1.0000	.9983	.9825	.9102
	10	1.0000	1.0000	.9998	.9961	.9713
	11	1.0000	1.0000	1.0000	.9994	.9935
	12	1.0000	1.0000	1.0000	.9999	.9991
	13	1.0000	1.0000	1.0000	1.0000	.9999
15	0	.2059	.0352	.0047	.0005	.0000
	1	.5490	.1671	.0353	.0052	.0005
	2	.8159	.3980	.1268	.0271	.0037
	3	.9444	.6482	.2969	.0905	.0176
	4	.9873	.8358	.5155	.2173	.0592
	5	.9978	.9389	.7216	.4032	.1509
	6	.9997	.9819	.8689	.6098	.3036
	7	1.0000	.9958	.9500	.7869	.5000
	8	1.0000	.9992	.9848	.9050	.6964
	9	1.0000	.9999	.9963	.9662	.8491
	10	1.0000	1.0000	.9993	.9907	.9408
	11	1.0000	1.0000	.9999	.9981	.9824
	12	1.0000	1.0000	1.0000	.9997	.9963
	13	1.0000	1.0000	1.0000	1.0000	.9995
	14	1.0000	1.0000	1.0000	1.0000	1.0000

TABLA 1 (continuación)

FUNCION DE DISTRIBUCION BINOMIAL

$$F_X(x) = \sum_{k=0}^{x} \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

n	X	p 0.10	0.20	0.30	0.40	0.50
10	5	.9999	.9936	.9527	.8338	.6230
	6	1.0000	.9991	.9894	.9452	.8281
	7	1.0000	.9999	.9984	.9877	.9453
	8	1.0000	1.0000	.9999	.9983	.9893
	9	1.0000	1.0000	1.0000	.9999	.9990
11	0	.3138	.0859	.0198	.0036	.0005
	1	.6974	.3221	.1130	.0302	.0059
	2	.9104	.6174	.3127	.1189	.0327
	3	.9815	.8389	.5696	.2963	.1133
	4	.9972	.9496	.7897	.5328	.2744
	5	.9997	.9883	.9218	.7535	.5000
	6	1.0000	.9980	.9784	.9006	.7256
	7	1.0000	.9998	.9957	.9707	.8867
	8	1.0000	1.0000	.9994	.9941	.9673
	9	1.0000	1.0000	1.0000	.9993	.9941
	10	1.0000	1.0000	1.0000	1.0000	.9995
12	0	.2824	.0687	.0138	.0022	.0002
	1	.6590	.2749	.0850	.0196	.0032
	2	.8891	.5583	.2528	.0834	.0193
	3	.9744	.7946	.4925	.2253	.0730
	4	.9957	.9274	.7237	.4382	.1938
	5	.9995	.9806	.8822	.6652	.3872
	6	.9999	.9961	.9614	.8418	.6128
	7	1.0000	.9994	.9905	.9427	.8062
	8	1.0000	.9999	.9983	.9847	.9270
	9	1.0000	1.0000	.9998	.9972	.9807
	10	1.0000	1.0000	1.0000	.9997	.9968
	11	1.0000	1.0000	1.0000	1.0000	.9998
13	0	.2542	.0550	.0097	.0013	.0001
	1	.6213	.2336	.0637	.0126	.0017
	2	.8661	.5017	.2025	.0579	.0112
	3	.9658	.7473	.4206	.1686	.0461
	4	.9935	.9009	.6543	.3530	.1334
	5	.9991	.9700	.8346	.5744	.2905
	6	.9999	.9930	.9376	.7712	.5000
	7	1.0000	.9988	.9818	.9023	.7095

TABLA 1 (continuación)

FUNCION DE DISTRIBUCION BINOMIAL

$$F_X(x) = \sum_{k=0}^{x} \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

				p		
n	x	0.10	0.20	0.30	0.40	0.50
18	0	.1501	.0180	.0016	.0001	.0000
	1	.4503	.0991	.0142	.0013	.0001
	2	.7338	.2713	.0600	.0082	.0007
	3	.9018	.5010	.1646	.0328	.0038
	4	.9718	.7164	.3327	.0942	.0154
	5	.9936	.8671	.5344	.2088	.0481
	6	.9988	.9487	.7217	.3743	.1189
	7	.9998	.9837	.8593	.5634	.2403
	8	1.0000	.9957	.9404	.7368	.4073
	9	1.0000	.9991	.9790	.8653	.5927
	10	1.0000	.9998	.9939	.9424	.7597
	11	1.0000	1.0000	.9986	.9797	.8811
	12	1.0000	1.0000	.9997	.9942	.9519
	13	1.0000	1.0000	1.0000	.9987	.9846
	14	1.0000	1.0000	1.0000	.9998	.9962
	15	1.0000	1.0000	1.0000	1.0000	.9993
	16	1.0000	1.0000	1.0000	1.0000	.9999
	17	1.0000	1.0000	1.0000	1.0000	1.0000
19	0	.1351	.0144	.0011	.0001	.0000
	1	.4203	.0829	.0104	.0008	.0000
	2	.7054	.2369	.0462	.0055	.0004
	3	.8850	.4551	.1332	.0230	.0022
	4	.9648	.6733	.2822	.0696	.0096
	5	.9914	.8369	.4739	.1629	.0318
	6	.9983	.9324	.6655	.3081	.0835
	7	.9997	.9767	.8180	.4878	.1796
	8	1.0000	.9933	.9161	.6675	.3238
	9	1.0000	.9984	.9674	.8139	.5000
	10	1.0000	.9997	.9895	.9115	.6762
	11	1.0000	1.0000	.9972	.9648	.8204
	12	1.0000	1.0000	.9994	.9884	.9165
	13	1.0000	1.0000	.9999	.9969	.9682
	14	1.0000	1.0000	1.0000	.9994	.9904
	15	1.0000	1.0000	1.0000	.9999	.9978
	16	1.0000	1.0000	1.0000	1.0000	.9996
	17	1.0000	1.0000	1.0000	1.0000	1.0000

TABLA 1 (continuación)

FUNCION DE DISTRIBUCION BINOMIAL

$$F_X(x) = \sum_{k=0}^{x} \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

n	x	p 0.10	0.20	0.30	0.40	0.50
20	0	.1216	.0115	.0008	.0000	.0000
	1	.3917	.0692	.0076	.0005	.0000
	2	.6769	.2061	.0355	.0036	.0002
	3	.8670	.4114	.1071	.0160	.0013
	4	.9568	.6296	.2375	.0510	.0059
	5	.9887	.8042	.4164	.1256	.0207
	6	.9976	.9133	.6080	.2500	.0577
	7	.9996	.9679	.7723	.4159	.1316
	8	.9999	.9900	.8867	.5956	.2517
	9	1.0000	.9974	.9520	.7553	.4119
	10	1.0000	.9994	.9829	.8725	.5881
	11	1.0000	.9999	.9949	.9435	.7483
	12	1.0000	1.0000	.9987	.9790	.8684
	13	1.0000	1.0000	.9997	.9935	.9423
	14	1.0000	1.0000	1.0000	.9984	.9793
	15	1.0000	1.0000	1.0000	.9997	.9941
	16	1.0000	1.0000	1.0000	1.0000	.9987
	17	1.0000	1.0000	1.0000	1.0000	.9998
	18	1.0000	1.0000	1.0000	1.0000	1.0000

EJEMPLO

SI SE LANZA AL AIRE SEIS VECES UNA MONEDA HOMOGENEA,

A) ¿CUAL ES LA PROBABILIDAD DE OBTENER DOS "CARAS"?

B) ¿CUAL ES LA PROBABILIDAD DE OBTENER POR LO MENOS CUATRO "CARAS" $(X \geq 4)$?

C) ¿CUANTO VALEN LA ESPERANZA Y LA DESVIACION ESTANDAR?

SOLUCION

A) PUESTO QUE LA MONEDA ES HOMOGENEA SE TIENE $p=1/2$ Y $q=1-1/2=1/2$, DONDE p ES LA PROBABILIDAD DE OBSERVAR "CARA" (CARA = EXITO) EN UN LANZAMIENTO. POR TANTO

$$P[X = 2] = f_x(2) = \frac{6!}{2!(6-2)!}\left(\frac{1}{2}\right)^2(1/2)^{6-2} = \frac{6!}{2!\ 4!}(1/2)^6 = \frac{15}{64} = 0.2344$$

B) PARA QUE SE CUMPLA $X \geq 4$ EN SEIS LANZAMIENTOS, SE NECESITA QUE SE OBSERVEN 4,5 o 6 CARAS. PUESTO QUE ESTOS TRES EVENTOS SON MUTUAMENTE EXCLUSIVOS, SE TIENE

$$P[X \geq 4] = f_x(4) + f_x(5) + f_x(6)$$

CALCULANDO LOS TRES SUMANDOS COMO EN LA PREGUNTA ANTERIOR, RESULTA

$$P[X \geq 4] = \frac{6!}{4!\ 2!}(1/2)^4(1/2)^{6-4} + \frac{6!}{5!\ 1!}(1/2)^5(1/2)^{6-5} + \frac{6!}{6!\ 0!}(1/2)^6(1/2)^{6-6}$$

$$= \frac{15}{64} + \frac{6}{64} + \frac{1}{64} = \frac{11}{32} = 0.3438$$

C) $E[X] = np = 6(1/2) = 3$

$\sigma^2[X] = npq = 6(1/2)(1/2) = 3/2$, $\sigma(X) = \sqrt{3/2} = 1.22$

DISTRIBUCION GEOMETRICA

SEA p LA PROBABILIDAD DE EXITO EN UN EXPERIMENTO. SI EL EXPERIMENTO ES CON REEMPLAZO Y SE REPITE SUCESIVAMENTE HASTA QUE SE OBSERVA UN EXITO SE TENDRA LA VARIABLE ALEATORIA X=NUMERO DE REPETICIONES DEL EXPERIMENTO HASTA QUE SE OBSERVA EL PRIMER EXITO. OBTENGAMOS LA DENSIDAD DE PROBABILIDADES DE X. $(S=\{1,2,3,\dots\})$

EL PRIMER EXITO OCURRIRA EN EL EXPERIMENTO NUMERO x SI, Y SOLO SI, EN LOS x-1 ANTERIORES HUBO PUROS FRACASOS. LA PROBABILIDAD DE ESTE EVENTO, DADO QUE LOS EXPERIMENTOS SON INDEPENDIENTES, ES

$$f_X(x) = (1-p)^{x-1}p$$

ESTA FUNCION SE DENOMINA *DISTRIBUCION GEOMETRICA* . SE PUEDE DEMOSTRAR QUE LA DISTRIBUCION DE PROBABILIDADES ACUMULADAS ES

$$F_X(x) = \sum_{x=1}^{n} p(1-p)^{x-1} = 1 - (1-p)^n$$

Y QUE

$$E(X) = \sum_{x=1}^{\infty} x(1-p)^{x-1}p = 1/p$$

$$\sigma^2(X) = \sum_{x=1}^{\infty} \left(x-\frac{1}{p}\right)^2 (1-p)^{x-1} p = (1-p)/p^2$$

EJEMPLO

¿CUAL ES LA PROBABILIDAD DE OBTENER UN TORNILLO DEFECTUOSO POR PRIMERA VEZ EN LA SEXTA EXTRACCION, SI EL PORCENTAJE DE DEFECTUOSOS DEL LOTE DEL CUAL SE MUESTREA ES DE 5 POR CIENTO?

$$P(X=6) = f_X(6) = (1-0.05)^5 \times 0.05 = 0.95^5 \times 0.05 = 0.03869$$

DISTRIBUCION HIPERGEOMETRICA

CUANDO SE TIENE UNA VARIABLE ALEATORIA DISCRETA CUYO ESPACIO DE EVENTOS TIENE SOLO DOS ELEMENTOS, DIGAMOS S={EXITO, FRACASO}, Y SE REALIZA UN MUESTREO *SIN REEMPLAZO*, ENTONCES LOS RESULTADOS DE CADA EXPERIMENTO *NO SON INDEPENDIENTES NI LA PROBABILIDAD DE EXITO PERMANECE CONSTANTE*, COMO EN LA DISTRIBUCION BINOMIAL, POR LO QUE ESTA ULTIMA NO ES APLICABLE.

SEA X LA VARIABLE ALEATORIA NUMERO DE EXITOS OBSERVADOS AL REPETIR n VECES EL EXPERIMENTO CONSISTENTE EN EXTRAER, SIN REEMPLAZO, ELEMENTOS DE UN LOTE QUE TIENE N OBJETOS DE LOS CUALES M SON "EXITOS". EL NUMERO DE ELEMENTOS QUE TIENE EL ESPACIO DE EVENTOS DEL EXPERIMENTO ES

$$N(S) = {}_{N}C_{n}$$

EL NUMERO, $N(\{X=x\})$, DE MANERAS POSIBLES E IGUALMENTE PROBABLES DE OBTENER x EXITOS ES

EN EL LOTE	EN LA MUESTRA
N ELEMENTOS	n ELEMENTOS

M EXITOS	(N-M) FRACASOS	x EXITOS QUE SE PUEDEN ELEGIR DE ${}_{M}C_{x}$ MANERAS	(n-x) FRACASOS QUE SE PUEDEN ELEGIR DE ${}_{N-M}C_{n-x}$ MANERAS

CADA ELECCION POSIBLE DE x EXITOS SE COMBINA CON CADA ELECCION POSIBLE DE (n-x) FRACASOS; POR LO TANTO, EL NUMERO TOTAL DE MANERAS DE OBTENER x EXITOS EN n EXTRACCIONES SIN REEMPLAZO ES

$$N(\{X=x\}) = ({}_{M}C_{x})({}_{N-M}C_{n-x})$$

POR LO TANTO

$$P(\{X=x\}) = f_X(x) = \frac{({}_MC_x)({}_{N-M}C_{n-x})}{{}_NC_n} = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \quad x=0,1,\ldots,n$$

EN DONDE $\binom{M}{x} = \frac{M!}{x!(M-x)!}$, $\binom{N-M}{n-x} = \frac{(N-M)!}{(n-x)!(N-M-n+x)!}$

Y $\binom{N}{n} = \frac{N!}{n!(N-n)!}$

QUE SE CONOCE COMO *DISTRIBUCION HIPERGEOMETRICA*. LA MEDIA Y LA VARIANCIA DE ESTA DISTRIBUCION SON

$$E(X) = \sum_{x=0}^{n} x\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = nM/N$$

Y

$$\sigma^2(X) = \sum_{x=0}^{n} \left(x-\frac{nM}{N}\right)^2 \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \frac{Mn(N-M)(N-n)}{N^2(N-1)}$$

RESPECTIVAMENTE.

EJEMPLO

EN UN PROBLEMA DE CONTROL ESTADISTICO DE CALIDAD, SE TIENE UN LOTE DE 100 TRANSFORMADORES DE CORRIENTE ELECTRICA, DE LOS CUALES 40 SON DEFECTUOSOS (NO CUMPLEN LAS NORMAS DE FABRICACION). ¿CUAL ES LA PROBABILIDAD DE OBTENER UNO DEFECTUOSO DE TRES SELECCIONADOS AL AZAR SIN REEMPLAZO?

$$P[X=1] = \frac{\binom{40}{1}\binom{100-40}{3\ -1}}{\binom{100}{3}} = \frac{\binom{40}{1}\binom{60}{2}}{\binom{100}{3}}$$

$$= \frac{\frac{40!}{39!\ x\ 1!} \ x\ \frac{60!}{58!\ x\ 2!}}{\frac{100!}{97!x3!}} = 0.438$$

APROXIMACION DE LA DISTRIBUCION HEPERGEOMETRICA MEDIANTE LA BINOMIAL

CUANDO *N ES GRANDE* Y *n PEQUEÑO* ($N \geq 10n$), LA DISTRIBUCION BINOMIAL SE PUEDE USAR *COMO APROXIMACION* DE LA HIPERGEOMETRICA. DE ESTA APROXIMACION SE HECHA MANO CUANDO LOS CALCULOS CON ESTA ULTIMA RESULTAN TEDIOSOS.

EN EL CASO DEL EJEMPLO ANTERIOR, SI SE USA LA DENSIDAD BINOMIAL SE OBTIENE, CON $p=40/100 = 0.40$ Y $n=3$

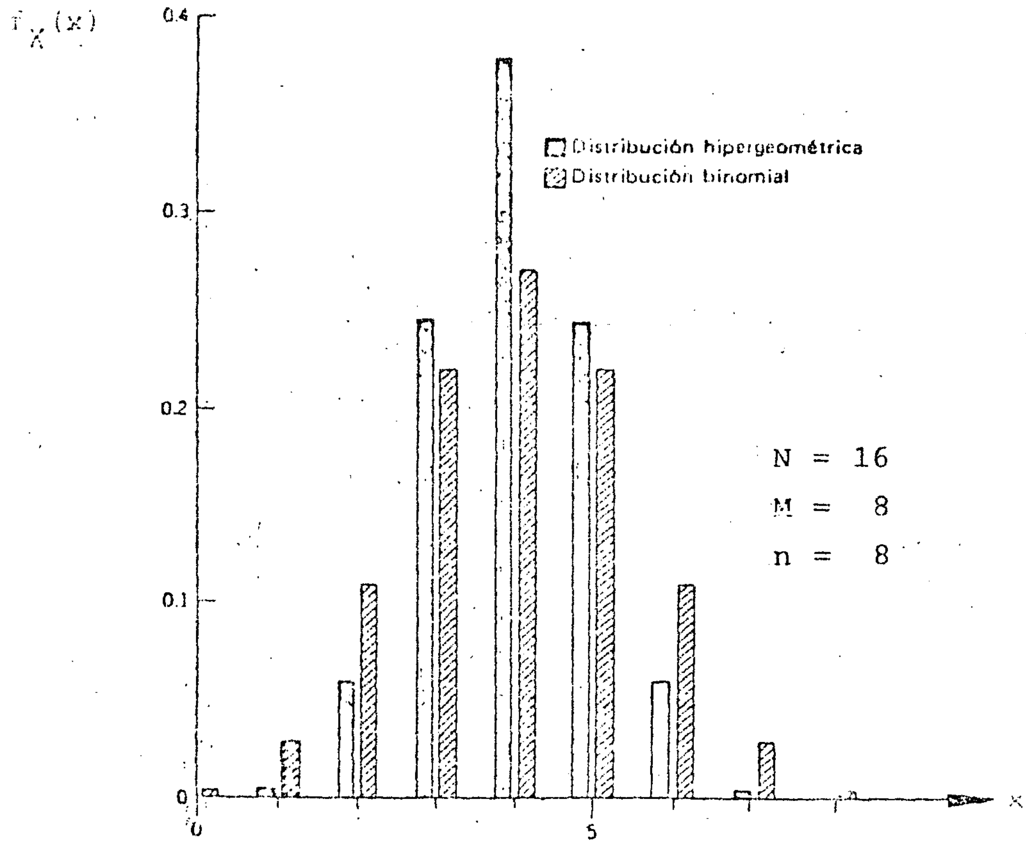
$$P[X=1] = \frac{3!}{1!\ 2!}(0.40)^1(0.60)^2 = 0.432$$

FORMULA DE STIRLING

$$N! \simeq \sqrt{2\pi n}\ \left(\frac{n}{e}\right)^n \quad ; \quad (e = 2.718...)$$

Error = 2% SI n = 4

Error = 0.8% SI n = 10



COMPARACION DE LAS DISTRIBUCIONES HIPERGEOMETRICA Y BINOMIAL

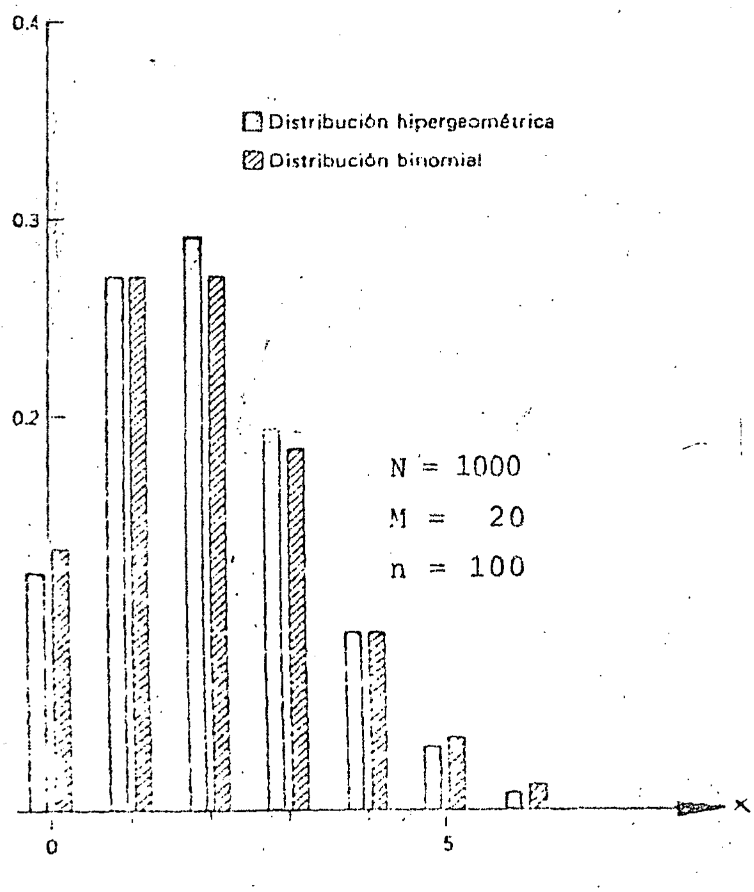


TABLA 2

FUNCION DE DISTRIBUCION DE POISSON

$$F_X(x) = \sum_{k=0}^{x} \frac{e^{-\lambda} \lambda^k}{k!}$$

x	λ .50	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0
0	.607	.368	.135	.050	.018	.007	.002	.001	.000	.000
1	.910	.736	.406	.199	.092	.040	.017	.007	.003	.001
2	.986	.920	.677	.423	.238	.125	.062	.030	.014	.006
3	.998	.981	.857	.647	.433	.265	.151	.082	.042	.021
4	1.000	.996	.947	.815	.629	.440	.285	.173	.100	.055
5	1.000	.999	.983	.961	.785	.616	.446	.301	.191	.116
6	1.000	1.000	.995	.966	.889	.762	.606	.450	.313	.207
7	1.000	1.000	.999	.988	.949	.867	.744	.599	.453	.324
8	1.000	1.000	1.000	.996	.979	.932	.847	.729	.593	.456
9	1.000	1.000	1.000	.999	.992	.968	.916	.830	.717	.587
10	1.000	1.000	1.000	1.000	.997	.986	.957	.901	.816	.706
11	1.000	1.000	1.000	1.000	.999	.995	.980	.947	.888	.803
12	1.000	1.000	1.000	1.000	1.000	.998	.991	.973	.936	.876
13	1.000	1.000	1.000	1.000	1.000	.999	.996	.987	.966	.926
14	1.000	1.000	1.000	1.000	1.000	1.000	.999	.994	.983	.959
15	1.000	1.000	1.000	1.000	1.000	1.000	.999	.998	.992	.978
16	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.996	.989
17	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.995
18	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.998
19	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

TABLA 2 (continuación)

FUNCION DE DISTRIBUCION DE POISSON

$$F_X(x) = \sum_{k=0}^{x} \frac{e^{-\lambda} \lambda^k}{k!}$$

	λ					
x	10.0	11.0	12.0	13.0	14.0	15.0
2	.003	.001	.001	.000	.000	.000
3	.010	.005	.002	.001	.000	.000
4	.029	.015	.008	.004	.002	.001
5	.067	.038	.020	.011	.006	.003
6	.130	.079	.046	.026	.014	.008
7	.220	.143	.090	.054	.032	.018
8	.333	.232	.155	.100	.062	.037
9	.458	.341	.242	.166	.109	.070
10	.583	.460	.347	.252	.176	.118
11	.697	.579	.462	.353	.260	.185
12	.792	.689	.576	.463	.358	.268
13	.864	.781	.682	.573	.464	.363
14	.917	.854	.772	.675	.570	.466
15	.951	.907	.844	.764	.669	.568
16	.973	.944	.899	.835	.756	.664
17	.986	.968	.937	.890	.827	.749
18	.993	.982	.963	.930	.883	.819
19	.997	.991	.979	.957	.923	.875
20	.998	.995	.988	.975	.952	.917
21	.999	.998	.994	.986	.971	.947
22	1.000	.999	.997	.992	.983	.967
23	1.000	1.000	.999	.996	.991	.981
24	1.000	1.000	.999	.998	.995	.989
25	1.000	1.000	1.000	.999	.997	.994
26	1.000	1.000	1.000	1.000	.999	.997
27	1.000	1.000	1.000	1.000	.999	.998
28	1.000	1.000	1.000	1.000	1.000	.999
29	1.000	1.000	1.000	1.000	1.000	1.000

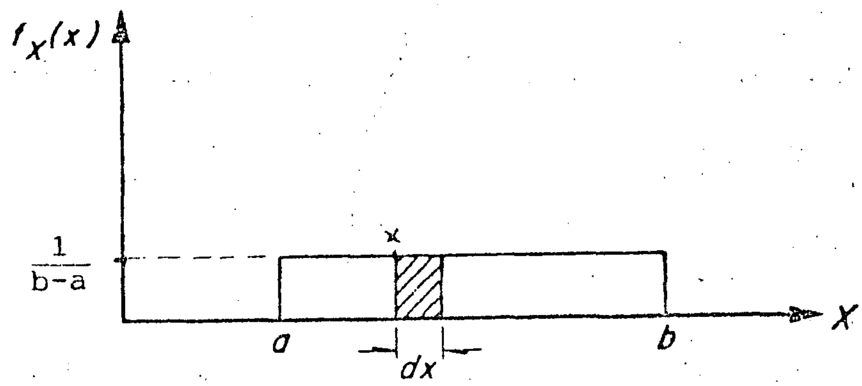
VARIABLES ALEATORIAS CONTINUAS

DISTRIBUCION UNIFORME

SE DICE QUE UNA VARIABLE ALEATORIA CONTINUA, X, TIENE *DISTRIBUCION UNIFORME* ENTRE X = a Y X = b(b>a) SI

$$f_X(x) = \text{CONSTANTE} = \frac{1}{b-a} \qquad a \le X \le b$$

LO QUE SIGNIFICA QUE LA PROBABILIDAD DE OBTENER UN VALOR ENTRE x Y x + dx ES LA MISMA PARA CUALQUIER x COMPRENDIDA ENTRE a Y b. LA GRAFICA DE DICHA DISTRIBUCION ES



Distribución uniforme de una variable aleatoria continua

LA ESPERANZA Y LA VARIANCIA DE LA DISTRIBUCION UNIFORME SE CALCULAN DE LA SIGUIENTE MANERA:

$$E[X] = \int_a^b x \frac{1}{b-a}\, dx = \left[\frac{x^2}{2(b-a)}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = (b+a)/2$$

$$\sigma^2(X) = \int_a^b (x-E[X])^2 \frac{1}{b-a}\, dx = \int_a^b \frac{x^2}{b-a}\, dx + \int_a^b \frac{(E[X])^2}{b-a}\, dx -$$

$$- \int_a^b \frac{2xE[X]}{a-b}\, dx$$

$$= \left[\frac{x^3}{3(b-a)}\right]_a^b + \left[\frac{(E[X])^2}{b-a}\, x\right]_a^b - \left[\frac{2\ E[X]}{b-a}\ \frac{x^2}{2}\right]_a^b =$$

$$= \frac{b^3-a^3}{3(b-a)} + (E[X])^2 - E[X]\,(b+a) = \frac{(b-a)^2}{12}$$

DISTRIBUCION NORMAL

UNA DE LAS DISTRIBUCIONES DE VARIABLES ALEATORIAS CONTINUAS MAS UTIL ES LA *DISTRIBUCION NORMAL* O DE *GAUSS*, DEFINIDA POR LA ECUACION

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

DONDE μ ES LA MEDIA Y σ LA DESVIACION ESTANDAR DE X.

SI SE HACE LA TRANSFORMACION

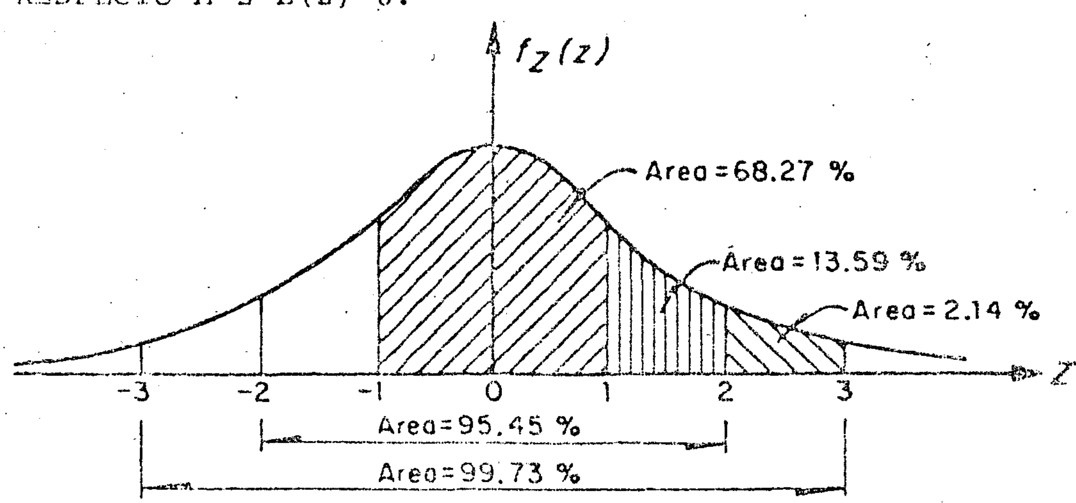
$$Z = (X-\mu)/\sigma$$

ENTONCES LA ECUACION ANTERIOR SE REDUCE A LA LLAMADA *FORMA ESTANDAR*, CUYA ECUACION ES

$$f_Z(z) = \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}$$

EN ESTE CASO LA VARIABLE ALEATORIA Z TIENE DISTRIBUCION NORMAL CON *MEDIA IGUAL A CERO Y VARIANCIA IGUAL A UNO.*

EXISTEN TABLAS PARA CALCULAR LAS PROBABILIDADES DE UNA VARIABLE ASOCIADA A UNA DISTRIBUCION NORMAL ESTANDAR. EN LA SIGUIENTE FIGURA SE MUESTRA LA FORMA DE CAMPANA DE ESTA DISTRIBUCION, OBSERVANDOSE LA SIMETRIA RESPECTO A Z=E(Z)=0.



Distribución normal de una variable aleatoria continua

LA UTILIDAD DE LA DISTRIBUCION NORMAL ESTANDAR RADICA EN QUE

$$P[x_1 \leq X \leq x_2] = \int_{x_1}^{x_2} f_X(x)\,dx = P[z_1 \leq Z \leq z_2] = \int_{z_1}^{z_2} f_Z(z)\,dz$$

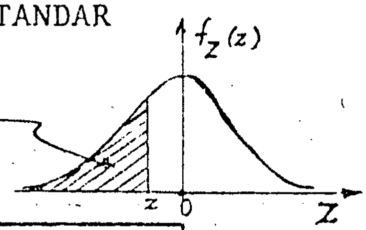
DONDE

$$z_1 = \frac{z_1 - \mu}{\sigma} \quad Y \quad z_2 = \frac{x_2 - \mu}{\sigma}$$

TABLA 3

FUNCION DE DISTRIBUCION NORMAL ESTANDAR

$$F_Z(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\, e^{-u^2/2}\, du$$

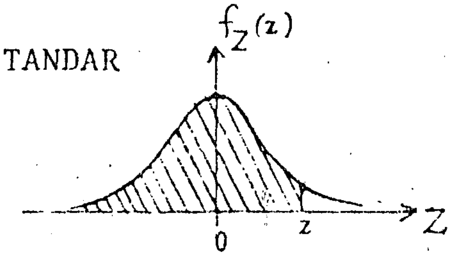


z	0	1	2	3	4	5	6	7	8	9
−3.	.0013									
−2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0227	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0300	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1921	.1894	.1867
−.7	.2420	.2389	.2358	.2326	.2297	.2266	.2236	.2206	.2177	.2148
−.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

TABLA 3 (continuación)

FUNCION DE DISTRIBUCION NORMAL ESTANDAR

$$F_Z(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$



z	0	1	2	3	4	5	6	7	8	9
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8079	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9700	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9773	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.	9987									

EJEMPLO

COMO RESULTADO DE UNA LARGA SERIE DE EXPERIMENTOS PROBANDO A COMPRESION SIMPLE CILINDROS DE CONCRETO, SE HA ESTIMADO QUE LA ESPERANZA DE LA RESISTENCIA ES DE 240 KG/CM^2 Y LA DESVIACION ESTANDAR DE 30 KG/CM^2.

A) ¿CUAL ES LA PROBABILIDAD DE QUE OTRO CILINDRO TOMADO AL AZAR RESISTA MENOS DE 240 KG/CM^2?

B. ¿CUAL ES LA PROBABILIDAD DE QUE RESISTA MAS DE 330 KG/CM^2?

C) ¿CUAL ES LA PROBABILIDAD DE QUE SU RESISTENCIA ESTE EN EL INTERVALO DE 210 A 240 KG/CM^2?

SUPONGASE QUE LA DISTRIBUCION DE PROBABILIDADES ES NORMAL.

SOLUCION

A) PARA EMPLEAR LAS TABLAS DE DISTRIBUCION NORMAL ES NECESARIO ESTANDARIZAR LA VARIABLE X, EMPLEANDO $\mu=240$ Y $\sigma=30$, CON $x_1=240$:

$$z_1 = \frac{240 - 240}{30} = 0$$

RECURRIENDO A LA TABLA DE LA DISTRIBUCION NORMAL SE OBTIENE

$$P[X \leq 240] = P[Z \leq 0] = 0.5$$

O SEA, LA PROBABILIDAD QUE CORRESPONDE AL AREA SOMBREADA DE LA SIGUIENTE FIGURA:

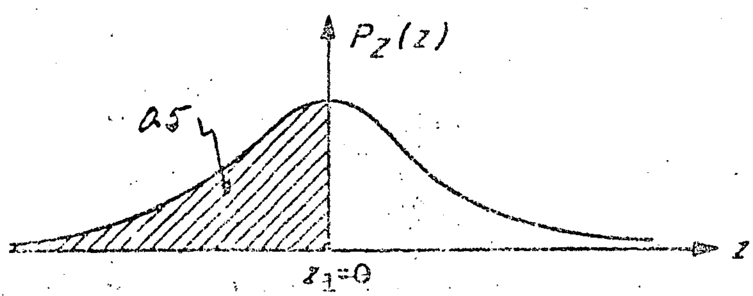


Fig 16. Distribución normal correspondiente inciso c del ejemplo

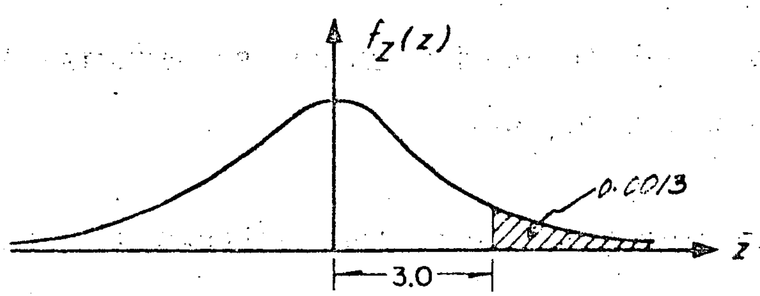
B) EL VALOR ESTANDARIZADO DE LA VARIABLE, PARA x_1=330 KG/CM^2, ES

$$z_1 = \frac{330 - 240}{30} = 3$$

POR LO QUE

$$P[X \geq 330] = P[Z \geq 3] = 1-0.9987 = 0.0013$$

QUE ES EL AREA SOMBREADA DE LA SIGUIENTE FIGURA:



Distribución normal correspondiente al inciso b del ejemplo

C) LOS VALORES ESTANDARIZADOS DE LA VARIABLE, PARA x_1=210 Y x_2=240 SON:

$$z_1 = \frac{210 - 240}{30} = -1$$

$$z_2 = \frac{240 - 240}{30} = 0$$

POR LO QUE

$$P[210 \leq X \leq 240] = P[-1 \leq Z \leq 0] = 0.3413$$

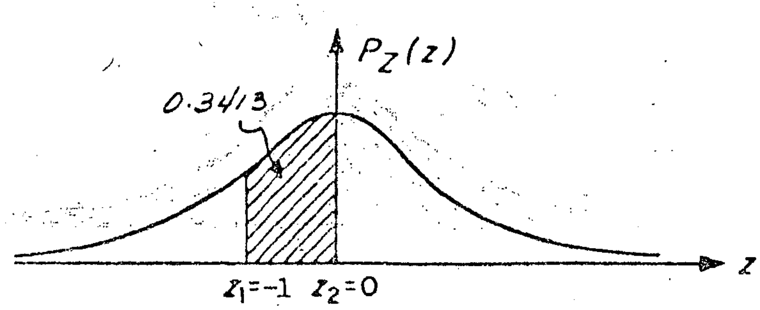


Fig 16. Distribución normal correspondiente al inciso c del ejemplo

TEOREMA CENTRAL DEL LIMITE

IDENTICAS

SEAN LAS VARIABLES ALEATORIAS $X_1, X_2, \ldots, X_k$, CON DENSIDADES DE PROBABILIDADES (ARBITRARIAS), CUYA SUMA SE DENOTARA COMO W, ES DECIR

$$W = X_1 + X_2 + \ldots + X_k$$

ES POSIBLE DEMOSTRAR EL TEOREMA DENOMINADO *TEOREMA CENTRAL DEL LIMITE*, CUYO ENUNCIADO INDICA QUE CONFORME AUMENTA EL NUMERO DE VARIABLES INVOLUCRADAS EN LA SUMA ANTERIOR (AL AUMENTAR k), LA DENSIDAD DE PROBABILIDADES DE W TIENDE A SER LA DISTRIBUCION NORMAL. ADEMAS SE PUEDE DEMOSTRAR QUE SI TODAS LAS VARIABLES $X_1, X_2, \ldots, X_k$ TIENEN DISTRIBUCION NORMAL, ENTONCES, RIGUROSAMENTE, W TAMBIEN LA TIENE, INDEPENDIENTEMENTE DEL NUMERO DE VARIABLES QUE APAREZCAN EN LA SUMA.

A PARTIR DEL TEOREMA DEL LIMITE CENTRAL SE DEMUESTRA QUE LA DISTRIBUCION DE BERNOULLI SE PUEDE APROXIMAR MEDIANTE LA NORMAL CUANDO EL NUMERO DE REPETICIONES DEL EXPERIMENTO ES GRANDE (30 O MAS), CON LO CUAL SE LOGRA UN AHORRO CONSIDERABLE DE LABOR NUMERICA EN LA SOLUCION DE ALGUNOS PROBLEMAS. PARA MEJORAR ESTA APROXIMACION, CONVIENE EFECTUAR UNA CORRECCION POR CONTINUIDAD, LA CUAL SE JUSTIFICA POR USAR UNA DISTRIBUCION CONTINUA EN VEZ DE UNA DISCRETA, SUMANDO O RESTANDO, SEGUN SEA EL CASO, 0.5 AL VALOR DE X QUE SE USE. POR EJEMPLO, SI SE DESEA CUANTIFICAR LA PROBABILIDAD DE QUE DE 2000 ENSAYES SE LOGREN DE 3 A 6 EXITOS, LOS LIMITES REALES QUE SE DEBEN USAR AL APLICAR LA DISTRIBUCION CONTINUA SON $x_1=2.5$ Y $x_2=6.5$.

EJEMPLO

SI LA PROBABILIDAD DE QUE FALLE UNA VARILLA DE ACERO AL APLICARLE CIERTA CARGA ES DE 0.001, DETERMINAR LA PROBABILIDAD DE QUE EN 2000 VARILLAS PROBADAS FALLEN MAS DE DOS.

USANDO LA DISTRIBUCION DE BERNOULLI SE OBTIENE

$$P[X > 2] = 1 - P[X \leq 2] = 1 - (P[X = 0] + P[X = 1] + P[X = 2]) =$$

$$= 1 - \left(\frac{2\,000!}{2\,000!\ 0!}(0.001)^{0}(0.999)^{2\,000} + \frac{2\,000!}{1\,999!\ 1!}(0.001)^{1}(0.999)^{1\,999} + \right.$$

$$\left. + \frac{2\,000!}{1\,998!\ 2!}(0.001)^{2}(0.999)^{1\,998}\right) = 0.3255$$

LOS CALCULOS NECESARIOS PARA OBTENER LA SOLUCION SON BASTANTE MAS TEDIOSOS QUE LOS QUE DEBEN EFECTUARSE APROVECHANDO QUE EL NUMERO DE REPETICIONES DEL EXPERIMENTO ES GRANDE, A FIN DE UTILIZAR LA DISTRIBUCION NORMAL. EN ESTAS CIRCUNSTANCIAS, LA PROBABILIDAD DE QUE $X \leq 2$ EN EL CASO DISCRETO, EQUIVALE A LA DE QUE $X \leq 2.5$ EN EL CONTINUO; ASI

$$\mu = np = 2\,000 \times 0.001 = 2$$

$$\sigma = \sqrt{npq} = \sqrt{2\,000 \times 0.001 \times 0.999} = 1.41$$

$$P[X \leq 2.5] = P\left[Z \leq \frac{2.5 - 2}{1.41}\right] = P[Z \leq 0.355] = 0.6387$$

DE DONDE

$$P[x > 2.5] = 1 - P[X \leq 2.5] = 1 - 0.6387 = 0.3613$$

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

FUNDAMENTOS DE LAS TECNICAS DE MUESTREO ESTADISTICO

ESTADISTICA DESCRIPTIVA

DR. OCTAVIO A. RASCON CHAVEZ

FEBRERO, 1979.

ESTADISTICA DESCRIPTIVA

POR [illegible]

DATO U OBSERVACION: ES EL RESULTADO DE REALIZAR UN EXPERIMENTO.

MUESTRA: ES UNA COLECCION DE DATOS

MUESTREO: PROCESO DE ADQUISICION DE UNA MUESTRA

MUESTREO
- *CON REEMPLAZO.-* CUANDO CADA ELEMENTO OBSERVADO SE REINTEGRA AL LOTE DEL CUAL FUE EXTRAIDO ANTES DE EXTRAER EL SIGUIENTE.
- *SIN REEMPLAZO.-* CUANDO CADA ELEMENTO OBSERVADO NO SE REINTEGRA AL LOTE.

POBLACION: TOTAL DE DATOS QUE SE PUEDEN OBTENER AL REALIZAR UNA SECUENCIA EXHAUSTIVA DE EXPERIMENTOS

POBLACION
- *DISCRETA.-* TIENE UN NUMERO FINITO O UN NUMERO INFINITO NUMERABLE DE DATOS POSIBLES
- *CONTINUA.-* TIENE UN NUMERO INFINITO NO NUMERABLE DE DATOS POSIBLES

EJEMPLOS

1. EXPERIMENTO: LANZAMIENTO DE UNA MONEDA DIEZ VECES

   POBLACION: SUCESION INFINITA NUMERABLE DE "CARAS" Y "CRUCES" (DISCRETA)

   MUESTRA: GRUPO DE 10 OBSERVACIONES

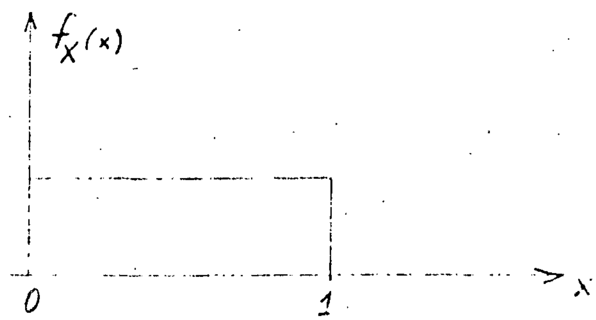
2. EXPERIMENTO: MEDICION DE LA PRECIPITACION PLUVIAL DIARIA EN LA CIUDAD DE MEXICO DURANTE DIEZ AÑOS

   POBLACION: SUCESION INFINITA NO NUMERABLE DE VALORES (CONTINUA)

   MUESTRA: GRUPO DE 3652 OBSERVACIONES (TOMANDO DOS AÑOS BISIESTOS DE 29 DIAS EN FEBRERO)

MUESTRA ALEATORIA: ES UNA MUESTRA OBTENIDA DE TAL MANERA QUE *TODOS* LOS ELEMENTOS DE LA POBLACION TIENEN LA *MISMA* PROBABILIDAD DE SER OBSERVADOS Y, ADEMAS, LA OBSERVACION DE UN ELEMENTO *NO AFECTA* LA PROBABILIDAD DE OBSERVAR CUALQUIER OTRO, ES DECIR, SI SON INDEPENDIENTES.

TABLA DE NUMEROS ALEATORIOS: ES UNA TABLA QUE CONTIENE NUMEROS QUE CONSTITUYEN UNA MUESTRA ALEATORIA OBTENIDA DE UNA DISTRIBUCION DE PROBABILIDADES UNIFORME, QUE GENERALMENTE CORRESPONDE A UNA VARIABLE ALEATORIA QUE PUEDE ASUMIR VALORES ENTRE 0 Y 1, MULTIPLICADOS POR 10^r, EN DONDE r ES EL NUMERO DE DIGITOS QUE SE DESEA TENGAN LOS NUMEROS.



LAS TABLAS QUE SE USEN PARA OBTENER UNA MUESTRA ALEATORIA DEBEN CONTENER NUMEROS CON MAYOR NUMERO DE DIGITOS QUE LOS QUE TIENE EL TOTAL DE ELEMENTOS DE LA POBLACION QUE SE VA A MUESTREAR. POR EJEMPLO, SI SE VA A OBTENER UNA MUESTRA ALEATORIA DE UN LOTE DE LENTES PARA MICROSCOPIO QUE TIENE 10,000 ELEMENTOS, LA TABLA QUE SE USE DEBERA TENER NUMEROS ALEATORIOS CON 5 O MAS DIGITOS.

METODO DE MUESTREO ALEATORIO

1. SE ENUMERAN LOS ELEMENTOS DE LA POBLACION
2. SE FIJA EL CRITERIO DE SELECCION DE LOS NUMEROS ALEATORIOS (POR EJEMPLO, SE DEFINE QUE RENGLONES Y QUE COLUMNAS SE VAN A LEER)
3. SE INDICA QUE DIGITOS SE VAN A ELIMINAR EN CASO DE QUE LOS NUMEROS DE LA TABLA TENGAN MAS DIGITOS QUE LOS NECESARIOS
4. SE LEEN LOS NUMEROS, DE ACUERDO CON LO FIJADO EN LOS PUNTOS 2 Y 3, Y SE EXTRAEN DEL LOTE LOS ELEMENTOS QUE TIENEN LOS NUMEROS LEIDOS. ESTOS CONSTITUYEN LA MUESTRA FISICA CON LA CUAL REALIZAR LOS EXPERIMENTOS. LAS OBSERVACIONES CONSTITUIRAN LA MUESTRA ALEATORIA DESEADA.

NOTA: TODOS LOS NUMEROS QUE SE REPITAN SE CONSIDERAN SOLO UNA VEZ. TAMBIEN SE ELIMINAN LOS NUMEROS MAYORES DEL TAMAÑO DEL LOTE.

<u>EJEMPLO</u>

SE TIENE UN LOTE DE 1,000 TRANSISTORES NUMERADOS DEL UNO AL MIL, CUYA CALIDAD SE VA A VERIFICAR ESTADISTICAMENTE, PARA LO CUAL SE DECIDE TOMAR UNA MUESTRA DE 40 ELEMENTOS Y MEDIR SU AMPLIFICACION, USANDO LA TABLA DE NUMEROS ALEATORIOS ANEXA, CON EL CRITERIO DE TOMAR TODOS LOS <u>RENGLONES IMPARES</u> ELIMINANDO EL ULTIMO DIGITO, LA MUESTRA FISICA SERIAN LOS TRANSISTORES CORRESPONDIENTES A LOS NUMEROS 0415, 0006, 0394, 0998, 0530, 0160, ETC.

TABLA DE NUMEROS ALEATORIOS

Renglón \ Columna	1	2	3	4	5	6	7	8	9	10	11
1	16408	81899	04153	53381	79401	21438	83035	92350	36693	31238	59649
2	18629	81953	05520	91962	04739	13092	37662	94822	94730	06496	35090
3	73115	47498	47498	87637	99016	00060	88824	71013	18735	20286	23153
4	57491	16703	23167	49323	45021	33132	12544	41035	80780	45393	44812
5	30405	03946	23792	14422	15059	45799	22716	19792	09983	74353	68668
6	16631	35006	85900	32388	52390	52390	16815	69298	38732	38480	73817
7	96773	20206	42559	78985	05300	22164	24369	54224	35083	19687	11052
8	38935	64202	14349	82674	66523	44133	00697	35552	35970	19124	63318
9	31624	76384	17403	03941	44167	64486	64758	75366	76554	01601	12614
10	78919	19474	23632	27889	47914	02584	37680	20801	72152	39339	34806

AGRUPAMIENTO DE DATOS

FRECUENCIA DE UN EVENTO:- ES EL NUMERO DE VECES QUE OCURRE EL EVENTO AL OBTENER UNA MUESTRA DE LA POBLACION CORRESPONDIENTE.

FRECUENCIA RELATIVA DE UN EVENTO:- ES EL COCIENTE DE SU FRECUENCIA ENTRE EL TOTAL DE ELEMENTOS (TAMAÑO) DE LA MUESTRA.

FRECUENCIA RELATIVA ACUMULADA:- ES LA ACUMULACION (SUMA) DE LAS FRECUENCIAS RELATIVAS HASTA UN VALOR DADO, PARTIENDO DEL VALOR (O DEL INTERVALO) MAS PEQUEÑO. EN OTRAS PALABRAS, ES LA FRECUENCIA DE VALORES MENORES O IGUALES QUE UN VALOR DADO.

FRECUENCIA COMPLEMENTARIA:- ES LA FRECUENCIA DE VALORES MAYORES QUE UN VALOR DADO = NUMERO DE DATOS - FRECUENCIA ACUMULADA.

DISTRIBUCION DE FRECUENCIAS

CON OBJETO DE FACILITAR LA INTERPRETACION DE LOS DATOS QUE SE TIENEN EN UNA MUESTRA, ES CONVENIENTE AGRUPARLOS POR VALORES O POR INTERVALOS DE VALORES, FORMANDO ASI UNA TABLA DE DISTRIBUCION DE FRECUENCIAS.

PARA FACILITAR EL CALCULO DE LAS FRECUENCIAS ES UTIL ORDENAR LOS DATOS EN FORMA CRECIENTE O DECRECIENTE DE VALORES, FORMANDO ASI UNA *TABLA DE DATOS ORDENADOS.*

EJEMPLO

EN UNA ESCUELA SECUNDARIA SE LES APLICO A 30 PROFESORES UN EXAMEN SOBRE PEDAGOGIA. LAS CALIFICACIONES (DATOS) QUE SE OBTUVIERON FUERON (YA ESTAN ORDENADOS EN FORMA CRECIENTE)

$\underbrace{57,\ 59,}_{A}\ \underbrace{65,\ 67,\ 67,\ 67,\ 69,}_{B}\ \underbrace{72,\ 73,\ 73,\ 77,\ 78,\ 78,}_{C}$

$\underbrace{81,\ 81,\ 83,\ 83,\ 83,\ 84,\ 84,\ 87,\ 88,\ 89,\ 89,}_{D}\ \underbrace{91,\ 91,\ 93,}_{E}$

$\underbrace{95,\ 97,\ 99}_{E}$

AGRUPAMIENTO DE VALORES

CALIFICACION	FRECUENCIA	FRECUENCIA RELATIVA	FRECUENCIA RELATIVA ACUMULADA
57	1	1/30	1/30
59	1	1/30	2/30
65	1	1/30	3/30
67	3	3/30	6/30
69	1	1/30	7/30
72	1	1/30	8/30
73	2	2/30	10/30
77	1	1/30	11/30
78	2	2/30	13/30
81	2	2/30	15/30
83	3	3/30	18/30
84	2	2/30	20/30
87	1	1/30	21/30
88	1	1/30	22/30
89	2	2/30	24/30
91	2	2/30	26/30
93	1	1/30	27/30
95	1	1/30	28/30
97	1	1/30	29/30
99	1	1/30	30/30=1
	$\Sigma=30$	$\Sigma=30/30=1$	

AGRUPAMIENTO POR INTERVALOS

LIMITES DE CLASES; SON LOS VALORES MINIMO Y MAXIMO DE CADA INTERVALO

MARCAS DE CLASE: SON LOS VALORES MEDIOS DE CADA INTERVALO DE CLASE

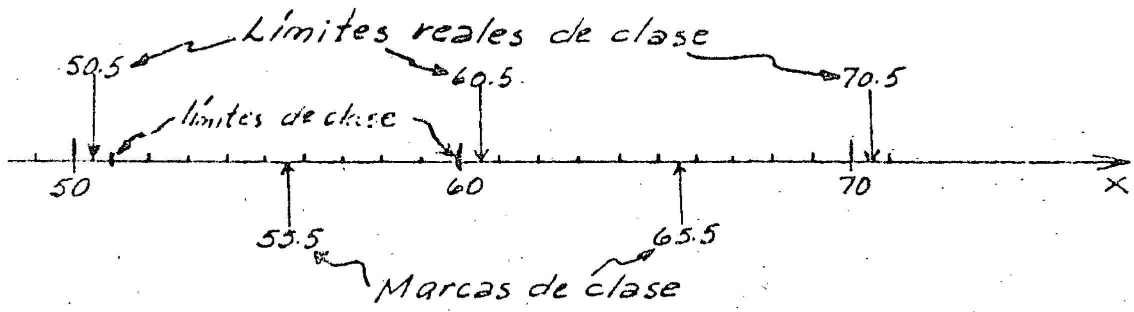
LIMITES REALES DE CLASE: SON LOS VALORES MINIMO Y MAXIMO QUE SON FRONTERA ENTRE LOS INTERVALOS. ESTOS DEBEN TENER UNA CIFRA DECIMAL MAS QUE LOS DATOS.

EVENTO (INTERVALO DE CALIFICACIONES)	ELEMENTOS OBSERVADOS	FRECUENCIA	FRECUENCIA RELATIVA
A = [51-60]	57,59	2	2/30
B = [61-70]	65,67,67,67,69	5	5/30
C = [71-80]	72,73,73,77,78,78	6	6/30
D = [81-90]	81,81,83,83,83,84, 84,87,88,89,89	11	11/30
E = [91-100]	91,91,93,95,97,99	6	6/30
		Σ=30	30/30=1

LIMITES INFERIORES DE CLASE | LIMITES SUPERIORES DE CLASE

EVENTO	LIMITES DE CLASE		LIMITES REALES DE CLASE		MARCAS DE CLASE
	INFERIOR	SUPERIOR	INFERIOR	SUPERIOR	
A	51	60	50.5	60.5	55.5
B	61	70	60.5	70.5	65.5
C	71	80	70.5	80.5	75.5
D	81	90	80.5	90.5	85.5
E	91	100	90.5	100.5	95.5

vento	Elementos corresp. a los intervalos	Frecuencia	Frecuencia relativa	Frecuencia acumulada	Frecuencia relativ acumulada
A: 51-60	59,57	2	2/30=0.067(6.7%)	2	0.067
B: 61-70	67,65,69,67,67	5	5/30=0.166(16.6%)	2+5=7	0.067+0.166=0.233
C: 71-80	72,73,73,77,78,78,	6	6/30=0.200(20%)	7+6=13	0.233+0.200=0.433
D: 81-90	83,88,84,89,83,84, 89,87,81,83,81	11	11/30=0.367(36.7%)	13+11=24	0.433+0.367=0.800
E: 91-100	99,91,97,95,91,93	6	6/30=0.200(20%)	24+6=30	0.800+0.200=1.000
		30	1.000		



$A = \{X:\ 50.5 < X \leq 60.5\}$

$B = \{X:\ 60.5 < X \leq 70.5\}$

$C = \{X:\ 70.5 < X \leq 80.5\}$

$D = \{X:\ 80.5 < X \leq 90.5\}$

$E = \{X:\ 90.5 < X \leq 100.5\}$

LIMITES REALES INFERIORES DE CLASE

LIMITES REALES SUPERIORES DE CLASE

A MAYOR NUMERO DE DATOS SE REQUIERE MAYOR NUMERO DE INTERVALOS, PERO SE RECOMIENDA QUE ESTE NUMERO ESTE ENTRE 5 Y 20, SUPONIENDO QUE EN PROMEDIO CAIGAN 5 O MAS ELEMENTOS EN CADA INTERVALO. ASI, SI SE TIENEN 30 DATOS, SE RECOMIENDA USAR 30/5=6 INTERVALOS.

EL PROCESO DE AGRUPAMIENTO SE INDICARA AL MISMO TIEMPO QUE SE REALIZA EL SIGUIENTE EJEMPLO.

EJEMPLO

EN UN ESTUDIO ANTROPOLOGICO SE OBTUVO UNA MUESTRA DE 30 ESTATURAS

DE LOS VARONES ADULTOS RESIDENTES EN UNA REGION. LOS DATOS, ORDENADOS EN FORMA CRECIENTE DE VALORES, FUERON LOS SIGUIENTES: 159,161,163,163,163,167,167,167,167,168,168,168,169,169,170, 171,171,173,174,175,175,175,178,179,181,181,183,184,187,191 CM. OBTENER LA TABLA DE DISTRIBUCION DE FRECUENCIAS.

SOLUCION:

1. DETERMINACION DEL RANGO DE LA MUESTRA

RANGO = VALOR MAXIMO - VALOR MINIMO = 191-159=32 CM

2. DETERMINACION DEL NUMERO DE INTERVALOS

$$\text{NUMERO DE INTERVALOS} = \frac{30}{5} = 6$$

3. DETERMINACION DE LOS LIMITES DE CLASE

$$\text{ANCHO DE LOS INTERVALOS} = \frac{\text{RANGO}}{\text{NUMERO}} = \frac{32}{6} = 5.3$$

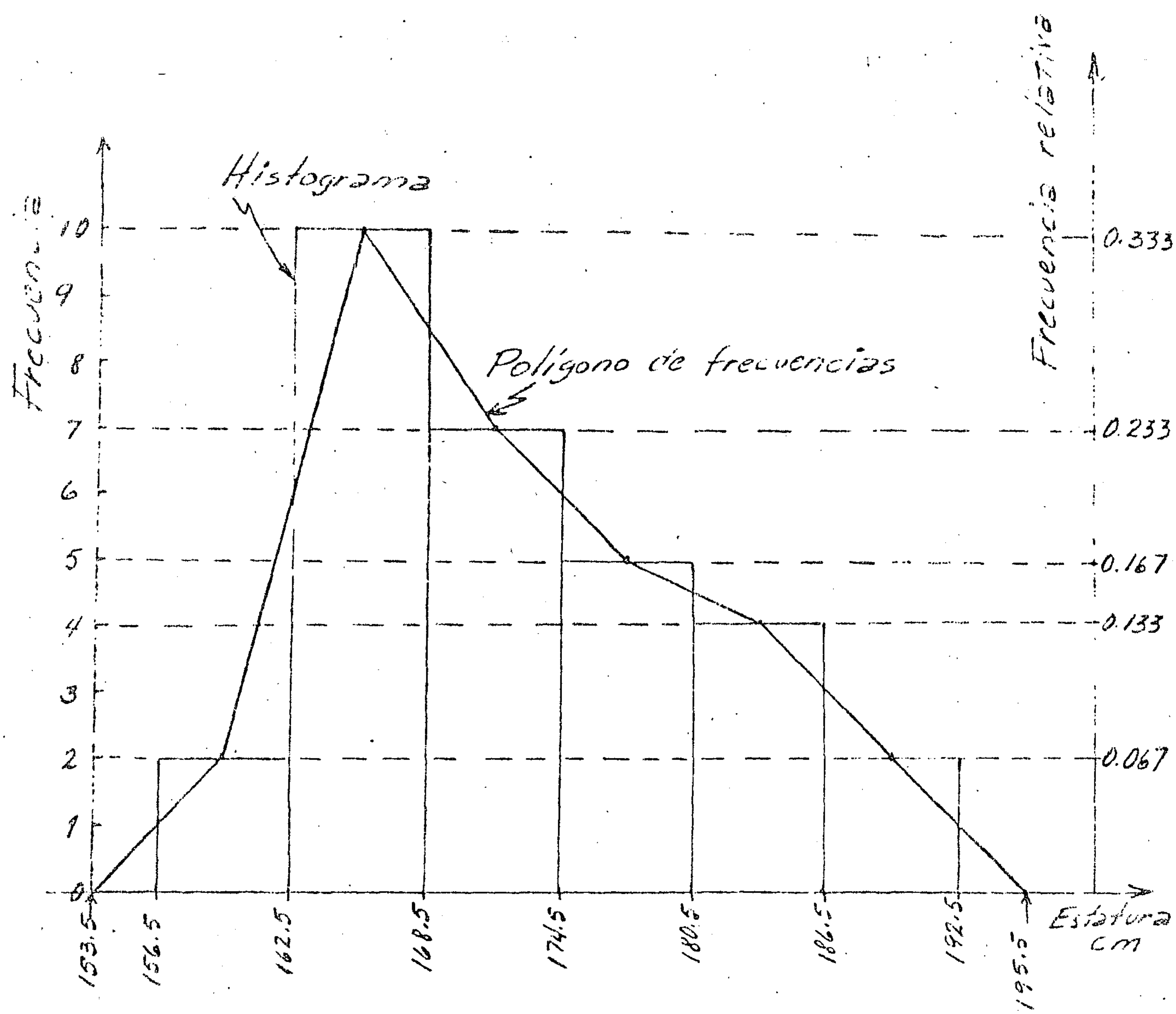
TOMAREMOS UN ANCHO DE 6 CM, CON LO CUAL EL RANGO DEL AGRUPAMIENTO ES 6 x 6 =36 CM. LA DIFERENCIA DE RANGOS ES 36-32=4, QUE SE REPARTE EN LOS DOS INTERVALOS EQUITATIVAMENTE. POR LO TANTO, LOS INTERVALOS RESULTAN SER:

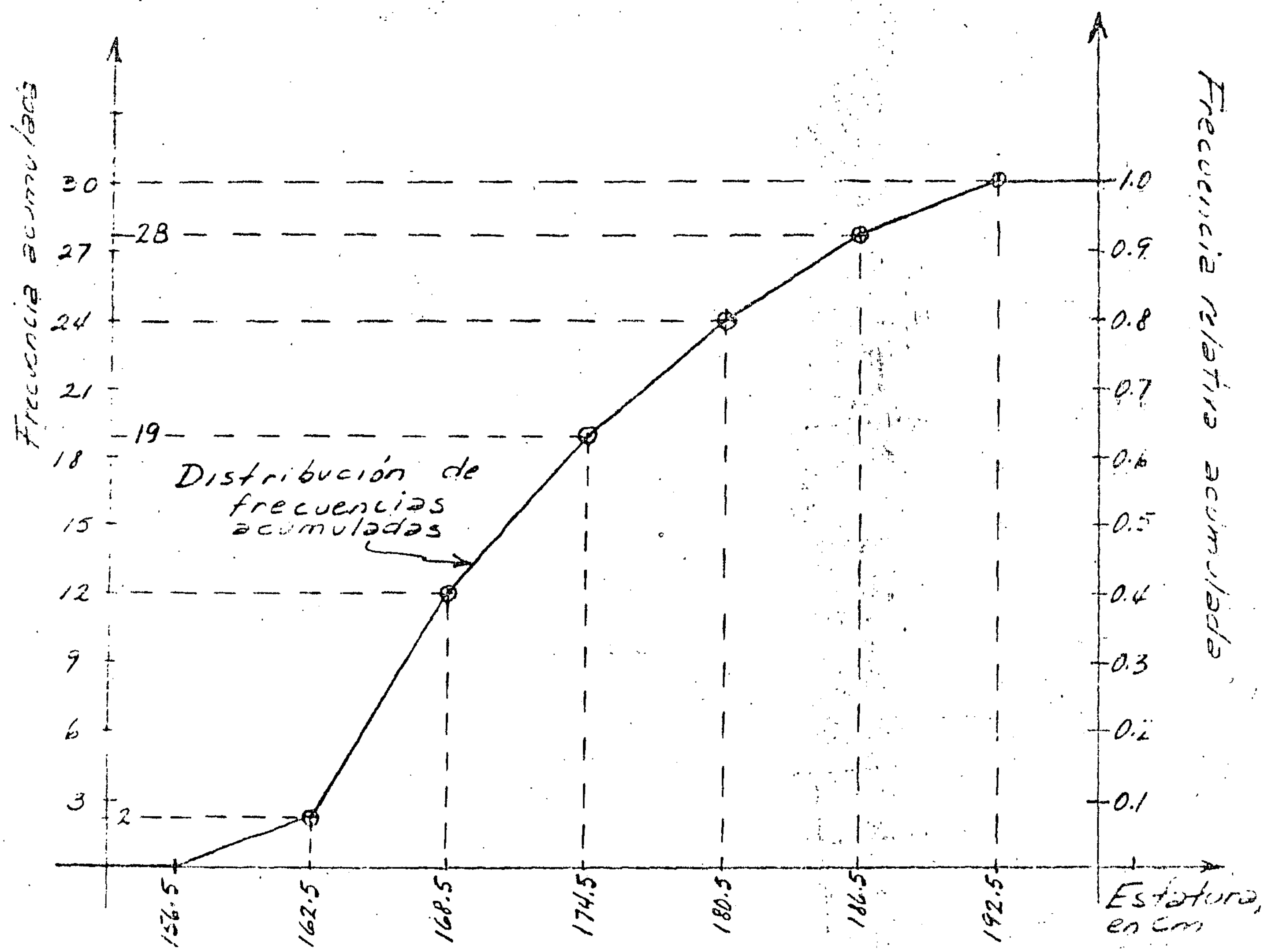
157-162, 163-168, 169-174, 175-180, 181-186, 187-192

4. INTEGRACION DE LA TABLA:

INTERVALO	LIMITES REALES		FREC.	FREC. REL.	FREC. ACUM.	FREC. REL. ACUM.
	INF.	SUP.				
157-162	156.5	162.5	2	$\frac{2}{30} = 0.067$	2	0.067
163-168	162.5	168.5	10	$\frac{10}{30} = 0.333$	12	0.400
169-174	168.5	174.5	7	$\frac{7}{30} = 0.233$	19	0.633
175-180	174.5	180.5	5	$\frac{5}{30} = 0.167$	24	0.800
181-186	180.5	186.5	4	$\frac{4}{30} = 0.133$	28	0.933
187-192	186.5	192.5	2	$\frac{2}{30} = 0.067$	30	1.000
			$\Sigma=30$	$\Sigma = 1.000$		

PRESENTACION GRAFICA DE LAS DISTRIBUCIONES DE FRECUENCIAS

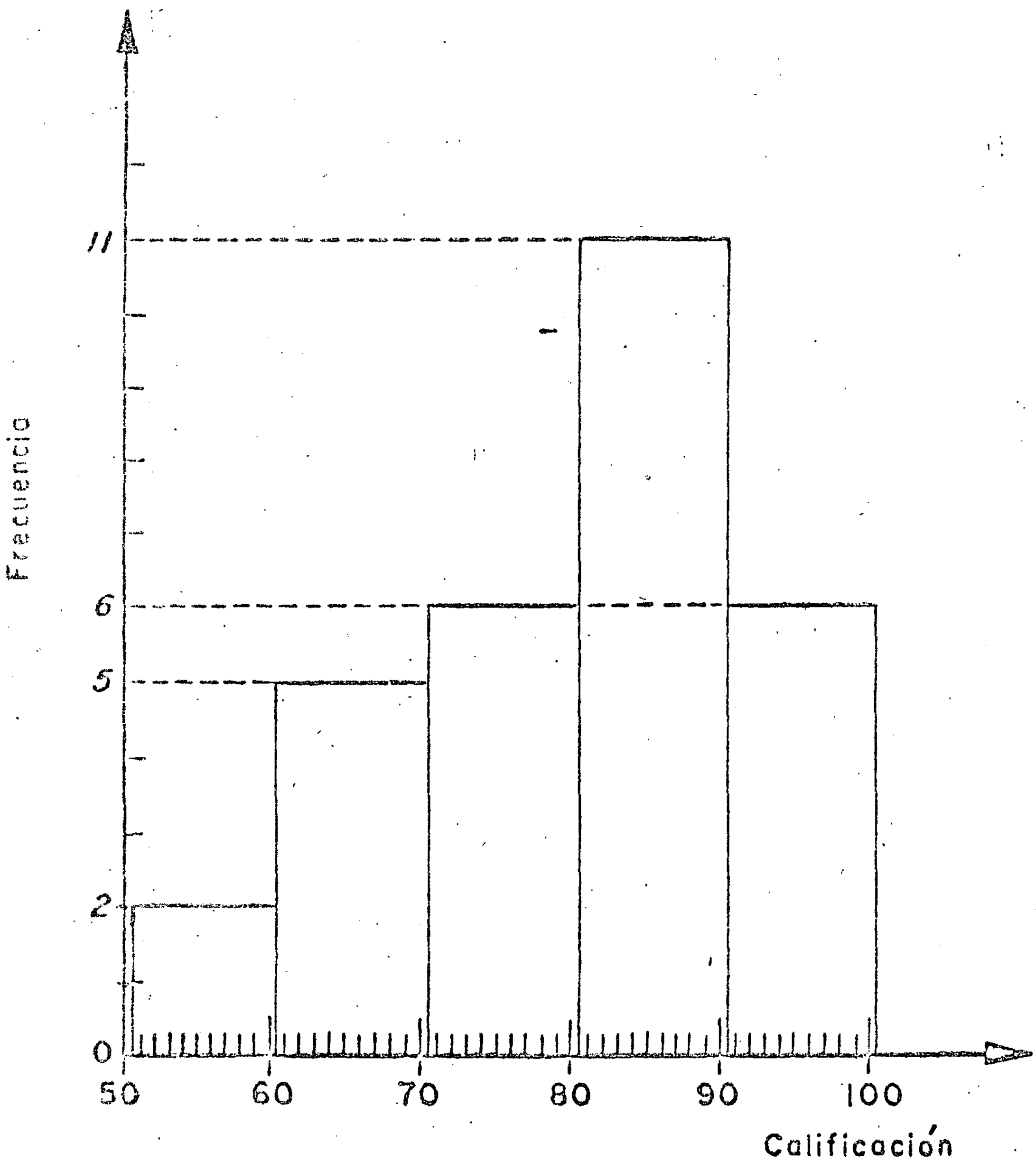




¿CUAL ES LA FRECUENCIA DE VALORES MAYORES QUE 180.5?: 30-24=6

LA FRECUENCIA RELATIVA ACUMULADA COMPLEMENTARIA ES: 1-0.800=0.200 (20%)

HISTOGRAMA DEL PROBLEMA DE LAS CALIFICACIONES EN PEDAGOGIA

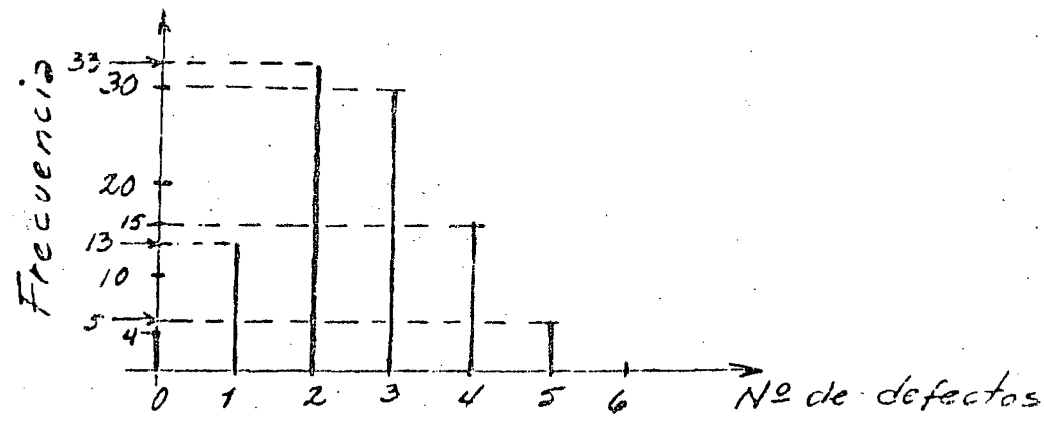


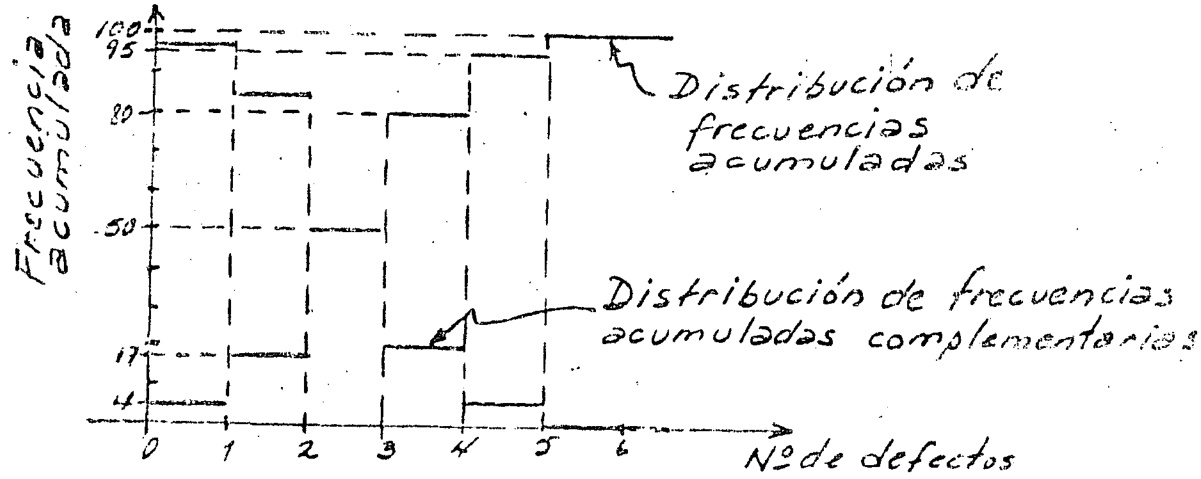
TAREA: DIBUJAR EL POLIGONO, DE FRECUENCIAS Y LAS CURVAS DE FRECUENCIAS RELATIVAS ACUMULADAS Y COMPLEMENTARIAS.

EJEMPLO

EN UN ESTUDIO SOBRE LA CALIDAD DE LOS MONOBLOCKS PRODUCIDOS POR UNA FABRICA, SE OBTUVO UNA MUESTRA ALEATORIA DE 100 ELEMENTOS, A LOS CUALES SE LES CONTO EL NUMERO DE DEFECTOS DE FABRICACION. LA DISTRIBUCION DE FRECUENCIAS QUE SE OBTUVO ES LA SIGUIENTE:

NUMERO DE DEFECTOS	FRECUENCIA	FRECUENCIA ACUMULADA	FRECUENCIA ACUMULADA COMPLEMENTARIA
0	4	4	96 (100-4)
1	13	17	83 (100-17)
2	33	50	50 (100-50)
3	30	80	20 (100-80)
4	15	95	5 (100-95)
5	5	100	0 (100-100)
	100		



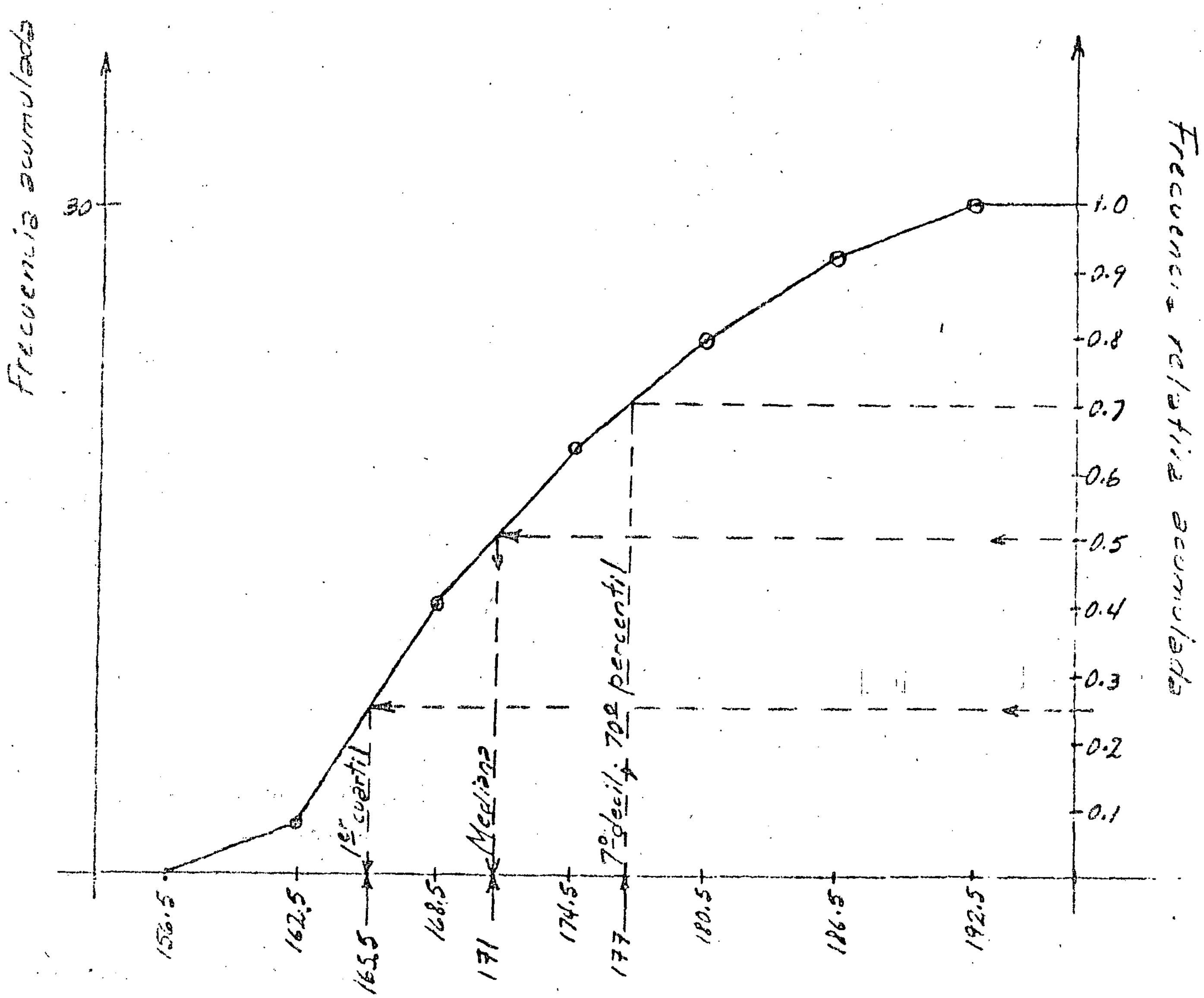


PERCENTILES:- SON LOS VALORES DE LA VARIABLE CORRESPONDIENTES A FRECUENCIAS RELATIVAS ACUMULADAS QUE SON MULTIPLOS DE 1 POR CIENTO.

DECILES:- SON LOS VALORES DE LA VARIABLE CORRESPONDIENTES A FRECUENCIAS RELATIVAS ACUMULADAS QUE SON MULTIPLOS DE 10 POR CIENTO.

CUARTILES:- SON LOS VALORES DE LA VARIABLE CORRESPONDIENTES A FRECUENCIAS RELATIVAS ACUMULADAS QUE SON MULTIPLOS DE 25 POR CIENTO.

MEDIANA:- VALOR DE LA VARIABLE CORRESPONDIENTE A LA FRECUENCIA RELATIVA ACUMULADA DE 50%.



MEDIDAS REPRESENTATIVAS DE LOS DATOS

MEDIDAS DE TENDENCIA CENTRAL

VALOR MEDIO O PROMEDIO ARITMETICO

PARA DATOS NO AGRUPADOS:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

DONDE x_i SON LOS VALORES DE LOS DATOS Y n ES EL TAMAÑO DE LA MUESTRA.

SI LOS DATOS ESTAN AGRUPADOS Y f_j ES LA FRECUENCIA DEL j-ESIMO INTERVALO Y x_j ES LA MARCA DE CLASE CORRESPONDIENTE, ENTONCES

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{K} f_j x_j \quad ; \quad K = \text{NUMERO DE INTERVALOS}$$

EJEMPLO

SEA EL EJEMPLO ENUNCIADO ANTERIORMENTE DE LOS DEFECTOS EN MONOBLOCKS. SE TENIA:

j	No. DE DEFECTOS x	FRECUENCIA f	fx
1	0	4	4 x 0 = 0
2	1	13	13 x 1 = 13
3	2	33	33 x 2 = 66
4	3	30	30 x 3 = 90
5	4	15	15 x 4 = 60
K=6	5	5	5 x 5 = 25
		Σ=100	$\sum_{j=1}^{6}$ 254

$$\bar{x} = \frac{254}{100}$$

$\bar{x}$ = 2.54 DEFECTOS POR MONOBLOCK

MODO.- ES EL VALOR DE LA VARIABLE QUE APARECE CON MAYOR FRECUENCIA EN UNA MUESTRA. SI LOS DATOS ESTAN AGRUPADOS, EL MODO ES LA MARCA DE CLASE DEL INTERVALO QUE TIENE LA MAYOR FRECUENCIA.

EJEMPLO

EN EL PROBLEMA DE LOS MONOBLOCKS EL MODO ES 2. EN EL PROBLEMA DE LAS ESTATURAS DE LOS VARONES ADULTOS DE UNA CIUDAD EL MODO ES 165.5 CM.

<u>MEDIANA</u>: ES EL VALOR DE LA VARIABLE QUE CORRESPONDE AL 50% DE LA FRECUENCIA RELATIVA ACUMULADA.

SI LOS DATOS ESTAN AGRUPADOS POR INTERVALOS, LA MEDIANA SE PUEDE CALCULAR CON LA FORMULA (ADEMAS DE GRAFICAMENTE, COMO YA SE VIO):

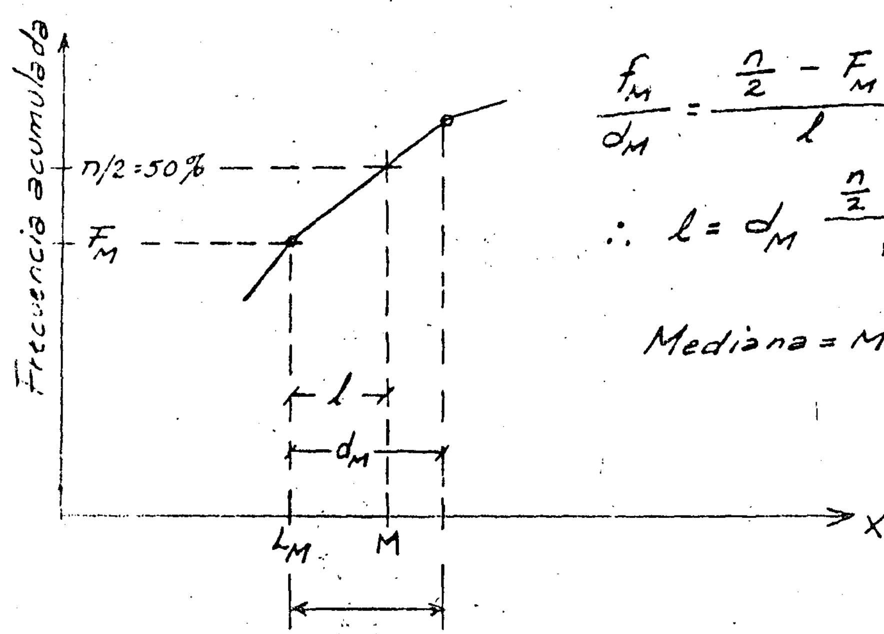
$$\text{MEDIANA} = M = L_M + \frac{\frac{n}{2} - F_M}{f_M} d_M$$

DONDE L_M = LIMITE INFERIOR REAL DEL INTERVALO QUE CONTIENE A LA MEDIANA

f_M Y d_M = RESPECTIVAMENTE, A LA FRECUENCIA Y ANCHO DEL INTERVALO QUE CONTIENE A LA MEDIANA

F_M = FRECUENCIA ACUMULADA HASTA EL INTERVALO QUE CONTIENE A LA MEDIANA EXCLUSIVE

n = TAMAÑO DE LA MUESTRA



Intervalo que contiene a la mediana

$$\frac{f_M}{d_M} = \frac{\frac{n}{2} - F_M}{l}$$

$$\therefore\ l = d_M \frac{\frac{n}{2} - F_M}{f_M}$$

$$\text{Mediana} = M = L_M + l$$

EJEMPLO

EN UN ESTUDIO PARA DETERMINAR LOS TIEMPOS EN QUE UNA MUESTRA ALEATORIA DE INDIVIDUOS REACCIONABA A CIERTOS ESTIMULOS PSICOLOGICOS SE OBTUVO LO SIGUIENTE:

j	MARCA DE CLASE x, EN SEG	LIMITES REALES	FRECUENCIA f	FRECUENCIA ACUMULADA, F	fx, SEG
1	0.10	0.075-0.125	2	2	0.20
2	0.15	0.125-0.175	7	9	1.05
3	0.20	0.175-0.225	14	23	2.80
4	0.25	0.225-0.275	4	27	1.00
K=5	0.30	0.275-0.325	3	30	0.90
			$\Sigma=30$		$\sum_{j=1}^{5} f_i x_i = 5.95$

$$\bar{x} = \frac{5.95}{30} = 0.198 \text{ SEG}$$

MODO = 0.20 SEG

$$d_M = 0.05, \quad L_M = 0.20 - \frac{0.05}{2} = 0.175, \quad F_M = 9$$

$$n/2 = 30/2 = 15, \quad f_M = 14$$

$$\text{MEDIANA} = M = 0.175 + \frac{15-9}{14}\,0.05$$

$$M = 0.175 + \frac{0.30}{14} = 0.175 + 0.021 = 0.196 \text{ SEG}$$

MEDIDAS DE DISPERSION

RANGO = MAXIMO VALOR OBSERVADO - MINIMO VALOR OBSERVADO

VARIANCIA : SI LOS DATOS *NO* ESTAN AGRUPADOS;

$$s_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \overline{x^2} - \bar{x}^2$$

SI LOS DATOS ESTAN AGRUPADOS:

$$s_X^2 = \frac{1}{n}\sum_{j=1}^{K} f_j(x_j - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{k} f_i\, x_i^2 - \bar{x}^2 = \overline{x^2} - \bar{x}^2$$

DONDE LAS x_j SON LOS VALORES DE LAS MARCAS DE CLASE DE LOS INTERVALOS.

DESVIACION ESTANDAR

$$s_X = \sqrt{s_X^2}$$

COEFICIENTE DE VARIACION

$$v_X = s_X/\bar{x}$$

EJEMPLO

EN UN ESTUDIO SOBRE LA TEMPERATURA MAXIMA DIARIA EN UNA CIUDAD SE OBTUVO LO SIGUIENTE DURANTE UNA PRIMAVERA:

j	INTERVALOS DE TEMPERATURA, °F	MARCA DE CLASE, °F	FRECUENCIA f	xf	$x-\bar{x}$	$(x-\bar{x})^2$	$(x-\bar{x})^2 f$
1	55 - 63	59	2	118	-21.3	453.7	907.4
2	64 - 72	68	6	408	-12.3	151.3	907.8
3	73 - 81	77	7	539	- 3.3	10.9	76.3
4	82 - 90	86	9	774	5.7	32.5	292.5
5	91 - 99	95	6	570	14.7	216.1	1296.6
			30	2409			3480.6

$$\bar{x} = \frac{2409}{30} = 80.3 \text{ °F}$$

$$S_X^2 = \frac{3480.6}{30} = 116 \text{ °F}^2$$

$$S_X = \sqrt{116} = 10.8 \text{ °F}$$

$$v_X = \frac{10.8}{80.3} = 0.134$$
(13.4%)

MODO = 86

$d_M = 9$, $L_M = 72.5$, $f_M = 7$, $F_M = 8$, $\frac{n}{2} = \frac{30}{2} = 15$

$$\text{MEDIANA} = M = 72.5 + \frac{15 - 8}{7}\, 9 = 72.5 + 9 = 81.5 \text{ °F}$$

IGUAL QUE LA DESVIACION ESTANDAR DE UNA MUESTRA SE DEFINE COMO LA RAIZ CUADRADA DE LA VARIANCIA, EL *ERROR ESTANDAR DE LA ESTIMACION O DE LA PREDICCION*, $s_{y|x}$, SE DEFINE COMO LA RAIZ CUADRADA DE LA VARIANCIA DE LA ESTIMACION, ES DECIR

$$s_{y|x} = \sqrt{s^2_{y|x}}$$

YA SE DIJO QUE LA DIFERENCIA $y_i - \tilde{y}_i$ REPRESENTA LA DESVIACION DE UNA ORDENADA OBSERVADA RESPECTO A SU ORDENADA PREDICHA MEDIANTE LA RECTA DE REGRESION. EXISTE OTRO TIPO DE DESVIACION: LA DE LAS ORDENADAS PREDICHAS MEDIANTE LA RECTA DE REGRESION, $\tilde{y}_i$, RESPECTO AL PROMEDIO ARITMETICO, $\bar{y}$, DE LAS ORDENADAS OBSERVADAS, y_i. ESTA DESVIACION, INDICADA COMO $\tilde{y}_i - \bar{y}$, SE LLAMA *DESVIACION EXPLICADA*, YA QUE DE LA ECUACION

$$\tilde{y}_i = mx_i + b = mx_i + \bar{y} - m\bar{x} = \bar{y} + m(x_i - \bar{x})$$

SE OBTIENE

$$\tilde{y}_i - \bar{y} = m(x_i - \bar{x})$$

LA CUAL INDICA QUE LAS DESVIACIONES DE $\tilde{y}_i$ RESPECTO A $\bar{y}$ SE EXPLICAN EXCLUSIVAMENTE POR (SON FUNCION DE) LA DESVIACION DE x_i RESPECTO A $\bar{x}$.

SI A LA DIFERENCIA $y_i - \bar{y}$ SE LE LLAMA *DESVIACION TOTAL* DE y_i CON RESPECTO AL PROMEDIO ARITMETICO, $\bar{y}$, LA ECUACION

$$y_i - \bar{y} = (\tilde{y}_i - \bar{y}) + (y_i - \tilde{y}_i)$$

INDICA QUE LA DESVIACION TOTAL ES IGUAL A LA DESVIACION EXPLICADA MAS $y_i - \tilde{y}_i$. LAS DESVIACIONES $y_i - \tilde{y}_i$ OCURREN AL AZAR, ES DECIR, EN FORMA INEXPLICABLE, DE AHI QUE SE LES CONOZCA CON EL NOMBRE DE

DESVIACIONES INEXPLICADAS (NO EXPLICADAS).

COMO CONSECUENCIA DE LA ECUACION ANTERIOR, SE PUEDE DEMOSTRAR QUE

$$\frac{\sum_{i=1}^{i=N} (y_i - \bar{y})^2}{N} = \frac{\sum_{i=1}^{i=N} (\tilde{y}_i - \bar{y})^2}{N} + \frac{\sum_{i=1}^{i=N} (y_i - \tilde{y}_i)^2}{N}$$

EL MIEMBRO IZQUIERDO DE ESTA ECUACION CORRESPONDE A LA VARIANCIA, $s^2(y)$, DE LOS DATOS DE y. EL SEGUNDO TERMINO DEL MIEMBRO DERECHO ES PRECISAMENTE LA VARIANCIA DE LA PREDICCION, $s^2_{y|x}$, CONOCIDA TAMBIEN COMO *VARIANCIA INEXPLICADA*. EL PRIMER TERMINO DEL MISMO MIEMBRO SE DENOMINA *VARIANCIA EXPLICADA*, Y SE REPRESENTA CON EL SIMBOLO $s^2(\tilde{y})$. EN CONSECUENCIA, SE PUEDE ESCRIBIR

$$s^2(y) = s^2(\tilde{y}) + s^2_{y|x}$$

MEDIDAS DE CORRELACION

CUANDO SE REALIZAN ESTUDIOS ESTADISTICOS EN QUE SE INVOLUCRAN DOS O MAS VARIABLES ES A MENUDO CONVENIENTE CONTAR CON UNA MEDIDA NUMERICA DEL GRADO DE ASOCIACION O RELACION QUE HAY ENTRE ELLAS.

UNA DE ESTAS MEDIDAS SE DENOMINA *COVARIANCIA*, S^2_{XY}, LA CUAL SE DEFINE COMO:

$$S^2_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \bar{x}\bar{y}$$

EN DONDE

(x_i, y_i) = PAREJAS DE DATOS

$\bar{x}$ = PROMEDIO DE LOS DATOS DE LA VARIABLE X

$\bar{y}$ = PROMEDIO DE LOS DATOS DE LA VARIABLE Y

N = TOTAL DE PAREJAS DE DATOS

OTRA MEDIDA DE CORRELACION, QUE RESULTA ADIMENSIONAL, ES EL *COEFICIENTE DE CORRELACION*, ρ_{XY}, QUE SE DEFINE COMO

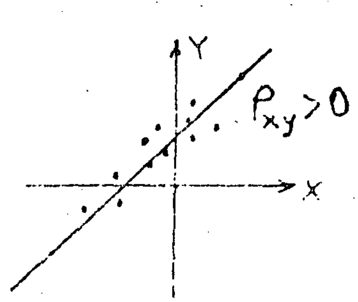
$$\rho_{xy} = \frac{S^2 xy}{S(x)\ S(y)}; \quad -1 \leq \rho_{xy} \leq 1$$

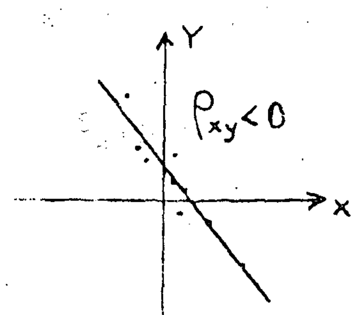
EN DONDE

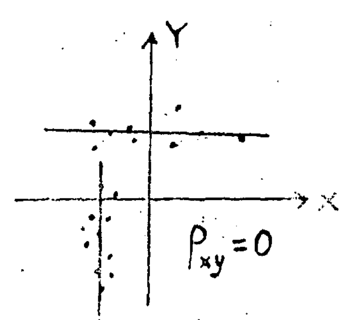
S^2_{XY} = COVARIANCIA ENTRE X y Y

S(x) = DESVIACION ESTANDAR DE LOS DATOS DE X

S(y) = DESVIACION ESTANDAR DE LOS DATOS DE Y







CASO DE CORRELACION PERFECTA

CUANDO SE PLANTEO EL METODO DE MINIMOS CUADRADOS PARA ESTIMAR LA RECTA DE REGRESION LINEAL ENTRE DOS VARIABLES, ESTE SE DESARROLLO SOBRE LA BASE DE HACER MINIMA LA SUMA DE LOS CUADRADOS DE LA DESVIACION VERTICAL DE CADA PUNTO RESPECTO A LA RECTA DE REGRESION, ESTO ES QUE

$$D = \sum_{i=1}^{N} (y_i - \tilde{y}_i)^2 = \text{MINIMO} \qquad (1)$$

EN DONDE

$$\tilde{y}_i = mx_i + b = mx_i + \bar{y} - m\bar{x} = \bar{y} + m(x_i - \bar{x}) \qquad (2)$$

SUSTITUYENDO A $\tilde{y}_i$ EN LA EC (1) Y AGRUPANDO TERMINOS SE OBTIENE

$$D = \sum_{i=1}^{N} \left[(y_i - \bar{y}) - m(x_i - \bar{x})\right]^2 \qquad (3)$$

OBSERVESE QUE D ES CERO SI, Y SOLO SI, CADA SUMANDO ES CERO, ES DECIR, SI

$$y_i - \bar{y} - m(x_i - \bar{x}) = 0 \quad \text{PARA TODO } i$$

PARA LO CUAL SE REQUIERE QUE *TODOS* LOS PUNTOS (x_i, y_i) QUEDEN SOBRE LA RECTA DE REGRESION, DADA POR LA EC (2), EN ESTE CASO SE DICE QUE LA *REGRESION ES PERFECTA.*

POR OTRA PARTE, DESARROLLANDO EL BINOMIO AL CUADRADO DE LA EC (3) OBTENEMOS

$$D = \sum_{i=1}^{N} \left[(y_i - \bar{y})^2 - 2m(y_i - \bar{y})(x_i - \bar{x}) + m^2(x_i - \bar{x})^2\right]$$

$$= NS^2(y) - 2mNS^2_{xy} + NS^2(x)m^2$$

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

FUNDAMENTOS DE LAS TECNICAS DE MUESTREO ESTADISTICO

E J E M P L O S

DR. OCTAVIO A. RASCON CHAVEZ

FEBRERO, 1979.

EJEMPLO

En un estudio sobre los nombres de las personas, se tomó una muestra aleatoria de 10 nombres de una colección de 100. Además del sexo, se consideraron las variables aleatorias Número de letras = Y, Número de vocales = Z, Número de consonantes = W. Los datos obtenidos son los siguientes:

Persona	Nombre	Nº de letras, Y	Nº de vocales, Z	Nº de consonantes, W	Sexo
1	Mario	5	3	2	M=0
2	Juan	4	2	2	M=0
3	Raúl	4	2	2	M=0
4	Gloria	6	3	3	F=1
5	Cosme	5	2	3	M=0
6	Carlos	6	2	4	M=0
7	Luz	3	1	2	F=1
8	Ernesto	7	3	4	M=0
9	Rosa	4	2	2	F=1
10	Norma	5	2	3	F=1

a. Estimar el número medio de letras por nombre; calcular el error estándar y el intervalo de confianza del 95%.

$$\hat{Y} = \bar{y} = \sum_{i=1}^{10} y_i / n = (5+4+4+6+5+6+3+7+4+5)/10 = 4.9$$

$$\hat{V}(\bar{y}) = (1-f)\hat{s}^2/n = \left(1-\frac{10}{100}\right)\frac{\sum_{i=1}^{10}(y_i-\bar{y})^2}{10-1}/10 = 0.129\left(\frac{\text{letras}}{\text{nombre}}\right)^2$$

$$\text{Error estándar} = \sqrt{\hat{V}(\bar{y})} = \sqrt{0.129} = 0.359\ \frac{\text{letras}}{\text{nombre}}$$

$$\text{Intervalo de confianza} = \bar{y} \mp t_c\sqrt{\hat{V}(\bar{y})} = 4.9 \mp 2.26 \times 0.359$$

(t_c se obtiene para $10-1=9$ grados de libertad)

$$\text{I. de C.} = \{4.09,\ 5.71\}$$

b. Estimar el número total de letras en los 100 nombres, U, tanto en forma puntual como con un intervalo de 90%.

$$\hat{U} = N\bar{y} = 100 \times 4.9 = 490 \text{ letras}$$

$$\text{I.C.} = 490 \mp \underbrace{t_c\sqrt{\hat{V}(N\bar{y})}}_{t_c\hat{\sigma}(N\bar{y})} = 490 \mp \underbrace{1.83}_{t_c} \times \underbrace{100 \times 0.359}_{\hat{\sigma}(N\bar{y})} = \{424.3,\ 555.7\}$$

c. Estimar el porcentaje de personas con sexo femenino y el error estándar de la estimación.

$$\hat{p} = p = 4/10 = 0.4 = 40\%$$

$$\hat{\sigma}(\hat{p}) = \sqrt{\frac{N-n}{(n-1)N}\hat{p}\hat{q}} = \sqrt{\frac{100-10}{(10-1)100}0.4 \times 0.6} = 0.155$$

El intervalo de confianza del 95% es

$$I.C. = 0.155 \mp 2.26 \times 0.155 = \{1.91, 2.61\}$$

d. Estimar el número de vocales por consonante.

$$\hat{R} = \frac{\bar{z}}{\bar{w}} = \frac{22/10}{27/10} = 0.81 \text{ vocales/consonante}$$

$$\hat{\sigma}^2(\hat{R}) = \frac{1-f}{n\bar{w}^2}\frac{\sum_{i=1}^{10}(z_i - \hat{R}w_i)^2}{n-1} = \frac{1-\frac{10}{100}}{10(2.7)^2}\frac{\sum_{i=1}^{10}(z_i - 0.81w_i)^2}{10-1}$$

$$= \frac{0.09}{(2.7)^2}\frac{\sum_{i=1}^{10} z_i^2 - 2 \times 0.81 \sum z_i w_i + (0.81)^2 \sum_{i=1}^{10} w_i^2}{9}$$

$$= \frac{0.09}{(2.7)^2}\frac{52 - 1.62(61) + (0.81)^2(79)}{9} = 0.006875$$

$$\hat{\sigma}(\hat{R}) = \sqrt{0.006875} = 0.0829$$

El intervalo de confianza del 95% es

$$I.C. = 0.81 \mp 1.83 \times 0.0829 = \{0.66, 0.96\}$$

EJEMPLO

En un estudio de mercadotecnia se tomó una muestra de 20 precios de carne en 20 tiendas distintas para estimar su variabilidad. Los datos arrojaron un promedio $\bar{x} = \$92.00$ y una desviación estándar $s = \$8.00$. Encontrar el 95% de nivel de confianza del 95 % de la varianza.

$$I.C. = \left\{ \frac{20(8)^2}{32.9}, \frac{20(8)^2}{8.91} \right\} = \{38.91, 143.66\} \$^2$$

El intervalo de confianza para la desviación estandar es

$$\left\{ \sqrt{38.91}, \sqrt{143.66} \right\} = \{6.24, 11.99\} \$$$

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

FUNDAMENTOS DE LAS TECNICAS DE MUESTREO ESTADISTICO

ESTIMACION DE TOTALES, PROPORCIONES Y RAZONES

DR. OCTAVIO A. RASCON CHAVEZ

22 FEBRERO, 1979.

ESTIMACION DE TOTALES

Puesto que $\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n} = \dfrac{y}{n} = \hat{\mu}$

es un estimador de la media, μ, de una población, entonces el total de las observaciones en la muestra es

$$y = x_1 + x_2 + \cdots + x_n = n\bar{x}$$

Por lo tanto, un estimador del total, Y, de la población

$$Y = X_1 + X_2 + \cdots + X_N$$

es $\qquad \hat{Y} = N\bar{x}$

en donde N = tamaño de la población

Este es un estimador insesgado ya que

$$E[\hat{Y}] = E[N\bar{x}] = N\,E[\bar{x}] = N\mu = Y$$

La varianza de $\hat{Y}$ es

$$\sigma^2(\hat{Y}) = \sigma^2(N\bar{x}) = N^2\sigma^2(\bar{x}) = N^2\frac{\sigma^2(x)}{n}(1-f)$$

en donde $f = n/N$ = proporción de muestreo.

El intervalo de confianza es

$$\left\{ N\bar{x} - Z_c\sigma(\hat{Y}),\ N\bar{x} + Z_c\sigma(\hat{Y}) \right\} \quad --- (1)$$

en donde Z_c es el valor crítico correspondiente al nivel de significancia α, el cual se obtiene de la distribución normal.

Si la muestra es pequeña, en vez de usar la distribución normal se usa la <u>t de Student</u> y, en tal caso, el intervalo de confianza es

$$\left\{ N\bar{x} - t_c\sqrt{N^2(1-f)}\,\frac{S(x)}{\sqrt{n-1}},\; N\bar{x} + t_c N\sqrt{1-f}\,\frac{S(x)}{\sqrt{n-1}} \right\}$$

en donde $S(x)$ es la desviación estándar de la muestra

$$S(x) = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

y t_c es el valor crítico correspondiente a un nivel de significancia α, que se obtiene de la distribución t de Student.

En la ecuación (1), $\sigma(\hat{Y})$ se estima usando $\sigma(\hat{Y}) = S(\hat{Y}) = N S(x)\sqrt{\frac{1-f}{n}}$

o sea, usando a $S(x)$ como estimador de $\sigma(x)$.

ESTIMACION DE PROPORCIONES

Las proporciones se estiman mediante las frecuencias relativas correspondientes con que aparece la observación de interés en la muestra. Así, si p es la proporción que es preciso estimar, su estimador, $\hat{p}$, será

$$\hat{p} = p = \frac{n(a)}{n} \times 100 \quad \text{, en porcentaje}$$

La variancia de p es

$$\sigma^2(p) = \frac{NPQ}{N-1} \; \frac{1-f}{N} \quad \text{, donde } Q = 1-P$$

Se puede demostrar que el estimador de $\sigma^2(p)$ es

$$\hat{\sigma}^2(p) = \hat{V}(p) = \frac{N-n}{(n-1)N} pq \text{, donde } q = 1-p$$

Asimismo, el estimador del <u>total</u> de elementos que tiene la población <u>con la característica</u> en cuestión es

$$\hat{A} = N\hat{p}$$

La variancia de este estimador es

$$\sigma^2(\hat{A}) = V(\hat{A}) = \frac{N^3PQ}{N-1} \; \frac{1-f}{n}$$

la cual se estima con $\hat{V}(\hat{A}) = \frac{N(N-n)}{n-1} pq$.

Para muestras grandes el intervalo de confianza es:

$$\left\{ \hat{A} - z_c\sqrt{\hat{V}(\hat{A})} \, , \, \hat{A} + z_c\sqrt{\hat{V}(\hat{A})} \right\}$$

ESTIMACION DE RAZONES O COCIENTES

Sean W y X dos variables aleatorias. La razón de la primera respecto a la segunda se estima como:

$$\hat{R} = \frac{\bar{W}}{\bar{X}}$$

Se puede demostrar que la varianza de este estimador es

$$V(\hat{R}) = \sigma^2(\hat{R}) \doteq \frac{1-f}{n\bar{X}^2} \frac{\sum_{i=1}^{N}(w_i - R y_i)^2}{N-1}$$

el cual se estima con

$$\hat{V}(\hat{R}) = \frac{1-f}{n\bar{x}^2} \frac{\sum_{i=1}^{n}(w_i - \hat{R} y_i)^2}{n-1}$$

Para muestras grandes, el intervalo de confianza es

$$\left\{\hat{R} - z_c \hat{V}(\hat{R}),\ \hat{R} + z_c \hat{V}(\hat{R})\right\}$$

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

FUNDAMENTOS DE LAS TECNICAS DE MUESTREO ESTADISTICO

INFERENCIA ESTADISTICA

M. EN I. AUGUSTO VILLARREAL ARANDA

FEBRERO, 1979

INFERENCIA ESTADISTICA

Por: M en I Augusto Villarreal Aranda*

1. Introducción

La parte de la estadística que proporciona las reglas para inferir ciertas características de una población a partir de muestras extraídas de ella, junto con indicaciones probabilísticas de la veracidad de tales inferencias, se llama *inferencia estadística*.

En la inferencia estadística se estudian las relaciones existentes entre una población, las muestras obtenidas de ella, y las técnicas para estimar parámetros, tales como la media y la variancia, o bien para determinar si las diferencias entre dos muestras son debidas al azar, etc.

2. Distribuciones muestrales

Si se consideran todas las muestras posibles de tamaño

* *Secretario Académico*, División de Estudios Superiores, Facultad de Ingeniería, UNAM y *Profesor investigador*, Instituto de Ingeniería, UNAM

n que pueden extraerse de una población, y para cada una se calcula el valor del promedio aritmético, este seguramente variará de una muestra a otra, ya que depende de los valores de los datos que se hayan obtenido en cada muestra. Por lo tanto, el promedio aritmético es en sí una variable aleatoria, como también lo son, por la misma razón, el rango y la variancia de la muestra.

A todo elemento que es función de los valores de los datos que se tienen en una muestra se le denomina *estadística;* toda estadística es, entonces, una variable aleatoria cuya distribución de probabilidades se conoce como *distribución muestral*. Si, por ejemplo, la estadística considerada es la variancia de la muestra, su densidad de probabilidades se llama *distribución muestral de la variancia*.

En forma similar se pueden obtener las distribuciones muestrales de la desviación estándar, del rango, etc., cada una de las cuales tendrá sus propios parámetros, lo que permite hablar de la media y la desviación estándar de la variancia, etc.

3. Muestreo con y sin remplazo

Cuando se efectúa un muestreo en una población de tal manera que cada elemento de la misma se pueda escoger más de una vez, se dice que el muestreo es *con remplazo*; en caso contrario, el muestreo es *sin remplazo*. Si de una urna se quiere extraer una muestra de bolas de colores, se puede proceder de dos maneras: se saca al azar una bola, se anota su color y se regresa a la urna antes de obtener otra, y así sucesivamente; en este caso el muestreo es *con remplazo*. La segunda forma consiste en extraer al azar todas las bolas que constituyen la muestra sin regresarlas a la urna, siendo entonces un muestreo *sin remplazo*.

4. Distribucion muestral del promedio aritmético

Supóngase que se extraen sin remplazo todas las muestras posibles de tamaño n de una población finita de tamaño $N_p > n$. Si la media y la desviación estándar de la distribución muestral del promedio aritmético se denotan con $\mu_{\bar{X}}$ y $\sigma_{\bar{X}}$, y la media y la desviación estándar de la población con μ y σ, respectivamente, entonces es posible demostrar que se cumplen las siguientes ecuaciones

$$\mu_{\bar{X}} = \mu$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}}$$

Además, si la población es infinita (o el muestreo es con remplazo), los resultados anteriores se reducen a

$$\mu_{\bar{X}} = \mu$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

puesto que

$$\lim_{N_p \to \infty} \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}} = \frac{\sigma}{\sqrt{n}}$$

Para valores grandes de $n(n \geqslant 30)$ se demuestra, empleando el teorema del límite central, que la distribución muestral del promedio aritmético es aproximadamente una distribución normal con media $\mu_{\bar{X}}$ y desviación estándar $\sigma_{\bar{X}}$, independientemente de cuál sea la densidad de probabilidades de X, la variable aleatoria asociada a la población. Si esta variable tiene distribución normal, la distribución muestral del promedio aritmético también es normal, aun para valores pequeños de n $(n < 30)$.

Ejemplo 4.1

Supóngase que se tiene una población finita formada por los datos 1,2,3,4,5. Se desea conocer la media y la desviación estándar de la distribución muestral del promedio aritmético, considerando las muestras de tamaño 3 obtenidas sin remplazo.

Primer procedimiento.

Siendo la población finita y el muestreo sin remplazo, es posible obtener la distribución muestral correspondiente para calcular después sus parámetros, considerando que el número total de muestras distintas de tamaño 3 que pueden obtenerse a partir de una población de 5 elementos es

$$\frac{5!}{3!(5-3)!} = 10$$

Dichas muestras son las siguientes, junto con sus promedios aritméticos correspondientes:

	$\bar{X}_i$		$\bar{X}_i$
1, 2, 3	6/3	3, 4, 5	12/3
1, 2, 4	7/3	3, 4, 1	8/3
1, 2, 5	8/3	4, 5, 1	10/3
2, 3, 4	9/3	4, 5, 2	11/3
2, 3, 5	10/3	5, 1, 3	9/3

Para calcular la media y la desviación estándar, se emplea la siguiente tabla

$\bar{X}_i$	6/3	7/3	8/3	8/3	9/3	9/3	10/3	10/3	11/3	12/3
$\bar{X}_i^2$	36/9	49/9	64/9	64/9	81/9	81/9	100/9	100/9	121/9	144/9

$$\sum_{i=1}^{10} \bar{X}_i = 90/3 \qquad\qquad \sum_{i=1}^{10} \bar{X}_i^2 = 840/9$$

$$\mu_{\bar{X}} = \bar{\bar{X}} = \frac{1}{10} \sum_{i=1}^{10} \bar{X}_i = \frac{1}{10} \cdot \frac{90}{3} = 3$$

$$\sigma_{\bar{X}}^2 = \frac{1}{10} \sum_{i=1}^{10} \bar{X}_i^2 - \bar{\bar{X}}^2 = \frac{1}{10} \cdot \frac{840}{9} - (3)^2 =$$

$$= 9.333 - 9.000 = 0.333 \Rightarrow \sigma_{\bar{X}} = \sqrt{0.333} = 0.577$$

Es decir, $\mu_{\bar{X}} = 3$ y $\sigma_{\bar{X}} = 0.577$

Segundo procedimiento.

Por tratarse de una población finita, se verifica que

$$\mu_{\bar{X}} = \mu \qquad \text{y} \qquad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$$

en donde $N_p = 5$, $n = 3$ y $\mu = 3$.

El valor de σ^2 de la población es

$$\sigma^2 = \frac{1+4+9+16+25}{5} - (3)^2 = \frac{55}{5} - 9 = 11-9 = 2$$

Por lo tanto, $\sigma = \sqrt{2} = 1.4145$ y

$$\sigma_{\bar{X}} = \frac{1.4145}{\sqrt{3}} \sqrt{\frac{5-3}{5-1}} = (0.8164)(0.7071) = 0.577$$

Es decir, $\mu_{\bar{X}} = 3$ y $\sigma_{\bar{X}} = 0.577$

Comparando los resultados, se puede observar que ambos procedimientos conducen a la obtención de los mismos valores de $\mu_{\bar{X}}$ y $\sigma_{\bar{X}}$ para la distribución muestral del promedio aritmético.

Ejemplo 4.2

En una bodega se tienen cinco mil varillas de acero; el valor medio del peso, X, de cada varilla es de 5.02 kg, y la desviación estándar 0.3 kg. Hallar la probabilidad de que una muestra de cien varillas, escogida al azar, tenga un peso total

a. entre 496 y 500 kg

b. de más de 510 kg.

Para la distribución muestral del promedio, se tiene que $\mu_{\bar{X}} = \mu = 5.02$ kg y, por tratarse de una población finita,

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}} = \frac{0.30}{\sqrt{100}} \sqrt{\frac{5000 - 100}{5000 - 1}} = 0.027$$

a. El peso total de la muestra estará entre 496 y 500 kg si el peso promedio de las cien varillas se encuentra entre 4.96 y 5.00 kg. Puesto que la muestra es mayor de 30 elementos se puede considerar como aproximadamente normal a la distribución muestral, y los valores estándar correspondientes a $\bar{X} = 4.96$ y a $\bar{X} = 5.00$ se obtienen mediante la transformación

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$$

es decir,

$$Z_1 = \frac{4.96 - 5.02}{0.027} = -2.22$$

$$Z_2 = \frac{5.00 - 5.02}{0.027} = -0.74$$

En la fig 4.1 se puede apreciar que

$$P[496 \leqslant X \leqslant 500] = P[-2.22 \leqslant Z \leqslant -0.74] =$$

$$= P[-2.22 \leqslant Z \leqslant 0] - P[-0.74 \leqslant Z \leqslant 0]$$

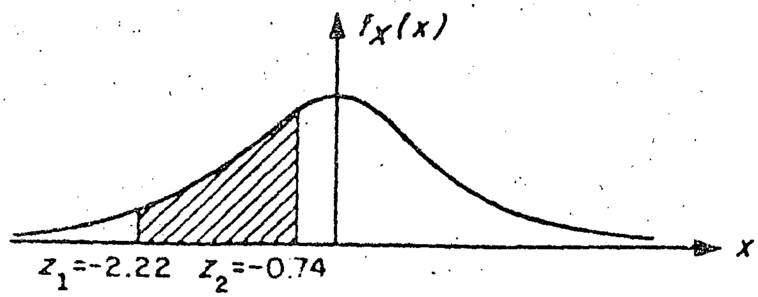


Fig 4.1 Distribución normal correspondiente al ejemplo

Recurriendo a la tabla de áreas bajo la curva normal estándar entre 0 y Z queda finalmente

$$P[496 \leq X \leq 500] = 0.4868 - 0.2704 = 0.2164$$

b. El peso total de la muestra excederá de 510 kg si el peso promedio de las cien varillas pasa de 5.10 kg.

Estandarizando dicho valor, queda

$$Z_3 = \frac{5.10 - 5.02}{0.027} = 2.96$$

Calculando el área bajo la curva normal a la derecha de este valor (fig 4.2), se tiene que

$$P[X \geq 510] = P[Z \geq 2.96] = P[Z \geq 0] - P[0 \leq Z \leq 2.96] =$$

$$= 0.5 - 0.4985 = 0.0015$$

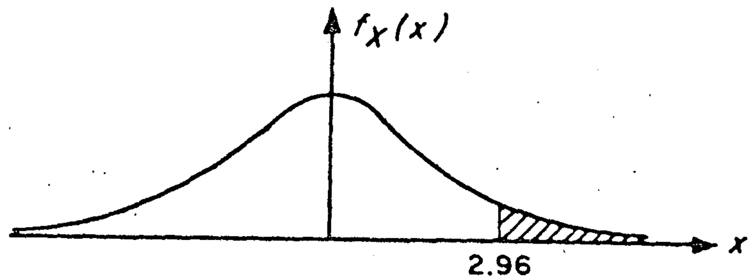


Fig 4.2 Distribución normal correspondiente al ejemplo

5. Distribución muestral de diferencias de promedios aritméticos

Con frecuencia se presenta el caso en el que se tienen datos de dos poblaciones con variables aleatorias asociadas X y Y, respectivamente, surgiendo la duda de si estas se pueden considerar como una sola, es decir, si X = Y. Para probar estadísticamente esta hipótesis (como se verá más adelante), es necesario obtener las distribuciones muestrales de la diferencia de los promedios y de las variancias de las muestras de ambas variables.

Sean $\bar{X}$ y $\bar{Y}$ los promedios aritméticos obtenidos de muestras aleatorias de tamaño n_X y n_Y de dos poblaciones con características X y Y, respectivamente. Se puede demostrar que la distribución muestral de la diferencia de los promedios correspondientes a poblaciones infinitas con medias μ_X y μ_Y y desviaciones estándar σ_X y σ_Y, tiene los siguientes parámetros:

$$\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} = \mu_X - \mu_Y$$

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\sigma_{\bar{X}}^2 + \sigma_{\bar{Y}}^2} = \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

si las muestras son independientes.

Esta distribución también es aplicable a poblaciones finitas si el muestreo es con remplazo. Para el caso de poblaciones finitas en las cuales el muestreo se hace sin remplazo, los parámetros de la distribución muestral de la diferencia de los promedios aritméticos son

$$\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} = \mu_X - \mu_Y$$

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}}} = \sqrt{\frac{\sigma^2_X}{n_X}\,\frac{N_X - n_X}{N_X - 1} + \frac{\sigma^2_X}{n_Y}\,\frac{N_Y - n_Y}{N_Y - 1}}$$

suponiendo que las muestras sean independientes.

Ejemplo 5.1

Considérese que de una población X se obtienen tres muestras posibles, cuyos correspondientes promedios aritméticos son 3, 7 y 8. De otra población Y se extraen dos muestras posibles, con promedios 2 y 4, respectivamente. Se deben obtener los parámetros de la distribución muestral de las diferencias de los promedios aritméticos.

Primer procedimiento

Todas las posibles diferencias de promedios aritméticos de X con los de Y serían

$$\begin{array}{ccc} 3-2 & 7-2 & 8-2 \\ 3-4 & 7-4 & 8-4 \end{array} \Longrightarrow \begin{array}{ccc} 1 & 5 & 6 \\ -1 & 3 & 4 \end{array}$$

Es decir,

$$\mu_{\bar{X}-\bar{Y}} = \frac{-1+1+3+4+5+6}{6} = \frac{18}{6} = 3$$

$$\sigma^2_{\bar{X}-\bar{Y}} = \frac{(-1-3)^2 + (1-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2 + (6-3)^2}{6} =$$

$$= \frac{34}{6} = \frac{17}{3}$$

Segundo procedimiento

Se sabe que

$$\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} \quad ; \quad \sigma^2_{\bar{X}-\bar{Y}} = \sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}}$$

Por ello,

$$\mu_{\bar{X}} = \frac{3+7+8}{3} = \frac{18}{3} = 6$$

$$\mu_{\bar{Y}} = \frac{2+4}{2} = \frac{6}{2} = 3$$

$$\sigma^2_{\bar{X}} = \frac{(3-6)^2 + (7-6)^2 + (8-6)^2}{3} = \frac{14}{3}$$

$$\sigma^2_{\bar{Y}} = \frac{(2-3)^2 + (4-3)^2}{2} = \frac{2}{2} = 1$$

$$\mu_{\bar{X}-\bar{Y}} = 6 - 3 = 3$$

$$\sigma^2_{\bar{X}-\bar{Y}} = \frac{14}{3} + 1 = \frac{17}{3}$$

Se observa que ambos procedimientos conducen a los mismos resultados.

Ejemplo 5.2

Las varillas de acero que fabrica una compañía A tienen un peso medio de 6.5 kg y una desviación estándar de 0.4, en tanto que las producidas por una empresa B tienen un peso medio de 6.3 kg y una desviación estándar de 0.3 kg. Si se toman muestras aleatorias de 100 varillas de cada fábrica, ¿cuál es la probabilidad de que las de la compañía A tengan un peso promedio de por lo menos

a. 0.35 kg

b. 0.10 kg

mayor que el de la compañía B?

Se puede suponer en este caso que las distribuciones muestrales involucradas son normales, en virtud de que el tamaño de ambas muestras es mayor de 30 elementos. También se puede suponer que ambas poblaciones son infinitas, y siendo $\bar{X}_A$ y $\bar{X}_B$ los pesos promedios de las muestras de las fábricas A y B, respectivamente, entonces

$$\mu_{\bar{X}_A - \bar{X}_B} = \mu_{\bar{X}_A} - \mu_{\bar{X}_B} = 6.5 - 6.3 = 0.20 \text{ kg}$$

$$\sigma_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = \sqrt{\frac{(0.4)^2}{100} + \frac{(0.3)^2}{100}} = 0.05 \text{ kg}$$

La variable estandarizada de la diferencia de los promedios es

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - \mu_{\bar{X}_A - \bar{X}_B}}{\sigma_{\bar{X}_A - \bar{X}_B}} = \frac{(\bar{X}_A - \bar{X}_B) - 0.20}{0.05}$$

a. Estandarizando la diferencia de 0.35 kg se llega a

$$Z_1 = \frac{0.35 - 0.20}{0.05} = \frac{0.15}{0.05} = 3$$

La probabilidad deseada es el área bajo la curva normal a la derecha de $Z = 3$, es decir

$$P\,[\bar{X}_A \geq \bar{X}_B + 0.35] = P\,[Z \geq 3] = 0.500 - 0.4987 = 0.0013$$

b. Al estandarizar la diferencia de 0.10 kg, la variable Z resulta

$$Z_2 = \frac{0.10 - 0.20}{0.05} = \frac{-0.1}{0.05} = -2$$

La probabilidad requerida es el área bajo la curva normal a la derecha de $Z= -2$, es decir

$$P[\bar{X}_A \geq \bar{X}_B + 0.10] = P[Z \geq -2] = 0.5 + 0.4772 = 0.9772$$

6. Teoría estadística de la estimación

En la práctica profesional a menudo resulta necesario inferir información acerca de una población mediante el uso de muestras extraídas de ella; una parte básica de dicha inferencia consiste en *estimar* los valores de los parámetros de la población (media, variancia, etc.) a partir de las estadísticas correspondientes de la muestra, como se explica a continuación.

7. Estimadores puntuales. Clasificación

Si un estimador de un parámetro de la población consiste en un solo valor de una estadística, se le conoce como *estimador puntual* del parámetro.

Cuando la media de la distribución muestral de una estadística es igual al parámetro que se está estimando de la población, entonces la estadística se conoce como *estimador insesgado* del parámetro; si no sucede así, entonces se denomina *estimador sesgado*. Ambos estimadores son puntuales, y sus valores correspondientes se llaman estimaciones insesgadas o sesgadas, respectivamente. Dicho de otra manera, si S es una estadística cuya distribución muestral tiene media μ_S, y el parámetro correspondiente de la población es θ, se dice que S es un estimador insesgado de θ si

$$\mu_S = \theta$$

Por otra parte, si la estadística S_n de la muestra tiende a ser igual al parámetro θ de la población a medida que se hace más grande el tamaño de la muestra, entonces la estadística recibe el nombre de *estimador consistente* del parámetro.

Empleando símbolos, si

$$\lim_{n\to\infty} S_n = \theta$$

resulta que la estadística S_n es un estimador consistente. Por ejemplo, el promedio aritmético es un estimador insesgado y consistente de la media, y la variancia de la muestra es un estimador sesgado y consistente de la variancia de la población.

Si las distribuciones muestrales de varias estadísticas tienen el mismo valor de la media, se dice que la estadística que cuenta con la menor variancia es un *estimador eficiente* de dicha media, en tanto que las estadísticas restantes se conocen como *estimadores ineficientes* del parámetro.

Por ejemplo, las distribuciones muestrales del promedio aritmético y de la mediana cuentan con medias que son, en ambos casos, iguales a la media de la población. Sin embargo, la variancia de la distribución muestral del promedio aritmético es menor que la de la distribución de la mediana, por lo que el promedio aritmético obtenido de una muestra aleatoria proporciona un estimador eficiente de la media de la población, en tanto que la mediana obtenida de la muestra proporciona un estimador ineficiente de dicho parámetro.

8. Estimación de intervalos de confianza para los parámetros de una población

La estimación de un parámetro de una población mediante un par de números entre los cuales se encuentra, con cierta probabilidad, el valor de dicho parámetro, se llama estimación del intervalo del mismo.

Sea S una estadística obtenida de una muestra de tamaño n para estimar el valor del parámetro θ, y sea σ_S la desviación estándar (conocida o estimada) de su distribución muestral. La probabilidad, $1 - \alpha$, de que el valor de θ se localice en el intervalo de $S - z_c \sigma_S$ a $S + z_c \sigma_S$, donde z_c es una constante, se escribe en la forma

$$P[S - z_c \sigma_S \leqslant \theta \leqslant S + z_c \sigma_S] = 1 - \alpha$$

Si se fija el valor de $1 - \alpha$, se puede obtener el valor de z_c necesario para que se satisfaga la ecuación anterior, con lo cual queda definido el *intervalo de confianza* del parámetro θ, $(S - z_c \sigma_S, S + z_c \sigma_S)$, correspondiente al nivel de *confianza* $1 - \alpha$.

La constante z_c que fija el intervalo de confianza se conoce como *valor crítico*. Si la distribución de S es normal, el valor de z_c correspondiente a uno de α se obtiene de la tabla de áreas bajo la curva normal o de la tabla 8.1 siguiente.

TABLA 8.1 VALORES DE z_c PARA DISTINTOS NIVELES DE CONFIANZA

Nivel de confianza, en porcentaje	z_c
99.73	3.00
99.00	2.58
98.00	2.33
96.00	2.05
95.45	2.00
95.00	1.96
90.00	1.64
80.00	1.28
68.27	1.00
50.00	0.674

Ejemplo 8.1

Sea el promedio aritmético $\bar{X}$ una estadística con distribución normal. Las probabilidades o niveles de confianza de que $\mu_{\bar{X}}$ (o μ de la población) se encuentre localizada entre los límites $\bar{X} \pm \sigma_{\bar{X}}$, $\bar{X} \pm 2\ \sigma_{\bar{X}}$ y $\bar{X} \pm 3\ \sigma_{\bar{X}}$ son 68.26, 95.44 y 99.73%, respectivamente, obteniéndose dichos valores de la tabla de áreas bajo la curva normal. Lo anterior significa que el intervalo $\bar{X} \pm 3\ \sigma_{\bar{X}}$ contendrá a $\mu_{\bar{X}}$ en el 99.73 por ciento de las muestras de tamaño n, por lo que los intervalos de confianza de 68.26, 95.44 y 99.73 por ciento para estimar a μ son $(\bar{X} - \sigma_{\bar{X}}, \bar{X} + \sigma_{\bar{X}})$ $(\bar{X} - 2\ \sigma_{\bar{X}}, \bar{X} + 2\ \sigma_{\bar{X}})$ y $(\bar{X} - 3\ \sigma_{\bar{X}}, \bar{X} + 3\ \sigma_{\bar{X}})$, lo cual se aprecia en la *fig* 8.1 siguiente.

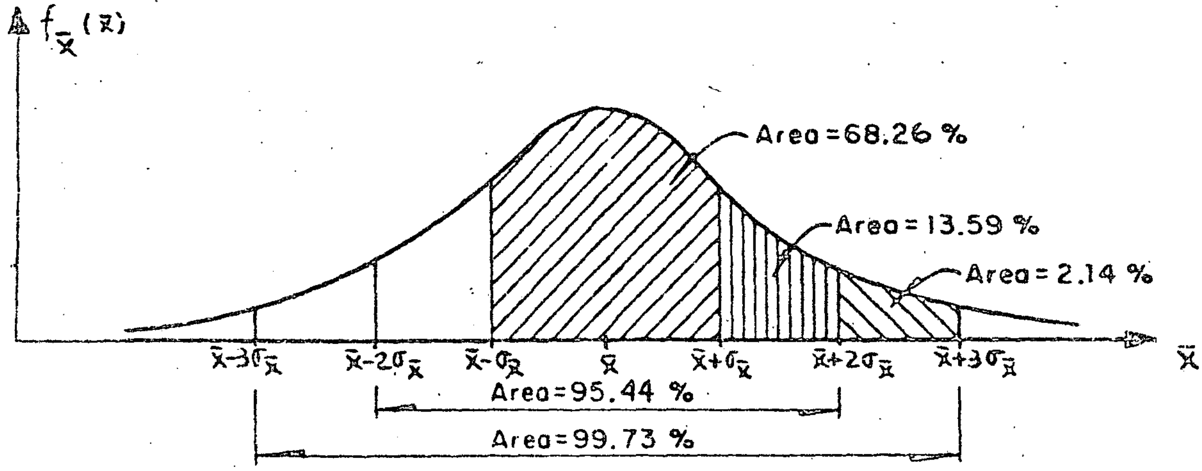


Fig 8.1

9. Estimación de intervalos de confianza para la media

Los límites de confianza para la media de una población con variable aleatoria X asociada están dados por

$$\bar{X} \pm z_c \, \sigma_{\bar{X}}$$

en donde z_c depende del nivel de confianza deseado. Si $\bar{X}$ tiene distribución normal, z_c puede obtenerse en forma directa de la tabla 8.1. Por ejemplo, los límites de confianza de 95 y 99 por ciento para estimar la media, μ, de la población son $\bar{X} \pm 1.96\sigma_{\bar{X}}$ y $\bar{X} \pm 2.58\,\sigma_{\bar{X}}$, respectivamente. Al obtener estos límites hay que usar el valor calculado de $\bar{X}$ para la muestra correspondiente.

Entonces, los límites de confianza para la media de la población quedan dados por

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{n}}$$

en caso de que el muestreo se haga a partir de una población infinita o de que se efectúe con remplazo a partir de una población finita, o por

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$$

si el muestreo es sin remplazo a partir de una población finita de tamaño N_p.

Ejemplo 9.1

Las mediciones de los diámetros de una muestra aleatoria de 100 tubos de albañal mostraron una media de 32 cm y una desviación estándar de 2 cm. Obténganse los límites de confianza de

a. 95 por ciento
b. 97 por ciento

para el diámetro medio de todos los tubos.

a. De la tabla 8.1, los límites de confianza del 95 por ciento son

$$\bar{X} \pm 1.96\sigma/\sqrt{n} = 32 \pm 1.96(2/\sqrt{100}) = 32 \pm 0.392 \text{ cm}$$

o sea 31.608 y 32.392, en donde se ha empleado el valor de S_X para estimar el de σ de la población, puesto que la muestra es suficientemente grande (mayor de 30 elementos). Esto significa que con una probabilidad de 95 por ciento, el valor de μ_X se encuentra entre 31.608 y 32.392 cm.

b. Si $Z = z_c$ es tal que el área bajo la curva normal a la derecha de z_c es el 1.5 por ciento del área total, entonces el área entre 0 y z_c es 0.5 - 0.015 = 0.485, por lo que de la tabla de áreas bajo la curva normal se obtiene $z_c = 2.17$. Por lo tanto, los límites de confianza del 97 por ciento son:

$$\bar{X} \pm 2.17\sigma/\sqrt{n} = 32 \pm 2.17(2/\sqrt{100}) = 32 \pm 0.434 \text{ cm}$$

y el intervalo de confianza respectivo es (31.566 cm, 32.434 cm).

Ejemplo 9.2

Una muestra aleatoria de 50 calificaciones de cierto examen de admisión tiene un promedio aritmético de 72 puntos, con desviación estándar igual a 10. Si el examen se aplicó a 1018 personas, obtener

a. El intervalo de confianza del 95% para la media del total de calificaciones.

b. El tamaño de muestra necesario para que el error en la estimación de la media no exceda de 2 puntos, considerando el mismo nivel de confianza.

c. El nivel de confianza para el cual la media de la población sea 72 ± 1 puntos.

a. Si se estima a σ de la población con S_X de la muestra y se considera que la población es finita, los límites de confianza son, puesto que $\bar{X} = 72$, $Z_c = 1.96$, $S_X = 10$, $N_p = 1018$ y $n = 50$,

$$72 \pm 1.96 \frac{10}{\sqrt{50}} \sqrt{\frac{1018 - 50}{1018 - 1}}$$

$$72 \pm 1.96 \; (1.4142) \; (0.9755)$$

$$72 \pm 2.704$$

y el intervalo de confianza respectivo es

$$(69.296, \quad 74.704)$$

b. Puesto que el error en la estimación de la media es, para población finita,

$$\text{Error en la estimación} = Z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$$

en este caso se tendría

$$Z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}} < 2$$

o sea, para un nivel de confianza de 95%,

$$1.96 \frac{10}{\sqrt{n}} \sqrt{\frac{1018 - n}{1018 - 1}} < 2$$

$$\frac{19.6}{\sqrt{n}} \sqrt{\frac{1018 - n}{1018 - 1}} < 2$$

Elevando al cuadrado la desigualdad, queda

$$\frac{384.16}{n} \quad \frac{1018 - n}{1017} < 4$$

o sea

$$87.85 < n$$

Por lo cual, se requieren al menos 88 elementos en la muestra para que el error en la estimación no exceda de 2 puntos, para $1 - \alpha = 0.95$.

c. Los límites de confianza son, en este caso

$$72 \pm Z_c \frac{10}{\sqrt{50}} \sqrt{\frac{1018 - 50}{1018 - 1}}$$

$$72 \pm Z_c \ (1.4142)(0.9755)$$

o sea

$$72 \pm 1.3795 \ Z_c$$

Puesto que se desea que el valor de la media sea 72 ± 1 puntos, se verifica que

$$1 = 1.3795 \ Z_c$$

Es decir

$$Z_c = \frac{1}{1.3795} = 0.725$$

El área bajo la curva normal estándar entre 0 y $Z_c = 0.725$ es, por interpolación lineal, igual a 0.2657. Por lo tanto, el nivel de confianza es igual al doble del área anterior, es decir, $2(0.2657) = 0.5314$ (ó 53.14%), tal como se muestra en la *fig* 9.1.

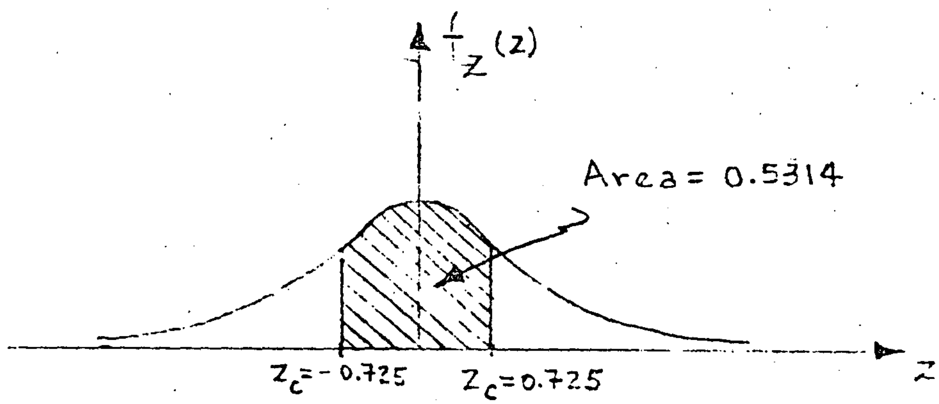


Fig 9.1

10. Intervalos de confianza para diferencias de medias

Los límites de confianza para la diferencia de las medias cuando las poblaciones X y Y son infinitas, o cuando el muestreo se realiza con remplazo de poblaciones finitas, se encuentran dados por

$$\bar{X} - \bar{Y} \pm Z_c \, \sigma_{\bar{X} - \bar{Y}} = \bar{X} - \bar{Y} \pm z_c \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

en donde $\bar{X}$, n_X y $\bar{Y}$, n_Y son los respectivos promedios aritméticos y tamaños de las dos muestras extraídas de las poblaciones, y σ_X y σ_Y las desviaciones estándar de estas últimas.

En el caso de que las poblaciones X y Y sean finitas y el muestreo sin remplazo, los límites de confianza son

$$\bar{X} - \bar{Y} \pm z_c \sigma_{\bar{X}-\bar{Y}} = \bar{X} - \bar{Y} \pm z_c \sqrt{\frac{\sigma_X^2}{n_X} \, \frac{N_X - n_X}{N_X - 1} + \frac{\sigma_Y^2}{n_Y} \, \frac{N_Y - n_Y}{N_Y - 1}}$$

en donde N_X y N_Y son los tamaños de las poblaciones X y Y, respectivamente.

Las dos ecuaciones anteriores son válidas únicamente si las muestras aleatorias seleccionadas son independientes.

Ejemplo 10.1

Para el ejemplo de las varillas tratado anteriormente (5.2), encontrar el intervalo de confianza del 95.45% para las diferencias de las medias de las poblaciones.

Siendo $\bar{X}_A = \mu_A = 6.5$ kg, $\sigma_A = 0.4$ kg, $\bar{X}_B = \mu_B = 6.3$ kg, $\sigma_B = 0.3$ kg y $n_A = n_B = 100$, los límites de confianza para la diferencia de las medias son, empleando la tabla 8.1

$$\bar{X}_A - \bar{X}_B \pm z_c \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = 6.5 - 6.3 \pm 2 \sqrt{\frac{(0.4)^2}{100} + \frac{(0.3)^2}{100}} =$$

$$= 0.2 \pm 0.1$$

Por lo tanto, el intervalo de confianza respectivo es (0.1, 0.3).

Ejemplo 10.2

Se tienen en una bodega 3000 focos de marca X, y 5000 de marca Y. Se extrae una muestra aleatoria de 150 focos de la marca X, y se obtiene una duración promedio de 1400 horas, con desviación estándar igual a 120 horas. Otra muestra aleatoria de 200 focos de la marca Y tuvo una duración promedio de 1200 horas, con desviación estándar igual a 80 horas. Obtener intervalos de confianza de

a. 95%

b. 99%

para la diferencia de los tiempos medios de duración de los focos de ambas marcas.

a. Puesto que se trata de poblaciones finitas y $\bar{X} = 1400$ h, $S_X = 120$ h, $N_X = 3000$, $n_X = 150$, $\bar{Y} = 1200$ h, $S_Y = 80$ h, $N_Y = 5000$ y $n_Y = 200$, se obtiene, estimando a σ_X y σ_Y con S_X y S_Y , respectivamente

$$1400 - 1200 \pm 1.96 \sqrt{\frac{(120)^2}{150} \; \frac{3000 - 150}{3000 - 1} + \frac{(80)^2}{200} \; \frac{5000 - 200}{5000 - 1}}$$

$$200 \pm 1.96 \; (11.04)$$

$$200 \pm 21.638$$

o sea, (178.362, 221.638), puesto que de la tabla 8.1, para un nivel de confianza de 95%, $Z_c = 1.96$.

b. En este caso, al emplear la tabla 8.1 se obtiene

$Z_c = 2.58$ para un nivel de confianza de 99%, por lo cual los límites son

$$1400 - 1200 \pm 2.58 \sqrt{\frac{(120)^2}{150}\,\frac{3000 - 150}{3000 - 1} + \frac{(80)^2}{200}\,\frac{5000 - 2000}{5000 - 1}}$$

$$200 \pm 2.58\ (11.04)$$

$$200 \pm 28.483$$

y el intervalo de confianza es

$$(171.517,\ 228.483)$$

11. Pruebas de hipótesis

Supóngase que una empresa armadora de automóviles está en la disyuntiva de emplear una nueva marca de bujías en sus unidades o la que regularmente utiliza, y que su departamento de control de calidad debe decidir, con base en la información de las muestras de las dos marcas distintas. Las decisiones de este tipo, es decir, que se basan en estudios estadísticos, reciben el nombre de *decisiones estadísticas*, y a los procedimientos que permiten decidir si se acepta o rechaza una hipótesis se les llama *pruebas de hipótesis*, *pruebas de significancia* o *reglas de decisión*.

Al tomar decisiones estadísticas, es necesario postular las diversas alternativas o cursos de acción que pueden adoptarse.

En el caso particular de una *prueba de hipótesis* solamente se tienen dos cursos de acción posibles, los que se denotarán como H_0 y H_1. A la acción H_0 se le llama *hipótesis nula*, y a la H_1, *hipótesis alternativa*. Por ejemplo, si la hipótesis nula establece que $\mu_1 = \mu_2$, la hipótesis alternativa puede ser una de las siguientes:

$$\mu_1 > \mu_2, \mu_1 < \mu_2 \text{ o } \mu_1 \neq \mu_2$$

Al realizar una prueba de hipótesis, se prueba siempre la verdad de la hipótesis nula H_0, aun cuando de antemano se desee rechazarla.

12. Errores de los tipos I y II. Nivel de significancia

En muchas ocasiones se presenta el caso de que se rechaza una hipótesis nula cuando en realidad debería ser aceptada; cuando esto sucede se dice que se ha cometido un *error de tipo I*. En otras ocasiones se acepta una hipótesis nula siendo en realidad falsa; en este caso se dice que se ha cometido un *error de tipo II*.

Al probar una hipótesis nula, a la máxima probabilidad con la que se está dispuesto a cometer un error del tipo I se le llama *nivel de significancia*, α, de la prueba, el cual dentro de la práctica se acostumbra establecer de 5 por ciento (0.05) o 10 por ciento (0.1). El complemento del nivel de significancia, $1 - \alpha$, se conoce como *nivel de confianza*.

Si, por ejemplo, al realizar una prueba de hipótesis se escoge un nivel de significancia de 10 por ciento, significa que existen 10 posibilidades en 100 de que se rechace ésta cuando debería ser aceptada; es decir, que se rechaza a un nivel de significancia del 10 por ciento, y que la probabilidad de que la decisión haya sido errónea es de 0.1.

13. Comportamiento de los errores tipos I y II

Supóngase que se trata de probar la hipótesis nula de que la media, μ_S, de la distribución muestral de la estadística S es μ_1, en contra de la hipótesis alternativa que establece que $\mu_S = \mu_2$, donde $\mu_2 > \mu_1$, es decir

$$H_0 : \mu_S = \mu_1$$

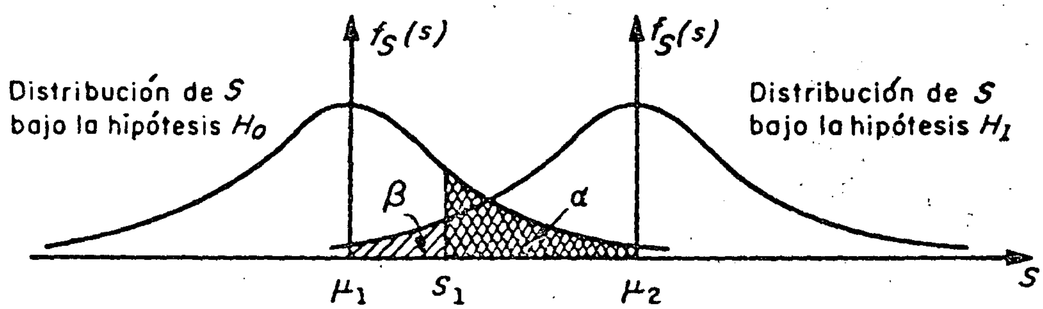
$$H_1 : \mu_S = \mu_2$$

En la fig 13.1 se muestra en forma gráfica la relación entre los errores tipos I y II en el caso en el que la regla de decisión para aceptar o rechazar H_0 es la siguiente:

Si el valor de la estadística S obtenido de una muestra excede de cierto valor crítico S_1, rechácese H_0; en caso contrario, acéptese.

Es evidente que si H_0 es verdadera, entonces α (área con rayado doble) es la probabilidad de que $S > S_1$, o sea la de rechazar a H_0 siendo verdadera (error tipo I). Por otro lado, si H_1 es verdadera, entonces β (área con rayado sencillo) es la probabilidad de que $S < S_1$, o sea la de aceptar H_0 siendo falsa (error tipo II).

Obsérvese que si se aumenta el valor de S_1 se reduce la probabilidad α, pero se incrementa la β; lo contrario sucede si se disminuye el valor de S_1.



$P[S>S_1] = \alpha$ (error tipo I)

$P[S<S_1] = \beta$ (error tipo II)

Fig 13.1 *Probabilidades de los errores tipos I y II en pruebas de hipótesis.*

En realidad, la única forma posible en la cual se pueden minimizar simultáneamente los errores de tipos I y II es aumentando el tamaño de la muestra, para hacer más "picudas" las distribuciones muestrales de la estadística bajo las hipótesis H_0 y H_1.

Al observar la fig 13.2 siguiente, es posible concluir que el tamaño de los errores I y II es menor para un tamaño de muestra igual a 100 que para un tamaño igual a 50, considerando la misma regla de decisión anterior.

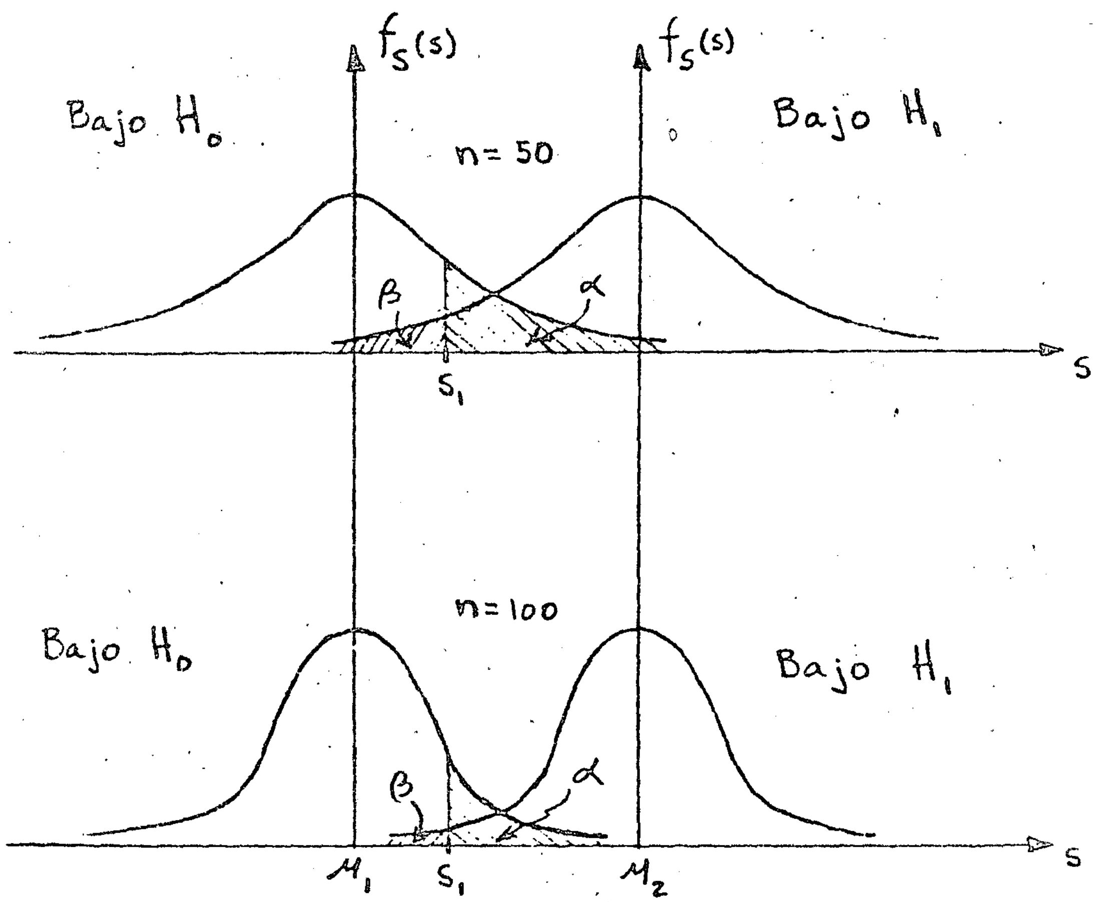


Fig 13.2

Sin embargo, esta técnica de reducción simultánea de ámbos tipos de errores no siempre puede ponerse en práctica, debido a razones de costo, tiempo, etc.

14. Regiones críticas, de rechazo ó de significancia. Regiones de aceptación.

Cuando una hipótesis nula no se acepta se dice que se *rechaza a un nivel de significancia del α por ciento*, o que el valor estandarizado de la estadística involucrada es *significativo a un nivel de significancia α*.

Al conjunto de los valores de la estadística en el que se rechaza la hipótesis nula se le denomina *región crítica, de rechazo, o de significancia*. Por el contrario, al conjunto de los valores de la estadística en que se acepta la hipótesis, se le llama *región de aceptación*.

Considérese que la distribución muestral de la estadística S es normal con desviación estándar σ_S, que la variable Z resulta de estandarizar a S, que la hipótesis nula, H_0, es que la media de S vale μ_S, y que la hipótesis alternativa H_1 es que dicha media es diferente de μ_S, es decir, que

$$Z = \frac{S - \mu_S}{\sigma_S}$$

H_0: media de la distribución muestral de $S = \mu_S$

H_1: media de la distribución muestral de $S \neq \mu_S$

Si se adopta la regla de decisión de aceptar la hipótesis H_0, si el valor de Z cae dentro del intervalo central que encierra al 99 por ciento del área de la distribución de probabilidades, entonces H_0 se aceptará en el caso en que

$$-2.58 \leqslant Z \leqslant 2.58$$

empleando la tabla de áreas bajo la curva normal estándar. Pero si el valor estandarizado de la estadística se encuentra fuera de dicho intervalo, se concluye que el evento puede ocurrir con probabilidad de 0.01 si la hipótesis H_0 es verdadera (área rayada total de la fig 14.1). En tal caso, el valor Z de la variable estándar difiere *significativamente* del que se podría esperar de acuerdo con la hipótesis nula, lo cual inclina a rechazarla a un nivel de confianza del 99 por ciento.

De lo anterior de deduce que el área total rayada de la fig 14.1 es el nivel de significancia α de la prueba, y representa la probabilidad de cometer un error del tipo I. Por ello, la región de aceptación de H_0 es $-2.58 \leqslant Z \leqslant 2.58$, y la de rechazo es $Z > 2.58$ y $Z < -2.58$.

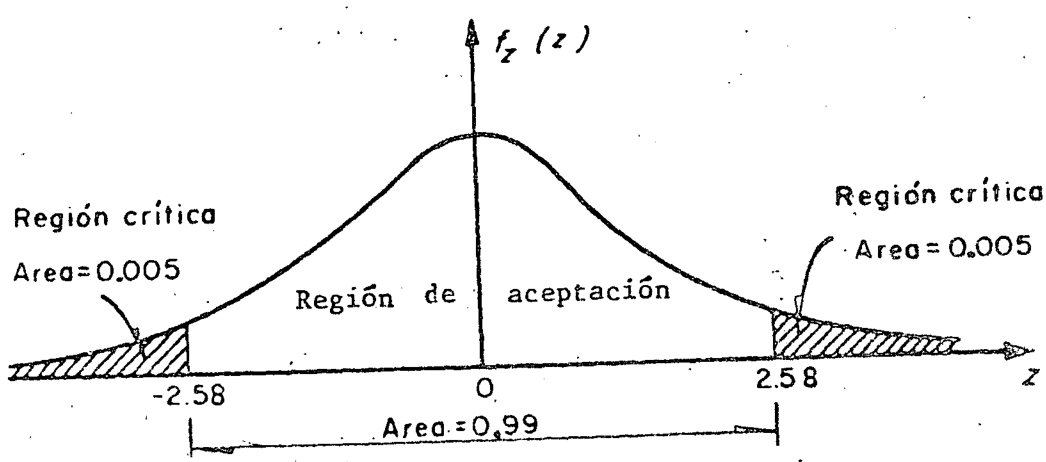


Fig 14.1 Región de significancia

En la tabla 14.1 se presentan los valores de la variable estandarizada, Z, que limitan las regiones de aceptación y de rechazo para el caso en el que la estadística involucrada en la prueba tenga distribución muestral normal. Cuando en alguna prueba de hipótesis se consideren niveles de significancia diferentes a los que aparecen en la tabla mencionada, resulta necesario emplear la de áreas bajo la curva normal estándar.

TABLA 14.1 VALORES CRITICOS DE z

Nivel de significancia, α	Valores de z para pruebas de una cola	Valores de z para pruebas de dos colas
0.1	−1.281 o 1.281	−1.645 y 1.645
0.05	−1.645 o 1.645	−1.960 y 1.960
0.01	−2.326 o 2.326	−2.575 y 2.575
0.005	−2.575 o 2.575	−2.810 y 2.810

15. Pruebas de una y de dos colas

En la prueba de hipótesis del ejemplo anterior, la región de rechazo de la hipótesis nula quedó en ambos extremos (colas) de la distribución muestral de la estadística involucrada en la prueba; a las pruebas de este tipo se les denomina *pruebas de dos colas*. Cuando la región de rechazo se encuentra solamente en un extremo de la distribución muestral en cuestión, se les llama *pruebas de una cola*.

Las pruebas de dos colas se presentan cuando en la hipótesis alternativa aparece el signo $\neq$ (diferente de), como en el siguiente caso

$$H_0 : \mu_S = \mu_1,$$

$$H_1 : \mu_S \neq \mu_1$$

en donde μ_S es la media de la estadística S, y μ_1 es un valor fijo.

En los casos

$$H_0 : \mu_S = \mu_1$$

$$H_1 : \mu_S < \mu_1$$

y

$$H_0 : \mu_S = \mu_1$$

$$H_1 : \mu_S > \mu_1$$

las pruebas resultan de una cola.

16. Pruebas de hipótesis para la media

Para el caso de una población infinita (o finita en que se muestree con remplazo), cuya desviación estándar σ se conoce o se puede estimar adecuadamente, si se tiene que la estadística S obtenida de la muestra es el promedio aritmético, entonces la media de su distribución muestral es $\mu_S = \mu_{\bar{X}} = \mu$, y su desviación estándar es $\sigma_S = \sigma_{\bar{X}} = \sigma/\sqrt{n}$, en donde μ y σ son, respectivamente, la media y la desviación estándar de la variable aleatoria X asociada a la población, y n es el tamaño de la muestra. En tal caso, si $\bar{X}$ tiene distribución normal, la variable estandarizada correspondiente será

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Para el caso de muestreo sin remplazo de población finita, se tiene que $\sigma_S' = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$, en donde N_p es el tamaño de la población, por lo que la variable estandarizada será

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}}$$

En los dos casos anteriores, el valor de Z correspondiente al de $\bar{X}$ de la muestra es el que se debe comparar con el valor crítico correspondiente al nivel de significancia fijado, para así aceptar o no la hipótesis nula (prueba de una cola). Si se trata de una prueba de dos colas, el valor de Z se debe comparar con los dos valores críticos que corresponden al valor de α seleccionado. En cualquiera de los casos anteriores, el valor o valores críticos se pueden obtener de la tabla 14.1, para valores comunes de α.

Ejemplo 16.1

Se sabe que el promedio de calificaciones de una muestra aleatoria de tamaño 100 de los estudiantes de tercer año de ingeniería civil es de 7.6, con una desviación estándar de 0.2. Si μ denota la media de la población de esas calificaciones, X, y si se supone que $\bar{X}$ tiene distribución normal, probar la hipótesis

$\mu = 7.65$ en contra de la hipótesis alternativa $\mu \neq 7.65$, usando un nivel de significancia de

a. 0.05

b. 0.01

Para la solución se deben considerar las hipótesis

$$H_0 : \mu = 7.65$$

$$H_1 : \mu \neq 7.65$$

Puesto que $\mu \neq 7.65$ incluye valores menores y mayores de 7.65, se trata de una prueba de dos colas.

La estadística bajo consideración es el promedio aritmético, $\bar{X}$, de la muestra, que se supone extraída de una población infinita. La distribución muestral de $\bar{X}$ tiene media $\mu_{\bar{X}} = \mu$, y desviación estándar $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, en donde μ y σ denotan, respectivamente, la media y la desviación estándar de la población de calificaciones.

Bajo la hipótesis H_0 (considerándola verdadera), se tiene que

$$\mu_{\bar{X}} = 7.65 = \mu$$

y utilizando la desviación estándar de la muestra como una estimación de σ, lo cual se supone razonable por tratarse de una muestra grande,

$$\sigma_{\bar{X}} = \sigma/\sqrt{n} = 0.2/\sqrt{100} = 0.2/10 = 0.02$$

a. Para la prueba de dos colas a un nivel de significancia de 0.05 se establece la siguiente regla de decisión

Aceptar H_0 si el valor Z correspondiente al valor del promedio de la muestra se encuentra dentro del intervalo de -1.96 a 1.96 (tabla 14.1). En caso contrario, rechazar H_0.

Puesto que

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{7.6 - 7.65}{0.02} = -2.5$$

se encuentra fuera del rango de -1.96 a 1.96, se rechaza la hipótesis H_0 a un nivel de significancia de 0.05.

b. Si el nivel de significancia es 0.01, el intervalo de -1.96 a 1.96 de la regla de decisión del inciso α se remplaza por el de -2.58 a 2.58 tabla (14.1). Entonces, puesto que el valor muestral $Z = -2.5$ se encuentra dentro de este intervalo, se acepta la hipótesis H_0 a un nivel de significancia de 0.01.

Ejemplo 16.2

La resistencia media a la ruptura de cables de acero fabricados por la empresa X es de 905 kg. Una empresa consultora sugiere a X que cambie su proceso de manufactura, con lo cual incrementará la resistencia de sus cables. Se prueba el nuevo proceso, y se extrae una muestra aleatoria de 50 cables, obteniéndose para ellos una resistencia promedio de 926 kg, con desviación estándar igual a 42 kg. ¿Se puede considerar que el nuevo proceso realmente incrementa la resistencia, con un nivel de confianza de 99%?

En este caso, se debe plantear una prueba de hipótesis de una cola, para la cual

$$H_0 : \mu = 905 \text{ kg}$$

$$H_1 : \mu > 905 \text{ kg}$$

Puesto que el tamaño de la muestra es suficientemente grande, se puede aproximar la distribución muestral de la resistencia promedio mediante una normal, y estimar el valor de σ de la población mediante S_X de la muestra.

Considerando a la población infinita, y suponiendo como verdadera a H_0, se tiene que

$$\mu_{\bar{X}} = \mu = 905 \text{ kg}$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{42}{\sqrt{50}} = 5.94$$

Para la prueba de una cola a un nivel de significancia de $\alpha = 1 - (1 - \alpha) = 1 - 0.99 = 0.01$, la regla de decisión es

Aceptar H_0 si el valor estandarizado de $\bar{X}$ de la muestra es menor o igual a $Z_c = 2.326$ (tabla 14.1); en caso contrario, rechazar H_0.

En virtud de que

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{926 - 905}{5.94} = 3.535$$

es mayor de 2.326, se rechaza H_0 a un nivel de significancia de 1%, concluyéndose que en realidad el nuevo proceso sí incrementa la resistencia de los cables.

17. Pruebas de diferencias de medias

Sean $\bar{X}$ y $\bar{Y}$ los promedios aritméticos obtenidos de dos muestras de tamaños n_X y n_Y, extraídas respectivamente de dos poblaciones con medias μ_X y μ_Y, y desviaciones estándar σ_X y σ_Y. Se trata de probar la hipótesis nula, H_0, de que no existe diferencia entre las medias, es decir, que $\mu_X = \mu_Y$. Si n_X y n_Y son suficientemente grandes (>30), la distribución muestral de las diferencias de los promedios es aproximadamente normal. Dicha distribución muestral es rigurosamente normal si las variables aleatorias X y Y asociadas a la población tienen distribución normal, aunque n_X y n_Y sean menores de 30. Para esta distribución muestral, la variable estandarizada Z, que se compara con los valores críticos correspondientes, se encuentra dada por

$$Z = \frac{X - Y - \mu_{\bar{X}-\bar{Y}}}{\sigma_{\bar{X}-\bar{Y}}} = \frac{X - Y - 0}{\sigma_{\bar{X}-\bar{Y}}} = \frac{\bar{X} - \bar{Y}}{\sigma_{\bar{X}-\bar{Y}}}$$

con la cual se puede probar la hipótesis nula H_0 en contra de otras hipótesis alternativas, H_1, a un nivel apropiado de significancia.

Ejemplo 17.1

En el laboratorio de pruebas de una empresa fabricante de aparatos electrónicos se ensayaron dos marcas de transistores, *A* y *B*, de características similares, con objeto de comprobar su ganancia de voltaje. Se tomaron muestras aleatorias de 100 transistores de cada marca, arrojando una ganancia promedio de 31 decibeles, con desviación estándar de 0.3 decibeles para la marca *A*, y 30.9 decibeles de ganancia promedio, con desviación estándar de 0.4 decibeles para la otra. ¿Existe una diferencia significativa entre las ganancias en voltaje de los transistores a un nivel de significancia de

a. 0.05

b. 0.01?

Si μ_A y μ_B son las medias respectivas de las dos poblaciones infinitas a las que corresponden las muestras, la prueba de hipótesis adopta la forma siguiente:

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A \neq \mu_B$$

Entonces, el valor de Z es, bajo la hipótesis H_0:

$$Z = \frac{\bar{x}_A - \bar{x}_B}{\sigma_{\bar{x}_A - \bar{x}_B}} = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} = \frac{31 - 30.9}{\sqrt{\frac{(0.3)^2}{100} + \frac{(0.4)^2}{100}}} = 2$$

a. Puesto que se trata de una prueba de dos colas a un nivel de significancia de 0.05, la diferencia es significativa si el valor de Z se encuentra fuera del intervalo de -1.96 a 1.96. Como este es el caso, puede concluirse que efectivamente existe diferencia significativa en la ganancia en voltaje de los transistores.

b. Si la prueba es a un nivel de significancia de 0.01, la diferencia es significativa si Z se encuentra fuera del rango de -2.58 a 2.58. Partiendo del hecho de que $Z = 2$, la diferencia entre las ganancias es producto del azar, y se acepta la hipótesis de que ambos tipos de transistores tienen igual ganancia media en voltaje a un nivel de confianza de 99 por ciento.

Ejemplo 17.2

La estatura promedio de 50 estudiantes varones tomados al azar que participan en actividades deportivas es de 173 cm, con desviación estándar de 6.3 cm. Otra muestra aleatoria de 50 estudiantes varones que no participan en ese tipo de actividades tiene promedio de estatura igual a 171 cm, con desviación estándar igual a 7.1 cm. Probar la hipótesis de que los estudiantes varones que practican deportes son más altos que los que no lo hacen, a un nivel de significancia de 0.05.

Se debe decidir entre las hipótesis

$$H_0 : \mu_X = \mu_Y$$

$$H_1 \quad \mu_X > \mu_Y$$

siendo X la variable aleatoria asociada a la población infinita de estaturas de alumnos que practican deportes, y Y la asociada a la de estudiantes que no lo hacen, que también es infinita.

Bajo la hipótesis H_0, ~~se tiene~~ que

$$\mu_{\bar{X}-\bar{Y}} = 0$$

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} = \sqrt{\frac{(6.3)^2}{50} + \frac{(7.1)^2}{50}} = 1.3424$$

Entonces, el valor de Z es

$$Z = \frac{\bar{X} - \bar{Y}}{\sigma_{\bar{X}-\bar{Y}}} = \frac{173 - 171}{1.3424} = \frac{2}{1.3424} = 1.489$$

Puesto que se trata de una prueba de hipótesis de una cola, a un nivel $\alpha = 0.05$, se rechazaría H_0 si el valor de Z muestral fuera mayor del valor crítico para dicho nivel, el cual es $Z_c = 1.645$. Puesto que $Z < Z_c$, en este caso se concluye que la diferencia en las estaturas de ambos grupos de estudiantes se debe únicamente al azar.

3.4 Muestras pequeñas

Como ya se indicó, para muestras grandes ($n \geqslant 30$) las distribuciones muestrales de muchas estadísticas son aproximadamente normales, siendo tanto mejor la aproximación cuanto mayor es el tamaño de n. Sin embargo, cuando se trata de muestras en las que $n < 30$, llamadas *muestras pequeñas*, la aproximación no es suficientemente buena, por lo que resulta necesario introducir una teoría apropiada para su estudio.

Al estudio de las distribuciones muestrales de las estadísticas para muestras pequeñas se le llama *teoría estadística de las muestras pequeñas*. Existen al respecto tres distribuciones importantes: *Ji cuadrada, F* y *t de Student.*

3.4.1 Distribución *Ji* cuadrada (χ^2)

Hasta ahora solo se ha tratado la distribución muestral de la media. En esta sección se verá lo concerniente a la distribución muestral de la variancia, S_X^2, para muestras aleatorias extraídas de poblaciones normales. Puesto que S_X no puede ser negativa, es de esperarse que su distribución muestral no sea una curva normal, ya que esta tiene ordenadas mayores de cero en el lado de las abscisas negativas. De hecho, la estadística S_X^2 se puede estudiar si se consideran muestras aleatorias de tamaño n extraídas de una población normal con desviación estándar σ_X y si para cada muestra se calcula el valor de la estadística.

$$\chi^2 = \frac{n\, S_X^2}{\sigma^2} \qquad (3.14)$$

donde S_X^2 es la variancia de la muestra.

El número de grados de libertad, ν, de una estadística se define como

$$\nu = n - k$$

siendo n el tamaño de la muestra y k el número de parámetros de la población que deben estimarse a partir de ella.

La distribución muestral de la estadística χ^2 está dada por la ecuación

$$f\,(\chi^2) = U\, \chi^{\nu - 2}\, e^{-\frac{1}{2} \chi^2}$$

en la que U es una constante que hace que el área total bajo la curva resulte igual a uno, y $\nu = n - 1$ es el número de grados de libertad. Esta distribución se llama *Ji cuadrada*, misma que se presenta en la fig 21 para distintos valores de ν.

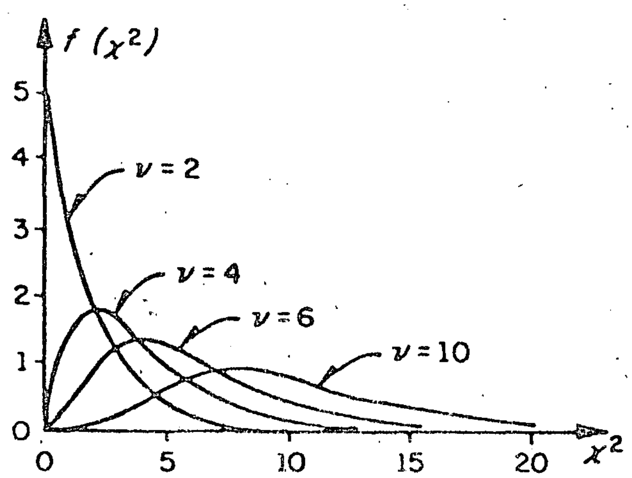
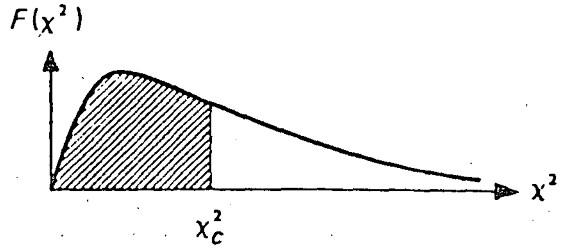


Fig 21. Distribución Ji cuadrada para distintos valores de ν

TABLA 8. VALORES CRITICOS χ^2_c



ν	$\chi^2_{.995}$	$\chi^2_{.99}$	$\chi^2_{.975}$	$\chi^2_{.95}$	$\chi^2_{.90}$	$\chi^2_{.75}$	$\chi^2_{.50}$	$\chi^2_{.25}$	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.01}$	$\chi^2_{.005}$
1	7.88	6.63	5.02	3.84	2.71	1.32	.455	.102	.016	.0039	.0010	.0002	.0000
2	10.6	9.21	7.38	5.99	4.61	2.77	1.39	.575	.211	.103	.0506	.0201	.0100
3	12.8	11.3	9.35	7.81	6.25	4.11	2.37	1.21	.584	.352	.216	.115	.072
4	14.9	13.3	11.1	9.49	7.76	5.39	3.36	1.92	1.06	.711	.483	.297	.207
5	16.7	15.2	12.8	11.15	9.2	6.63	4.35	2.67	1.61	1.15	.831	.554	.413
6	18.5	16.8	14.4	12.6	10.6	7.84	5.35	3.45	2.20	1.64	1.24	.872	.676
7	20.3	18.5	16.0	14.1	12.0	9.04	6.35	4.25	2.83	2.18	1.69	1.24	.989
8	22.0	20.1	17.5	15.5	13.4	10.2	7.34	5.07	3.49	2.73	2.18	1.65	1.34
9	23.6	21.7	19.0	16.9	14.7	11.4	8.34	5.90	4.17	3.33	2.70	2.09	1.73
10	25.2	23.2	20.5	18.3	16.0	12.5	9.34	6.74	4.87	3.94	3.25	2.56	2.16
11	26.8	24.7	21.9	19.7	17.3	13.7	10.35	7.57	5.58	4.57	3.82	3.05	2.60
12	28.3	26.2	23.2	21.0	18.5	14.8	11.3	8.44	6.30	5.23	4.40	3.57	3.07
13	29.8	27.7	24.7	22.4	19.8	16.0	12.3	9.30	7.04	5.89	5.01	4.11	3.57
14	31.3	29.1	26.1	23.7	21.1	17.2	13.3	10.2	7.79	6.57	5.63	4.66	4.07
15	32.7	30.6	27.5	25.1	22.3	18.2	14.3	11.0	8.55	7.26	6.25	5.22	4.60
16	34.3	32.0	28.8	26.3	23.5	19.4	15.3	11.9	9.31	7.96	6.91	5.81	5.14
17	35.7	33.4	30.2	27.6	24.8	20.5	16.3	12.8	10.1	8.67	7.56	6.41	5.70
18	37.2	34.8	31.5	28.9	26.0	21.6	17.3	13.7	10.9	9.39	8.23	7.01	6.26
19	38.6	36.2	32.9	30.1	27.2	22.7	18.3	14.6	11.73	10.1	8.91	7.63	6.84
20	40.0	37.6	34.2	31.4	28.45	23.8	19.3	15.5	12.4	10.9	9.59	8.26	7.43
21	41.4	38.8	35.6	32.7	29.6	24.9	20.3	16.3	13.2	11.6	10.3	8.90	8.02
22	42.8	40.3	36.8	33.9	30.8	26.0	21.3	17.2	14.0	12.3	11.0	9.54	8.64
23	44.2	41.6	38.1	35.2	32.0	27.1	22.3	18.1	14.8	13.1	11.7	10.2	9.26
24	45.6	43.0	39.4	36.4	33.2	28.2	23.3	19.0	15.7	13.8	12.4	10.9	9.89
25	46.9	44.3	40.6	37.7	34.4	29.3	24.3	19.9	16.5	14.5	13.15	11.5	10.5
26	48.3	45.6	41.9	38.9	35.6	30.4	25.3	20.8	17.3	15.4	13.8	12.2	11.2
27	49.6	47.0	43.2	40.1	36.7	31.5	26.3	21.7	18.1	16.2	14.6	12.9	11.8
28	51.0	48.3	44.5	41.3	37.9	32.6	27.3	22.7	18.9	16.9	15.3	13.6	12.5
29	52.3	49.6	45.7	42.5	39.1	33.7	28.3	23.6	19.8	17.7	16.0	14.3	13.1
30	53.7	50.9	47.0	43.8	40.3	34.8	29.3	24.5	20.6	18.5	16.8	15.0	13.8
40	66.8	63.7	59.3	55.8	51.8	45.7	39.3	33.7	29.1	26.5	24.4	22.2	20.7
50	79.5	76.2	71.4	67.5	63.2	56.3	49.3	43.0	37.7	34.8	32.4	29.7	28.0
60	92.0	88.4	83.3	79.1	74.4	67.0	59.3	52.3	46.5	43.2	40.5	37.5	35.5
70	104.2	100.4	95.0	90.5	85.5	77.6	69.3	61.7	55.3	51.7	48.8	45.4	43.3
80	116.3	112.3	106.6	101.9	96.6	88.1	79.3	71.1	64.3	60.4	57.2	53.5	51.2
90	128.3	124.1	118.1	113.1	107.6	98.6	89.3	80.6	73.3	69.1	65.6	61.8	59.2
100	140.2	135.8	129.6	124.3	118.5	109.1	99.3	90.12	82.4	77.9	74.2	70.1	67.3

No obstante que la distribución *Ji* cuadrada solo se ha presentado en el estudio de las muestras pequeñas, cabe aclarar que es válida para aquellas mayores de 30 si la variable aleatoria involucrada tiene distribución normal.

3.4.1.1 Intervalo de confianza para la variancia

Tal como se hizo para la distribución normal, se pueden establecer intervalos de confianza para la variancia de la población en términos de la variancia de una muestra extraída de ella, a un nivel de confianza dado $1 - \alpha$, si se hace uso de los valores críticos χ_c^2 de la tabla 8. Por lo tanto, un intervalo de confianza para la estadística χ^2, estaría dado por

$$\chi_{c'}^2 < \frac{n S_X^2}{\sigma^2} < \chi_c^2$$

donde $\chi_{c'}^2$ y χ_c^2 son los valores críticos para los cuales el $(1 - \alpha)/2$ por ciento del área se encuentra en los extremos izquierdo y derecho de la distribución, respectivamente.

Con base en lo anterior, se concluye que

$$\frac{n S_X^2}{\chi_c^2} < \sigma^2 < \frac{n S_X^2}{\chi_{c'}^2}$$

es un intervalo de confianza para estimar a σ^2 a un nivel de confianza $1 - \alpha$.

3.4.1.2 Prueba de hipótesis para la variancia

La prueba de hipótesis para la variancia de una población normal se efectúa calculando el valor de la estadística χ^2 y estableciendo las hipótesis H_0 y H_1 apropiadas, es decir, se adoptan reglas de decisión similares a las usadas para la estadística Z.

Ejemplo

La variancia del tiempo de elaboración de cierto producto es igual a 40 min; sin embargo, su proceso de manufactura se modifica y se toma una muestra de veinte tiempos, para la cual la variancia resulta ser igual a 62 min. ¿Es significativo el aumento del tiempo de elaboración a un nivel de significancia de

a) 0.05

b) 0.01?

Se debe decidir de entre las hipótesis

$$H_0 : \sigma^2 = 40 \text{ min}$$

$$H_1 : \sigma^2 > 40 \text{ min}$$

Suponiendo que la hipótesis nula es correcta, el valor de la estadística χ^2 para la muestra considerada es

$$\chi^2 = \frac{n S_X^2}{\sigma^2} = \frac{(20)(62)}{40} = 31$$

a) Como se trata de una prueba de una cola, la hipótesis H_0 se rechazaría si el valor de la estadística χ^2 fuera mayor que el de χ^2 para un nivel de significancia igual a 0.05, el cual, para $\nu = 20 - 1 = 19$ grados de libertad resulta ser 30.1 (tabla 8). Como $31 > 30.1$, H_0 se rechaza a un nivel de significancia de 0.05.

b) En este caso, el valor de χ^2 para un nivel de significancia de 0.01 y 19 grados de libertad es igual a 36.2. Puesto que $31 < 36.2$, se acepta H_0 a un nivel de significancia de 0.01.

3.4.2 Distribución F

Al efectuar la prueba de hipótesis de igualdad de medias para muestras pequeñas, en la siguiente sección se supondrá que las variancias de las poblaciones a las que corresponden tales muestras son iguales. Por lo tanto, es necesario probar antes si tal suposición es correcta. Para ello, debe considerarse que si S_X^2, n_X y S_Y^2, n_Y son respectivamente la variancia y el tamaño de dos muestras extraídas de poblaciones normales que tienen igual variancia, entonces

$$F = \frac{S_X^2}{S_Y^2} \tag{3.15}$$

TABLA 9. VALORES F_c PARA $\alpha = 0.01$

087

v_2 = Grados de libertad del denominador	v_1 = Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4.052	5.000	5.403	5.625	5.764	5.859	5.928	5.982	6.023	6.056	6.106	6.157	6.209	6.235	6.261	6.287	6.313	6.339	6.366
2	98.50	99.00	99.20	99.20	99.30	99.30	99.40	99.40	99.40	99.40	99.40	99.40	99.40	99.50	99.50	99.50	99.50	99.50	99.50
3	34.10	30.80	29.50	28.70	28.20	27.90	27.70	27.50	27.30	27.20	27.10	26.90	26.70	26.60	26.50	26.40	26.30	26.20	26.10
4	21.20	18.00	16.70	16.00	15.50	15.50	15.00	14.80	14.70	14.50	14.40	14.20	14.00	13.90	13.80	13.70	13.70	13.60	13.50
5	16.30	13.30	12.10	11.40	11.00	10.70	10.50	10.30	10.20	10.10	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.70	10.90	9.77	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.87
7	12.20	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.30	8.65	7.59	7.01	6.63	6.37	6.17	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.60	8.02	6.99	6.42	6.06	5.81	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.00	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.66	7.22	6.22	5.68	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.03	3.93	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.95	2.87
16	8.53	6.23	5.29	4.77	4.43	4.20	4.03	3.89	3.78	3.69	3.55	3.40	3.26	3.18	3.10	3.02	2.93	2.84	2.76
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	2.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	5.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.03	5.78	4.87	4.36	4.04	3.81	3.64	3.50	3.41	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.83	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.53	2.45	2.36	2.27	2.17
30	7.56	5.39	4.51	4.02	3.70	3.46	3.30	3.17	3.07	2.98	2.84	2.71	2.54	2.47	2.39	2.29	2.20	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.14	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.65	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

87

resulta ser el valor de una variable aleatoria (estadística) que tiene distribución F, con parámetros $\nu_X = n_X - 1$ y $\nu_Y = n_Y - 1$. Esta distribución (fig 22) cuenta con dos parámetros, ν_X y ν_Y, que son los grados de libertad que corresponden a la variancia del numerador y del denominador de la ec 3.15, respectivamente. Cuando se hace referencia a una distribución F en particular, siempre se dan primero los grados de libertad para la variancia del numerador; es decir, $F(\nu_X, \nu_Y)$. En la tabla 9 se presentan los valores críticos F_c para distintos valores de ν_X y ν_Y y un nivel de significancia de 0.01. Cuando los grados de libertad ν_X o ν_Y no se encuentren en dicha tabla, el valor de F se puede obtener mediante interpolación lineal. Si se desea probar la hipótesis a otros niveles de significancia, es factible emplear las tablas de la distribución F (refs 9 y 11).

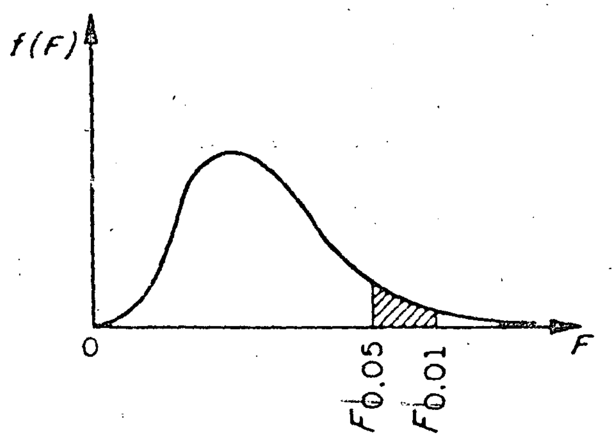


Fig 22. Distribución F

De acuerdo con lo anterior, se puede probar la hipótesis nula

$$H_0: \sigma_X^2 = \sigma_Y^2$$

en contra de alguna hipótesis alternativa adecuada haciendo uso del hecho de que el cociente S_x^2/S_Y^2 es una estadística que tiene distribución F.

Ejemplo

Una empresa manufacturera de cartón prensado va a decidir acerca del empleo de una prensadora A o una B a fin de obtener un grosor determinado en su producto. El problema estriba en que ambas prensadoras proporcionan grosores muy similares, es decir, que la variancia de los grosores para las dos máquinas es la misma. Para decidir acertadamente, se toma una muestra aleatoria de 31 cartones prensados por la máquina A y otra de 41 por la B. Como las variancias del grosor para los cartones de las muestras resultan ser de 12 y de 5 micras, respectivamente, se establecen las hipótesis

$$H_0 : \sigma_A^2 = \sigma_B^2$$

$$H_1 : \sigma_A^2 > \sigma_B^2$$

con objeto de probarlas a un nivel de significancia de 0.01.

El valor de la estadística F resulta

$$F = \frac{S_A^2}{S_B^2} = \frac{12}{5} = 2.4$$

Puesto que $\nu_A = 31 - 1 = 30$ y $\nu_B = 41 - 1 = 40$, en la tabla 9 se puede ver que para un nivel de significancia de 0.01 el valor, F_c, de $F(30, 40)$ es 2.11. De acuerdo con estos valores, la hipótesis H_0 se rechazaría si el valor de F fuera mayor que F_c (30, 40).

Puesto que lo anterior resulta ser cierto, se rechaza H_0, concluyéndose que la prensadora B sería la mejor elección.

3.4.3 Distribución t de Student

Si se consideran muestras de tamaño n extraídas de una población normal con media μ y variancia desconocida, para cada muestra se puede calcular la estadística T definida mediante la fórmula

$$T = \frac{\overline{X} - \mu}{S_X} \sqrt{n - 1} \tag{3.16}$$

donde $\overline{X}$ es el promedio y S_X la desviación estándar de la muestra.

La distribución muestral de T (fig 23) está dada por la ecuación

$$f(t) = \frac{U}{(1 + \frac{t^2}{\nu})^{\underbrace{(\nu + 1)/2}}}$$

ojo: va de exponente

en la ~~que U es una constante que~~ hace que el área bajo la curva sea igual a uno, y $\nu = n - 1$ es el número de grados de libertad.

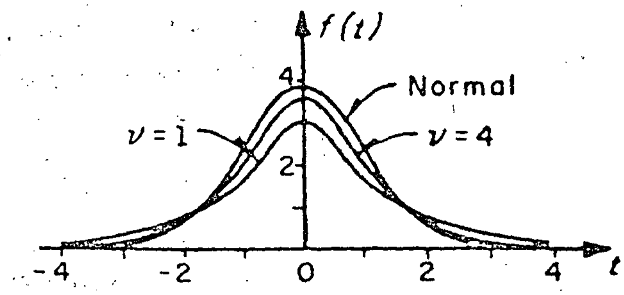


Fig 23. Distribución t de Student para distintos valores de ν

En la fig 23 se aprecia que conforme ν (o n, el tamaño de la muestra) aumenta, la distribución de $f(t)$ se aproxima a la distribución normal.

3.4.3.1 Límites e intervalos de confianza

De manera similar a como se hizo con la distribución normal, es posible estimar los límites de confianza de la media, μ, de una población mediante los *valores críticos*, t_c, de la distribución t, que dependen del tamaño de la muestra y del nivel de confianza deseado, encontrándose dichos valores en la tabla 10.

Así pues,

$$- t_c \leqslant \frac{\bar{X} - \mu}{S_X} \sqrt{n-1} \leqslant t_c \; .$$

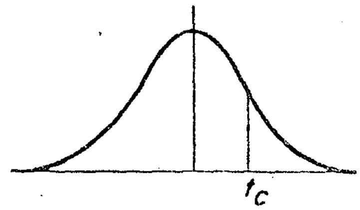
representa un intervalo de confianza para t, a partir del cual se puede estimar que μ se encuentra dentro del intervalo

$$\bar{X} - t_c \frac{\sigma_X}{\sqrt{n-1}} < \mu < \bar{X} + t_c \frac{\sigma_X}{\sqrt{n-1}}$$

En términos generales, los límites de confianza para la media de la población se representan como

$$\bar{X} \pm t_c \frac{\sigma_X}{\sqrt{n-1}}$$

TABLA 10. VALORES t_c PARA LA DISTRIBUCION t DE STUDENT



ν	$t_{.995}$	$t_{.99}$	$t_{.975}$	$t_{.95}$	$t_{.90}$	$t_{.80}$	$t_{.75}$	$t_{.70}$	$t_{.60}$	$t_{.55}$
1	63.66	31.82	12.71	6.31	3.07	1.376	1.000	.727	.325	.158
2	9.92	6.96	4.30	2.92	1.89	1.061	.816	.617	.289	.142
3	5.84	4.54	3.18	2.35	1.64	.978	.765	.584	.275	.138
4	4.60	3.75	2.78	2.13	1.53	.941	.741	.569	.271	.134
5	4.04	3.36	2.58	2.02	1.48	.920	.727	.560	.267	.132
6	3.71	3.14	2.45	1.94	1.44	.906	.718	.553	.265	.131
7	3.50	3.00	2.36	1.91	1.43	.896	.711	.549	.263	.130
8	3.36	2.90	2.31	1.86	1.40	.889	.706	.546	.262	.130
9	3.25	2.82	2.26	1.83	1.38	.883	.703	.543	.261	.129
10	3.17	2.76	2.23	1.81	1.37	.879	.700	.542	.260	.129
11	3.11	2.72	2.20	1.80	1.36	.876	.697	.540	.260	.129
12	3.06	2.68	2.18	1.78	1.36	.873	.695	.539	.259	.128
13	3.01	2.65	2.16	1.77	1.36	.871	.694	.538	.259	.128
14	2.98	2.62	2.14	1.76	1.34	.868	.693	.537	.258	.128
15	2.95	2.61	2.13	1.75	1.34	.866	.691	.536	.258	.128
16	2.92	2.58	2.12	1.75	1.34	.865	.690	.535	.258	.128
17	2.90	2.57	2.11	1.74	1.33	.863	.689	.534	.257	.128
18	2.88	2.55	2.10	1.73	1.33	.862	.688	.534	.257	.128
19	2.87	2.54	2.09	1.73	1.33	.861	.688	.533	.257	.127
20	2.84	2.53	2.09	1.72	1.32	.860	.687	.533	.257	.127
21	2.83	2.52	2.08	1.72	1.32	.859	.686	.532	.256	.127
22	2.82	2.51	2.07	1.72	1.32	.858	.686	.532	.256	.127
23	2.81	2.50	2.07	1.71	1.32	.858	.685	.532	.256	.127
24	2.80	2.49	2.06	1.71	1.32	.857	.685	.531	.256	.127
25	2.79	2.48	2.06	1.71	1.32	.856	.684	.531	.256	.127
26	2.78	2.48	2.05	1.71	1.32	.856	.684	.531	.256	.127
27	2.77	2.47	2.05	1.71	1.31	.855	.683	.531	.256	.127
28	2.76	2.47	2.05	1.70	1.31	.855	.683	.530	.256	.127
29	2.76	2.46	2.04	1.70	1.31	.854	.683	.530	.256	.127
30	2.75	2.46	2.04	1.70	1.30	.853	.683	.530	.256	.127
40	2.70	2.43	2.02	1.68	1.30	.851	.681	.529	.255	.126
60	2.66	2.39	2.00	1.67	1.30	.848	.679	.528	.254	.126
120	2.62	2.36	1.98	1.66	1.29	.845	.677	.526	.254	.126
∞	2.58	2.33	1.96	1.645	1.28	.842	.674	.524	.253	.126

3.4.3.2 Pruebas de hipótesis

La prueba de hipótesis para la media de una población se puede efectuar con muestras pequeñas en forma análoga a la de muestras de tamaño mayor de 30 si en lugar de utilizar a la estadística Z se emplea la T. Entonces, si se consideran dos muestras aleatorias cuyos tamaños, desviaciones estándar y promedios son n_X, S_X, $\overline{X}$ y n_Y, S_Y, $\overline{Y}$, respectivamente, extraídas de poblaciones normales de igual variancia ($\sigma_X^2 = \sigma_Y^2$), se puede probar la hipótesis, H_0, de que las muestras provienen de una misma población, es decir, de que también sus medias son iguales, utilizando la estadística T definida por

$$T = \frac{\overline{X} - \overline{Y}}{\epsilon \sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}} \qquad (3.17)$$

donde

$$\epsilon = \sqrt{\frac{n_X S_X^2 + n_Y S_Y^2}{n_X + n_Y - 2}} \qquad (3.18)$$

cuya distribución es la t de Student, con $\nu = n_X + n_Y - 2$ grados de libertad.

Ejemplo

Conforme al plan de desarrollo agrícola de una región, se probó un nuevo fertilizante para maíz. Para ello se escogieron 24 ha de terreno, aplicándose dicho producto a la mitad de ellas. El promedio de producción de maíz en la zona que se usó fertilizante fue de 5.3 ton, con una desviación estándar de 0.40 ton, en tanto que en la otra zona el promedio fue de 5.0 ton, con desviación estándar de 0.36 ton.

De acuerdo con los resultados, ¿se puede concluir que existe un aumento significativo en la producción de maíz al usar fertilizante, si se utiliza un nivel de significancia de

a) 0.01

b) 0.05?

Solución

Para probar la hipótesis de igualdad de medias es indispensable saber primero si las muestras provienen de dos poblaciones normales de igual variancia. En ese caso, si σ_X^2 y σ_Y^2 denotan a las variancias de la producción de maíz en la zona tratada y en la no tratada, respectivamente, se debe probar la hipótesis nula H_0: $\sigma_X^2 = \sigma_Y^2$ en contra de la hipótesis alternativa H_1: $\sigma_X^2 > \sigma_Y^2$ a los dos niveles de significancia establecidos.

El valor de la estadística F es, de la ec 3.15,

$$F = \frac{S_X^2}{S_Y^2} = \frac{(0.40)^2}{(0.36)^2} = 1.27$$

y el valor crítico de F (11, 11), obtenido de la tabla 9 mediante interpolación lineal, resulta 4.47. Por lo tanto, como $1.27 < 4.47$, se acepta la hipótesis nula a un nivel de significancia de 0.01.

El valor crítico de F (11, 11) a un nivel de significancia de 0.05 (ref 9) es 2.82, de ahí que como $1.27 < 2.82$, también se acepta la hipótesis H_0.

Con base en lo anterior, se debe decidir entre las hipótesis

H_0: $\mu_X = \mu_Y$ (la diferencia en los promedios se debe al azar)

H_1: $\mu_X > \mu_Y$ (el fertilizante mejora la producción)

Bajo la hipótesis H_0, se tiene que

$$\epsilon = \sqrt{\frac{n_X S_X^2 + n_Y S_Y^2}{n_X + n_Y - 2}} = \sqrt{\frac{12(0.40)^2 + 12(0.36)^2}{12 + 12 - 2}} = 0.397$$

por lo cual

$$t = \frac{5.3 - 5.0}{0.397 \sqrt{\frac{1}{12} + \frac{1}{12}}} = 1.85$$

a) Puesto que se trata de una prueba de una cola a un nivel de significancia de 0.01, se rechaza la hipótesis H_0 si t es mayor que el valor crítico, t_c, correspondiente a dicho nivel, el cual para $\nu = n_X + n_Y - 2 = 12 + 12 - 2 = 22$ grados de libertad, se obtiene de la tabla 8 como $t_c = 2.51$. Como $t < t_c$, la hipótesis H_0 no se puede rechazar a un nivel de significancia de 0.01.

b) Si el nivel de significancia de la prueba es de 0.05, se rechaza H_0 si t es mayor que el valor t_c respectivo que para 22 grados de libertad es $t_c = 1.72$, por lo que de acuerdo con lo anterior, H_0 se rechaza a un nivel de significancia de 0.05.

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

FUNDAMENTOS DE LAS TECNICAS DE MUESTREO ESTADISTICO

TAMAÑO DE LA MUESTRA

M. en I. Augusto Villarreal A.

27 febrero, 1979

Palacio de Minería Calle de Tacuba 5, primer piso. México 1, D. F.

4. TAMAÑO DE LA MUESTRA

Por: M en I Augusto Villarreal Aranda

INTRODUCCION

Dentro de un plan de muestreo, cuando ya se ha establecido la característica (o características) a estimar, así como el nivel de confianza y el grado de precisión requeridos, se debe decidir cuál debe ser el tamaño de la muestra o número de elementos a seleccionar por el procedimiento de muestreo que vaya a emplearse, en forma tal que los resultados que se obtengan no sean en exceso costosos o imprecisos.

Una vez que se ha fijado el error máximo admisible, que representa la precisión mínima que se exige tengan los resultados, así como el nivel de confianza $P_K = 1 - \alpha$, se requiere conocer además, en la forma más precisa posible, la variabilidad de la población, ya que cuanto más dispersos estén los valores de la variable asociada a ella más arriesgado será el utilizar una muestra de tamaño pequeño.

A continuación se expondrá el procedimiento para seleccionar el tamaño de muestra más adecuado en el caso del muestreo aleatorio simple o irrestrictamente aleatorio (sin remplazo). Más adelante se estudiarán los métodos para calcular el tamaño de la muestra para otros procedimientos de muestreo.

4.1 Tamaño de una muestra aleatoria simple (Medias)

En este caso se trata de estimar la media μ de una población con variable aleatoria asociada X mediante el empleo del promedio aritmético $\overline{X}$, obtenido de una muestra aleatoria de tamaño n con un error máximo admisible absoluto e y un nivel de confianza P_K. Es natural que a la probabilidad P_K le corresponderá un cierto valor de desviación K, obtenido a partir de la desigualdad de Chebyshev, o bien considerando a K como el número de desviaciones estándar para una distribución normal o para una t de Student.

El procedimiento para obtener el tamaño de la muestra se fundamenta en el hecho de que

$$P\ \{\ \overline{X} - K\sigma_{\overline{X}} \leqslant \mu \leqslant \overline{X} + K\sigma_{\overline{X}}\ \} = P_K = 1 - \alpha$$

o sea que con probabilidad o nivel de confianza P_K se puede asegurar que el valor de μ de una población se encuentra dentro del

$(1-\alpha)$ % de los intervalos formados a partir de muestras de tamaño n, de la forma siguien e

$$(\bar{X} - K\sigma_{\bar{X}} \; , \; \bar{X} + K\sigma_{\bar{X}})$$

Lo anterior implica que los límites de confianza del P_K % para estimar a μ son

$$\bar{X} \pm K\sigma_{\bar{X}}$$

es decir, que el error en la estimación del valor de μ es, en valor absoluto,

$$|\text{error en la estimación de } \mu| = K\sigma_{\bar{X}} \qquad (4.1)$$

Por lo tanto, es posible escribir

$$|\text{error máximo admisible}| = |\text{error en la estimación de } \mu| = e$$

4.1.1 Muestreo de una población finita

De la inferencia estadística, el valor de $\sigma_{\bar{X}}$, la desviación estándar de la distribución muestral de $\bar{X}$ (o error estándar de $\bar{X}$) cuando la población es finita es

$$\sigma_{\bar{X}} = \sqrt{\frac{N_p - n}{N_p - 1} \; \frac{\sigma_X^2}{n}}$$

pudiéndose escribir entonces

$$e = K\sigma_{\bar{X}} = K\sqrt{\frac{N_p - n}{N_p - 1} \; \frac{\sigma_X^2}{n}}$$

siendo K la desviación correspondiente al nivel de confianza P_k N_p el tamaño de la población, σ_x^2 la variancia de esta última y n el tamaño de la muestra.

Puesto que se desea conocer el tamaño de la muestra, éste se puede obtener despejando de la ecuación anterior el valor de n. Para ello, se requiere elevar al cuadrado ambos miembros, es decir

$$e^2 = K^2 \frac{N_p - n}{N_p - 1} \frac{\sigma_x^2}{n}$$

$$e^2 = \frac{K^2 \sigma_x^2 N_p - K^2 \sigma_x^2 n}{(N_p - 1) n}$$

despejando a n:

$$ne^2 (N_p - 1) = K^2\sigma_x^2N_p - K^2\sigma_x^2n$$

$$ne^2N_p - ne^2 = K^2\sigma_x^2N_p - K^2\sigma_x^2n$$

$$ne^2N_p - ne^2 + K^2\sigma_x^2n = k^2\sigma_x^2N_p$$

$$n(e^2N_p - e^2 + K^2\sigma_x^2) = k^2\sigma_x^2N_p$$

$$\therefore n = \frac{K^2\sigma_x^2N_p}{e^2N_p - e^2 + K^2\sigma_x^2} \qquad (4.2)$$

La fórmula anterior permite obtener el tamaño de la muestra considerando conocidos K, e, N_p y σ_x^2. Puesto que el valor de σ_x^2 de la población usualmente se desconoce, se debe estimar previamente en forma adecuada considerando la información disponible de poblaciones semejantes a la que deberá muestrearse, o tomando una muestra preliminar suficientemente grande de dicha población.

Puesto que el tamaño de la muestra debe corresponder a un número entero positivo, se deberá asignar a n el valor entero más próximo por exceso al obtenido mediante la fórmula 4.2.

4.1.2 Muestreo de una población infinita

Cuando el muestreo se realiza a partir de una población infinita, el valor de $\sigma_{\bar{x}}$, la desviación estándar de la distribución muestral de $\bar{X}$, es

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

en donde σ_x es la desviación estándar de la población y n el tamaño de la muestra.

considerando la ecuación 4.1, se puede escribir en este caso

$$|\text{error en la estimación de } \mu| = e = K\sigma_{\bar{x}} = K\frac{\sigma_x}{\sqrt{n}}$$

Para obtener el valor de n, se elevan al cuadrado ambos miembros de la expresión anterior, es decir,

$$e^2 = \frac{K^2\ \sigma_x^2}{n}$$

Por lo cual

$$n = \frac{K^2\ \sigma_x^2}{e^2}$$

Para resaltar el hecho de que en este caso el tamaño de la muestra se obtiene a partir de una población infinita, en lugar de emplear n se puede emplear n_∞ , es decir

$$n_\infty = \frac{K^2\ \sigma_x^2}{e^2} \qquad (4.3)$$

Al igual que en el caso de una población finita, el tamaño de la muestra dado por la ec 4.3 debe corresponder a un número natural, por lo cual se debe aproximar por exceso al valor entero más cercano.

4.1.3 Comparación entre n y n_∞

Si se divide entre $N_p e^2$ el numerador y el denominador del miembro izquierdo de la ecuación 4.2, se obtiene

$$n = \frac{\dfrac{K^2\sigma_x^2 N_p}{N_p e^2}}{\dfrac{e^2 N_p - e^2 + K^2\sigma_x^2}{N_p e^2}} = \frac{\dfrac{K^2\sigma_x^2}{e^2}}{1 - \dfrac{1}{N_p} + \dfrac{K^2\sigma_x^2}{N_p e^2}}$$

$$n = \frac{\dfrac{K^2\sigma_x^2}{e^2}}{1 + \dfrac{1}{N_p}\left(\dfrac{K^2\sigma_x^2}{e^2} - 1\right)}$$

y, considerando el valor de n_∞ dado por la ec 4.3, se obtiene finalmente

$$n = \frac{n_\infty}{1 + \frac{1}{N_p}\ (n_\infty - 1)} \qquad (4.4)$$

Como se puede apreciar de la ec 4.4, el valor de n es menor que el de n_∞, a menos que $N_p = \infty$.

4.1.4 Empleo adecuado de n y n_∞

Para una población finita, se definirá la <u>fracción de muestreo</u> como

$$\text{fracción de muestreo} = fm = \frac{n_\infty}{N_p}$$

siendo n_∞ el tamaño de la muestra calculada con la ec 4.3, y N_p el tamaño de la población.

Al obtener el tamaño de la muestra cuando se trata de una población finita, usualmente se acostumbra emplear la fórmula 4.3, que proporciona dicho tamaño para población infinita, y considerar como bueno dicho valor siempre que se cumpla la condición

$$fm \leqslant 0.05$$

Lo anterior quiere decir que en la práctica se calcula el valor de n_∞, y si n_∞/N_p cumple con la condición mencionada, entonces se considera que n_∞ es una aproximación satisfactoria de n. Si la

condición no se cumple, entonces se emplea la ec 4.4 para obtener el valor de n.

Es claro que tomando como tamaño de la muestra a n_∞ siempre se estará del lado más prudente, en el sentido de que se toma una muestra igual o mayor que la necesaria. Sin embargo, la eficiencia del diseño exige que el gasto y el tiempo de muestreo no sean superiores a los que haya que efectuar.

Ejemplo 4.1

Sea una población normal finita con variancia aproximadamente igual a 500. Se desea obtener una muestra aleatoria para estimar mediante $\bar{X}$ a la media poblacional μ_X, con error en la estimación no mayor de 10 y nivel de confianza igual a 90%. Obténgase el valor de n considerando que el tamaño de la población es igual a

a. 1000
b. 100

Solución

a. Puesto que $\sigma_X^2 = 500$, $e = 10$ y $1 - \alpha = 0.90$, tratándose de una población normal se tiene que

$$K = Z_{0.45} = 1.645$$

por lo cual

$$n_\infty = \frac{K^2 \sigma_X^2}{e^2} = \frac{(1.645)^2 \ (500)}{10^2}$$

$$= (2.706)\ (5) = 13.53$$

$$\therefore n_\infty = 14$$

En virtud de que en este caso

$$fm = \frac{n_\infty}{N_p} = \frac{14}{1000} = 0.014 < 0.05$$

se considera que n = 14.

b. En este caso

$$fm = \frac{14}{100} = 0.14 > 0.05$$

por lo cual se emplea la ec 4.4 para obtener el valor de n, es decir,

$$n = \frac{n_\infty}{1 + \frac{1}{N_p}\ (n_\infty - 1)} = \frac{14}{1 + \frac{1}{100}\ (14 - 1)}$$

$$= \frac{14}{1 + \frac{13}{100}} = \frac{14}{1.13} = 12.389$$

$$\therefore n = 13$$

Ejemplo 4.2

Cierta universidad cuenta con 4726 estudiantes, y se desea conocer el rendimiento académico medio de todos ellos, en términos de una escala de calificación que va de cero a cien puntos. En estudios semejantes en otras universidades, se obtuvo que la desviación estándar de las calificaciones es aproximadamente igual a 7 puntos. Si el error en la estimación de la media de calificaciones no debe ser mayor de un punto en valor absoluto, y el nivel de confianza es igual a 99%, ¿cuál debe ser el tamaño de la muestra para realizar la estimación?

Solución

En este caso, aproximando la distribución muestral de $\bar{X}$ mediante la distribución normal, se debe considerar que

$$P_K = 1 - \alpha = 0.99 \quad \therefore \quad K = Z_{0.495} = 2.58$$

$$\sigma_X^2 = (7)^2 = 49 \quad ; \quad e = 1 \text{ punto}$$

Por lo tanto,

$$n_\infty = \frac{Z_C^2 \ \sigma_X^2}{e^2} = \frac{(2.58)^2 \ (49)}{(1)^2}$$

$$= \frac{(6.656) \ (49)}{1} = 326.144$$

O sea $n_\infty = 327$

Puesto que

$$fm = \frac{n_\infty}{N_p} = \frac{327}{4726} = 0.0692 > 0.05$$

se procede a calcular n, es decir,

$$n = \frac{n_\infty}{1 + \frac{1}{N_p}(n_\infty - 1)} = \frac{327}{1 + \frac{1}{4726}(327 - 1)}$$

$$= \frac{327}{1 + \frac{326}{4726}} = \frac{327}{1.069} = 305.89$$

$$\therefore n = 306$$

Ejemplo 4.3

Una muestra aleatoria de 14 observaciones de la altura alcanzada por cierto tipo de planta arrojó los siguientes datos:

Nº de elemento	Altura, X, en pulgadas
1	52.3
2	48.1
3	55.7
4	56.8
5	50.1
6	49.2
7	47.7
8	50.8
9	57.9
10	52.5
11	54.7
12	49.6
13	53.9
14	56.0

Obténgase el tamaño de muestra necesario para asegurar, con una probabilidad igual a 0.95, que el error en la estimación de la media de alturas de esta variedad de planta no sea mayor del 2.86%.

Solución

Se deben obtener primero los valores de $\bar{X}$ y S_X^2 de la muestra, con los cuales se estimarán los de μ_X y σ_X^2 de la población. Para ello, se dispone la información en la forma siguiente :

X_i	X_i^2
52.3	2735.3
48.1	2313.6
55.7	3102.5
56.8	3226.2
50.1	2510.0
49.2	2420.6
47.7	2275.3
50.8	2580.6
57.9	3352.4
52.5	2756.2
54.7	2992.1
49.6	2460.2
53.9	2905.2
56.0	3136.0
Σ 735.3	38766.2

Por lo tanto,

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{n} X_i = \frac{1}{14}(735.3) = 52.52 \text{ pulgadas}$$

$$S_X^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{14}(38766.2) - (52.52)^2$$

$$= 2769.01 - 2758.35 = 10{,}66 \text{ pulgadas}$$

Puesto que el error en la estimación de la media no debe ser mayor del 2.86%, y el estimador de μ_X es $\overline{X} = 52.52$, se tiene que

$$e = 52{,}52\ (0{,}0286) = 1{,}5 \text{ pulgadas}$$

Por otra parte, se desconoce el valor real de σ_X^2 de la población, además de que S_X^2, su estimador, se ha obtenido de una muestra menor de 30 elementos. Por lo tanto, la distribución teórica a la cual se debe aproximar la muestral debe ser la t de Student, siendo en este caso $K = t_c$. Sin embargo, puesto que en este caso se estima σ_X^2 mediante S_X^2 de la muestra, se debe tener presente que el error en la estimación de μ_X es

$$e = K\ \sigma_{\bar{X}} = t_c\ \sigma_{\bar{X}} = t_c\ \frac{S_X}{\sqrt{n-1}}$$

O sea, elevando al cuadrado

$$e^2 = t_c^2\ \frac{S_X^2}{n-1}$$

y, despejando a n,

$$n - 1 = \frac{t_c^2\ S_X^2}{e^2}$$

$$n = \frac{t_c^2\ S_X^2}{e^2} + 1$$

Por ser muestreo de población infinita, se puede escribir finalmente

$$n_\infty = \frac{t_c^2\, S_x^2}{e^2} + 1 \qquad (4.5)$$

Ya que el valor de t_c depende del número de grados de libertad de la muestra υ, y este último depende del tamaño de la muestra (ya que $\upsilon = n - 1$), la fórmula anterior para obtener el valor de n_∞ contiene dos incógnitas. Por ello, se sigue el siguiente proceso iterativo para obtener el valor de n_∞ :

1. Se hace $t_{0.025} = Z_{0.475}$, es decir

$$t_{0.025} = 1.96$$

Con dicho valor de t_c se obtiene

$$n_\infty = \frac{(1.96)^2\,(10.66)}{(1.5)^2} + 1 = 18.2 + 1 = 19.3 \Rightarrow 20$$

De la tabla de la distribución t, se obtiene $t_{0.025} = 2.09$, para $\upsilon = 20-1 = 19$ grados de libertad.

2. Se toma ahora $t_{0.025} = 2.09$, y se obtiene

$$n_\infty = \frac{(2.09)^2\,(10.66)}{(1.5)^2} + 1 = 20.7 + 1 = 21.7 \Rightarrow 22$$

De la tabla de la distribución t, se obtiene $t_{0,025} = 2.08$, para $\upsilon = 22-1 = 21$ grados de libertad.

3. Se toma ahora $t_{0.025} = 2.08$, y se obtiene

$$n_\infty = \frac{(2.08)^2 \; (10.66)}{(1.5)^2} + 1 = 20.5 + 1 = 21.5 \Rightarrow 22$$

En este paso se obtiene un valor de n_∞ igual al del paso anterior, por lo que se puede considerar que el tamaño de muestra adecuado es igual a 22 plantas.

En este caso la población es infinita, por lo cual no se requiere hacer la corrección para población finita con la ec 4.4. Sin embargo, debe aclararse que es posible emplear la ec 4.5 para obtener n_∞ primero y, si la población de la que se muestrea es finita, usar después la ec 4.4 para obtener el valor de n corregido.

4.2 Tamaño de una muestra aleatoria simple (Totales)

Una característica o parámetro poblacional de gran interés es el total, que corresponde a la suma de todos los valores y_i que constituyen la población, es decir,

$$Y = \sum_{i=1}^{N_p} y_i$$

en donde Y denota al total, y N_p es el número de elementos de la misma.

Si se multiplica y divide por N_p el 2º miembro de la ecuación ante

rior, se obtiene

$$Y = \frac{N_p}{N_p} \sum_{i=1}^{N_p} Y_i = N_p \, \mu_Y$$

Es decir, el total de una población es igual al tamaño de la misma multiplicado por la media correspondiente.

Cono estimador puntual del total de la población se puede tomar el de la estadística

$$\hat{Y} = N_p \, \bar{Y}$$

en donde $\bar{Y}$ es el promedio aritmético de la muestra, y $\hat{Y}$ un estimador insesgado en virtud de que

$$E\{\hat{Y}\} = E\{N_p \bar{Y}\} = N_p E\{\bar{Y}\} = N_p \, \mu_Y = Y$$

Por otra parte, la variancia de la distribución muestral de $\hat{Y}$ es

$$\sigma^2_{\hat{Y}} = \sigma^2_{N_p \bar{Y}} = \mathrm{Var}\{N_p \bar{Y}\} = N_p^2 \,\mathrm{Var}\{\bar{Y}\} = N_p^2 \, \sigma^2_{\bar{Y}}$$

y la desviación estándar es

$$\sigma_{\hat{Y}} = \sigma_{N_p \bar{Y}} = N_p \sigma_{\bar{Y}} = N_p \frac{\sigma_Y}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$$

De igual manera a como se hizo para las medias, el valor del tamaño de muestra para estimar el total con un nivel de confianza y un error absoluto dados, se obtiene en la forma siguiente

$$e = K\,\sigma_{\hat{Y}} = K\,N_p\,\frac{\sigma_Y}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}}$$

Elevando al cuadrado y realizando operaciones algebraicas,

$$e^2 = K^2\,N_p^2\,\frac{\sigma_Y^2}{n}\,\frac{N_p - n}{N_p - 1}$$

$$e^2 = \frac{K^2\,N_p^3\,\sigma_Y^2 - K^2\,N_p^2\,\sigma_Y^2\,n}{n(N_p - 1)}$$

$$n\left(1 + \frac{K^2\,N_p^2\,\sigma_Y^2}{e^2(N_p - 1)}\right) = \frac{K^2\,N_p^3\,\sigma_Y^2}{e^2(N_p - 1)}$$

O sea

$$n = \frac{K^2\,N_p^3\,\sigma_Y^2}{e^2(N_p - 1) + K^2\,N_p^2\,\sigma_Y^2}$$

Dividiendo el numerador y denominador de la expresión anterior entre $N_p e^2$, se obtiene

$$n = \frac{\dfrac{K^2 N_p^3\,\sigma_Y^2}{N_p e^2}}{\dfrac{e^2 N_p - e^2 + K^2 N_p{}^2\,\sigma_Y^2}{N_p e^2}}$$

$$= \frac{N_p^2\,\dfrac{K^2\sigma_Y^2}{e^2}}{1 - \dfrac{1}{N_p} + \dfrac{N_p^2}{N_p}\,\dfrac{K^2\,\sigma_Y^2}{e^2}}$$

Considerando la ec 4.3, queda finalmente

$$n = \frac{N_p^2\,n_\infty}{1 + \dfrac{1}{N_p}\,(N_p^2\,n_\infty - 1)} \qquad (4.6)$$

Ejemplo 4.4

Con el fin de hacer una solicitud al Gobierno, se recogieron firmas de habitantes de una ciudad en 676 hojas. Cada hoja tenía espacio suficiente para 42 firmas, pero en varias hojas se recolectó un número menor de ellas. Para obtener una estimación del total de firmas, se contó el número de firmas por hoja en una muestra aleatoria de 50 hojas, obteniéndose los datos que aparecen en la tabla siguiente:

Número de firmas, y_i	Número de hojas, f_i
42	23
41	4
36	1
32	1
29	1
27	2
23	1
19	1
16	2
15	2
14	1
11	1
10	1
9	1
7	1
6	3
5	2
4	1
3	1

Obtener el tamaño de muestra necesario para estimar el valor del total de firmas con un error absoluto igual al 5%, considerando un nivel de confianza igual a 95%.

Solución : Por conveniencia para realizar los cálculos, se dispone la información en la forma siguiente:

y_i	f_i	y_i^2	$f_i y_i$	$f_i y_i^2$
42	23	1764	966	40572
41	4	1681	164	6724
36	1	1296	36	1296
32	1	1024	32	1024
29	1	841	29	841
27	2	729	54	1458
23	1	529	23	529
19	1	361	19	361
16	2	256	32	512
15	2	225	30	450
14	1	196	14	196
11	1	121	11	121
10	1	100	10	100
9	1	81	9	81
7	1	49	7	49
6	3	36	18	108
5	2	25	10	50
4	1	16	4	16
3	1	9	3	9
Σ	50		1471	54497

$$\bar{Y} = \frac{1}{50} \sum_{i=1}^{19} f_i \, y_i = \frac{1471}{50} = 29.42$$

$$S_Y^2 = \frac{1}{50} \sum_{i=1}^{19} f_i \, y_i^2 - (\bar{Y})^2 = \frac{54497}{50} - (29.42)^2 =$$

$$= 1089.94 - 865.44 = 224.5$$

Entonces

$$\hat{Y} = N_p \, \bar{Y} = 676 \times 29.42 = 19888 \text{ firmas}$$

y, puesto que el error absoluto debe ser igual al 5%, se tendría

$$e = (0.05)\ (19888) = 995$$

Por otra parte, el tamaño inicial de muestra igual a 50 permite suponer que la estimación de σ_Y^2 de la población es suficientemente buena con S_Y^2 , y que la distribución muestral de totales puede aproximarse mediante la normal. Por lo tanto,

$$K = Z_{0.475} = 1.96$$

$$N_p = 676$$

$$\sigma_Y^2 \doteq S_Y^2 = 224.5$$

$$N_p^2 n_\infty = N_p^2 \frac{K^2\sigma_Y^2}{e^2} = \frac{(676)^2\ (1.96)^2\ (224.5)}{(995)^2} = 397.9$$

$$n = \frac{N_p^2 n_\infty}{1 + \frac{1}{N_p}(N_p^2 n_\infty - 1)} = \frac{397.9}{1 + \frac{1}{676}(397.9 - 1)}$$

$$= \frac{397.9}{1 + 0.58} = \frac{397.9}{1.58} = 251.83$$

$$\therefore\ n = 252 \text{ hojas}$$

4.3 Tamaño de una muestra aleatoria simple (Proporciones)

4.3.1 Antecedentes

Supóngase una población binomial de tamaño N_p tal que cada uno de sus elementos únicamente puede estar en una de dos clases: A o B (buenos o malos, negros o blancos, grandes o chicos, etc). La proporción de elementos de la población que están en la clase A es

$$P = \frac{A}{N_p}$$

y la proporción de elemtos que están en B es

$$Q = \frac{B}{N_p}$$

por lo cual

$$P + Q = \frac{A}{N_p} + \frac{B}{N_p} = 1 \quad ; \quad (A + B = N_p)$$

Si a todos los elementos X_i de la población que están en A se les asigna el valor 1 y a los de B el 0, se obtiene

$$P = \frac{A}{N_p} = \frac{\sum_{i=1}^{N_p} X_i}{N_p} = \mu_X$$

Es decir, la proporción puede considerarse un caso particular de la media cuando los elementos de la población son unos y ceros.

La variancia es

$$\sigma_X^2 = \frac{1}{N_p} \sum_{i=1}^{N_p} (X_i - P)^2$$

o sea

$$\sigma_X^2 = \frac{1}{N_p} \sum_{i=1}^{N_p} X_i^2 - P^2$$

Sin embargo, como X_i sólo puede ser igual a uno o cero, se tiene que $X_i = X_i^2$, por lo cual

$$\sigma_X^2 = \frac{1}{N_p} \sum_{i=1}^{N_p} X_i - P^2 = P - P^2 = P(1 - P) = PQ$$

En virtud de lo anterior, si se muestrea sin remplazo y con tamaño n de una población binomial finita, para estimar la proporción de elementos con cierta característica, se obtienen, considerando que la proporción se puede calcular como una media, los siguientes parámetros de la distribución muestral de proporciones

$$\mu_p = P$$

$$\sigma_p = \frac{\sigma_X}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}} = \sqrt{\frac{PQ}{n}} \cdot \sqrt{\frac{N_p - n}{N_p - 1}}$$

Si la población es infinita, se obtiene

$$\mu_p = P$$

$$\sigma_p = \frac{\sigma_x}{\sqrt{n}} = \sqrt{\frac{PQ}{n}}$$

estimándose P en ambos casos con el valor de p de la muestra, si se desconoce P de la población.

En la práctica se considera que la distribución muestral de proporciones es aproximadamente igual a la normal para tamaños de muestra mayores o iguales a 30 elementos.

4.3.2 Obtención del tamaño de la muestra

Aprovechando el hecho de que la proporción se puede calcular como una media simple, las ecs 4.3 y 4.4 se pueden emplear en este caso para obtener el tamaño de la muestra haciendo $\sigma_x^2 = PQ$. Entonces,

$$n_\infty = \frac{K^2\ PQ}{e^2} \tag{4.7}$$

para muestreo de población infinita, y

$$n = \frac{n_\infty}{1 + \frac{1}{N_p}(n_\infty - 1)}$$

para muestreo de población finita con tamaño N_p.

Usualmente se calcula primero el valor de n_∞, y si la fracción de muestreo es mayor de 0.05, se calcula a continuación el valor de n.

Ejemplo 4.5

En una colonia con 4000 casas se desea estimar el porcentaje de inquilinos que son a la vez propietarios de su casa, con un error estándar en la estimación no mayor del 1%. Se supone, de estudios semejantes, que el porcentaje real de inquilinos-propietarios se acerca al 10%. ¿Cuántas casas se deben muestrear para que se satisfaga la condición establecida?

Solución

El error estándar en la estimación de P de la población es

$$\sigma_p = \sqrt{\frac{PQ}{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$$

y no debe ser mayor en este caso del 1%. Por lo tanto, siendo $N_p = 4000$, $P = 0.1$ y $Q = 1 - P = 0.9$, se obtiene

$$0.01 = \sqrt{\frac{(0.1)\ (0.9)}{n}} \sqrt{\frac{4000 - n}{4000 - 1}}$$

Elevando al cuadrado y realizando operaciones algebraicas

$$0.0001 = \frac{0.09}{n} \ \frac{4000 - n}{3999}$$

$$0.0001 = \frac{360 - 0.09\ n}{3999\ n}$$

$$0.3999\ n = 360 - 0.09\ n$$

$$n(0.3999 + 0.09) = 360$$

$$n = \frac{360}{0.4899} = 734.84$$

$$\therefore\ n = 735 \text{ casas}$$

Ejemplo 4.6

En un estudio antropológico para estimar el porcentaje de habitantes de una isla con sangre del grupo O, se obtuvo una muestra aleatoria de 50 isleños, en la cual 22 de ellos pertenecen al grupo sanguíneo mencionado. Si en la isla habitan 3208 gentes, ¿cuál debe ser el tamaño de muestra mínimo para estimar con un error absoluto del 5% el valor real de P, suponiendo que el nivel de confianza es del 95%?

Solución

En este caso la proporción de la muestra es

$$p = \frac{22}{50} = 0.44$$

o sea

$$q = 1 - p = 1 - 0.44 = 0.56$$

Considerando que la muestra inicial es suficientemente grande, se aproxima mediante la distribución normal, obteniéndose

$$K = Z_{0.475} = 1.96$$

por lo cual

$$n_\infty = \frac{K^2\ PQ}{e^2} = \frac{K^2\ pq}{e^2} = \frac{(1.96)^2\ (0.44)\ (0.50)}{(0.05)^2}$$

$$= \frac{0.84515}{0.0025} = 338.06$$

$$\therefore\ n_\infty = 339$$

Como

$$fm = \frac{n_\infty}{N_p} = \frac{339}{3208} = 0.106 > 0.05$$

se corrige el valor anterior, obteniéndose finalmente

$$n = \frac{n_\infty}{1 + \frac{1}{N_p}\ (n_\infty - 1)} = \frac{339}{1 + \frac{1}{3208}\ (339-1)}$$

$$= \frac{339}{1{,}105} = 306{,}787$$

$\therefore$ n = 307 habitantes

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

FUNDAMENTOS DE LAS TECNICAS DE MUESTREO ESTADISTICO

PRACTICA DE MUESTREO

M. EN C. ADELA ABAD DE SERVIN

27 FEBRERO, 1979

PRACTICA DE MUESTREO.

ESTIMADORES DE RAZON

1. De una lista de 468 academias de 2 años de estudio fue sacada una muestra aleatoria simple de 100. La muestra contenía 54 instalaciones públicas y 46 privadas. Los datos para el número de estudiantes (y) y el número de profesores (x) se muestran a continuación.

	n	$\sum(y)$	$\sum(x)$
Pública	54	31 281	2 024
Privada	46	13 707	1 075

	$\sum(y^2)$	$\sum(yx)$	$\sum(x^2)$
Pública	29 881 219	1 729 349	111 090
Privada	6 366 785	431 041	33 119

a. Para cada tipo de instalación en la población, estimar la proporción (número de estudiantes/número de profesores)
b. Calcular los errores estándar de los estimadores.
c. Para las instalaciones públicas encontrar los límites de confianza del 95% para la proporción de estudiante/maestro en la población.

2. En un estudio realizado en una zona formada por 70 manzanas, se listaron las 3000 familias que la componían y se eligieron aleatoriamente 30 familias. A cada familia se le preguntó el número de miembros y el número de autos que tenían. Se obtuvieron los siguientes resultados, donde la y indica número de miembros y la x número de coches.

$$\sum y_i = 236 \qquad \sum x_i = 115 \qquad \sum y_i x_i = 685$$

$$\sum y_i^2 = 1494 \qquad \sum x_i^2 = 401$$

a. Estimar el número de miembros por auto en la población.
b. Encontrar el error estándar de su estimador.
c. Calcular intervalos de confianza del 95% para el número de miembros por auto en la población.

3. En una zona formada por 2450 manzanas se desea estimar el número total de niños en edad pre-escolar en 1977. De acuerdo con el Censo de Población de 1970, este total fue de 25,000 niños. Se selecciona una muestra aleatoria simple de 10 manzanas y se obtienen los siguieates datos;

i	1	2	3	4	5	6	7	8	9	10
y_i	15	10	15	14	16	10	44	10	5	5
x_i	20	15	12	13	10	5	6	4	0	5

Estimar el número total de niños en esta zona en 1977, su varianza e intervalos de confianza del 95%.

4. Una compañía desea estimar el monto promedio en dinero pagado a sus empleados en gastos médicos, durante los 3 primeros meses del presente año. Los promedios trimestrales están disponibles en los reportes fiscales del año anterior. De la población de 1000 empleados fue seleccionada una muestra aleatoria de 100 registros de empleados. Los resultados de la muestra son totalizados y se encuentra que el total para el presente trimestre es de $\sum_{i=1}^{100} y_i = 1750$ y el total correspondiente al trimestre del año anterior es de $\sum_{i=1}^{100} x_i = 1200$. Los gastos médicos de la población total, correspondientes al mismo trimestre en el año anterior fue de 12500. Estime el monto promedio en dinero pagado por la compañía en el primer trimestre del presente año y su error estándar.

Se proporcionan los siguientes datos:

$$\sum_{i=1}^{100} y_i^2 = 30650, \qquad \sum_{i=1}^{100} x_i^2 = 15620, \qquad \sum_{i=1}^{100} y_i x_i = 21859.35$$

5. En un campo de cebada se pesaron el grano, y_i, y el grano más la paja, x_i, en cada una de un gran número de unidades de muestreo localizadas al azar por todo el campo.

También se pesó la cosecha total (grano más paja) del campo completo. Se obtuvieron los siguientes datos:

$c_{yy} = 1.13 \qquad c_{yx} = 0.78 \qquad c_{xx} = 1.11$

Calcule la ganancia en precisión obtenida estimando el rendimiento en grano del campo, de la razón grano a cosecha total en lugar de usar el rendimiento medio de grano por unidad.

6. Se desea estimar el número total de niños en edad pre escolar en el Estado de México en 1979. De acuerdo a información obtenido del censo de Población de 1970, se sabe que este total fue de 25,000 niños. Se selecciona una muestra aleatoria simple de 10 manzanas y se investiga el número de niños en edad pre escolar en el año de 1979 por manzana. El mismo dato se obtiene del censo para 1970.

MANZANA	1	2	3	4	5	6	7	8	9	10
1979	15	10	15	14	16	10	4	10	5	5
1970	20	15	12	13	10	5	6	4	0	5

a. Estimar el número total de niños en edad pre escolar en 1979 em el Estado de México.

b. Calcular intervalos del 95% de confianza.

PRACTICA DE MUESTREO

MUESTREO ESTRATIFICADO ALEATORIO

1. La Primera Cirujana del Ejército de E. U. Privada de su Presea

Números aleatorios

04433	80674	24520	18222	10610	05794
60298	47829	72648	37414	75755	04717
67884	59651	67533	68123	17730	95862
89512	32155	51906	61662	64130	16688
32653	01895	12506	88535	36553	23757

★ Fue Sufragista y Usaba Traje Masculino

★ Una Vieja Historia de la Guerra Civil

★ Iniciativa Para Restituirle la Medalla

WASHINGTON, 24 de enero. (NYT)—Hace sesenta años la doctora Mary Edwards Walker, cirujana con nombramiento durante la Guerra Civil de Estados Unidos, fue privada de su Medalla de Honor por una junta revisora del gobierno. La doctora Walker es la única mujer que jamás haya recibido la presea, lo cual también puede ser la razón para que la perdiera.

No es la primera vez que se envía una petición al Comité de los Servicios Armados del Senado para que devuelva póstumamente a la doctora Walker la medalla que es la más alta distinción del país por valor en combate.

El mes pasado la Junta del Ejército para la Corrección de los Expedientes Militares, junta revisora que actúa en nombre del secretario del Ejército, Clifford L. Alexander Jr., celebró una audiencia para examinar el caso de la doctora Walker. El grupo envió su recomendación al secretario Alexander, si bien los portavoces del ejército se niegan a decir qué se recomendó ni cuál será la decisión del secretario.

La medalla fue conferida a la doctora Walker por el Presidente Andrew Johnson el 11 de noviembre de 1865. Los generales William T. Sherman y George H. Tomas habían recomendado la presea y el Presidente Lincoln había firmado el testimonio poco antes de su muerte.

La doctora Walker fue citada por su papel como la primera cirujana del Ejército de Estados Unidos. La cita original se ha perdido y no se sabe que existan copias.

1917 la Junta de Acción Adversa de la Medalla de Honor revocó la medalla, arguyendo que había encontrado ambigüedades en la situación de la doctora Walker como miembro del ejército. caso fue expuesto a Brooke por la señora Anne Walker, de Mt. Vernon, Virginia, quien se dice "sobrina lejana" de la doctora Walker y cuya campaña para la restitución de la medalla le lleva casi todo su tiempo.

"La doctora Mary perdió la medalla —dijo recientemente la señora Walker— sólo porque se había adelantado cien años a su época y eso nadie lo podía aceptar".

La señora Walker quizás tenga razón. La doctora Walker fue sufragista toda su vida y partidaria de la reforma del vestido de la mujer. Desde la época de la Guerra Civil usó pantalones de hombre y sacos de levita. Daba conferencias feministas ataviada con traje de gala masculino, con la Medalla de Honor pendiente de las anchas solapas.

Durante el decenio de... 1870, trabajó en la sede de las sufragistas en Washington, al lado de Susan B. Anthony, Lucy Stone, Mary Livermore y Belva Lockwood. Las mujeres se convirtieron en blanco favorito de los impugnadores. "Ese curioso antropoide", llamó un reportero de The New York Times a la doctora

1. Estimar el número medio de letras por renglón en el artículo de periódico que se adjunta, utilizando muestreo estratificado (2 estratos) y m.a.s. dentro de cada estrato. Una muestra de 10 renglones debe ser repartida con afijación proporcional.

Calcular intervalos del 95% de confianza para la media en consideración.

De manera muy breve vaya narrando en cada paso el procedimiento utilizado.

2. Uttilizando la misma muestra del ejercicio 1, estimar:

a. La proporción de vocales en el artículo de periódico que se adjunta e intervalos de confianza del 95%.

b. El número total de vocales en el artículo y su error estándar.

2. Una compañía desea estimar el número total de horas hombres perdidas, para un mes dado, a causa de accidentes entre sus empleados. Pero obreros, técnicos y administradores tienen distintas tasas de accidentes, por lo que se decide usar muestreo aleatorio estratificado considerando cada grupo un estrato separado. Información de años anteriores proporciona las varianzas para número de horas hombres perdidas por empleado en los 3 grupos y además se proporciona el tamaño de los estratos.

I(obreros)	II(técnicos)	III(administradores)
$s_1^2 = 36$	$s_2^2 = 25$	$s_3^2 = 9$
$N_1 = 132$	$N_2 = 92$	$N_3 = 27$

a. Repartir en los estratos una muestra de 30 empleados con afijación proporcional

b. Estimar el número total de horas hombres perdidas durante el mes dado y establezca intervalos de confianza del 95% para este total. Utilice la siguiente información obtenida de una muestra de 18 obreros, 10 técnicos y 2 administradores.

I(Obreros)			II(Técnicos)		III(Administradores)
8	24	0	4	5	1
0	16	32	0	24	8
6	0	16	8	12	
7	4	4	3	2	
9	5	8	1	8	
18	2	0			

3. Un muestrista tiene una población de $N=5$ elementos y lo máximo que puede tomar es una muestra de tamaño 2.
A continuación se indican los valores de las características para esas cinco unidades:

Unidad i	y_i
1	0
2	1
3	0
4	1
5	1

El muestrista tiene completa libertad para utilizar muestreo aleatorio simple, muestreo estratificado, etc.

a. Qué plan le dará una varianza mínima para el el estimador de $\bar{Y}$?

b. Qué formato toma la varianza de $\bar{y}_{est}$ en el caso de afijación proporcional si se supone que la variabilidad por unidad en cada estrato es constante?

c. Proponga un estimador insesgado de la variabilidad por unidad en el caso b.

d. Proponga las fórmulas para la estimación de una proporción y su varianza en el caso de muestreo estratificado con muestreo aleatorio simple en cada estrato.

4. *En una estratificación con dos estratos, los valores de W_h y S_h son como sigue:*

Estrato	W_h	S_h
1	0.8	2
2	0.2	4

Calcule los tamaños de muestra n_1 y n_2 en los dos estratos necesarios para satisfacer las siguientes condiciones. Cada caso requiere un cálculo separado(ignore el cpf).

a. El error estándar del estimador de la media de la población $\bar{y}_{est}$ debe ser 0.1 y el tamaño de la muestra total $n = n_1 + n_2$ debe ser minimizado.

b. El error estándar de la media estimada de cada estrato debe ser de 0.1.

5. ***Cuatro recipientes, que contienen un número igual (y muy grande) de repuestos representan cuatro turnos de producción de una fábrica. El número de unidades correspondientes a muestras tomadas al azar de cada recipiente y el número de defectuosos encontrados se incluyen a continuación:***

Recipiente	h	1	2	3	4
	n_h	200	200	200	200
	a_h	4	2	10	8

a. Compute una estimación insesgada de la proporción de defectuosos en la población total (4 recipientes).

b. Compute una estimación de la varianza de su estimación en (a).

6. Dada una población de N = 500 recipientes con 200 unidades cada uno y una muestra de n = 5 recipientes., se obtiene la siguiente información con p_i, proporción de defectuosos en el recipiente i de la muestra:

i	1	2	3	4	5
p_i	$\frac{2}{200}$	$\frac{3}{200}$	$\frac{8}{200}$	$\frac{1}{200}$	$\frac{2}{200}$

a. Estime el porcentaje de defectuosos para ese período de producción y la varianza de su estimación.

b. Si en la situación anterior se toma una muestra al azar de tamaño 2 en cada uno de los 500 recipientes y a_i indicase el número de defectuosos encontrados en cada recipiente (i = 1,..,500), proponga Ud. un estimador para el porcentaje de defectuosos y una expresión para la varianza de seu estimador.

c. ¿Cuál de los dos esquemas estimaría más adecuado?

1. En una ciudad pequeña, para efecto de estimación del ingreso medio por varón adulto, se define cada manzana como un conglomerado. Al numerar las manzanas se encuentra que en total hay 415. De ellas se extrae una m.a.s. de 25 manzanas. Los resultados obtenidos se presentan en el cuadro siguiente.
   a. Estime el ingreso medio por varón adulto en la ciudad y su e.e.
   b. Estime el ingreso total de los varones adultos en la ciudad y su e.e.
   c. Sabiendo que hay 2500 varones adultos en la ciudad, estime el ingreso total de los varones adultos en la ciudad y su e.e.

2. En adición a la información sobre ingreso de varones adultos, se les pregunta si viven en casa propia o rentada. Los resultados se presentan en el mismo cuadro. Estime la proporción de varones adultos en la ciudad que viven en casa rentada y establezca su e.e.

Conglomerado i	Nº de varones adultos m_i	ingreso total x congl. y_i	No. de varones adultos en casas rentadas (a_i)
1	8	96,000	4
2	12	121,000	7
3	4	42,000	1
4	5	65,000	3
5	6	52,000	3
6	6	40,000	4
7	7	75,000	4
8	5	65,000	2
9	8	45,000	3
10	3	50,000	2
11	2	85,000	1
12	6	43,000	3
13	5	54,000	2
14	10	49,000	5
15	9	53,000	4
16	3	50,000	1
17	6	32,000	4
18	5	22,000	2
19	5	45,000	3
20	4	37,000	1
221	6	51,000	3
22	8	30,000	3
23	7	39,000	4
24	3	47,000	0
25	8	41,000	3
	$\sum m_i = 151$	$\sum y_i = 1,329,000$	

3. El gerente de una editorial periodística desea estimar el número promedio de periódicos comprados por casa en una comunidad dada. Los costos de transportes de casa a casa son considerables, por lo que las 4000 casas en la comunidad se listan en 400 conglomerados geográficos de 10 casas cada uno, y se selecciona una m.a.s. de 4 conglomerados. Los resultados de la entrevistas son:

Conglom.	No. de periódicos										Total
1	1	2	1	3	3	2	1	4	1	1	19
2	1	3	2	2	3	1	4	1	1	2	20
3	2	1	1	1	1	3	2	1	3	1	16
4	1	1	3	2	1	5	1	2	3	1	20

Estime el número medio de periódico por casa en la comunidad y obtenga su error estandar.

4. En una ciudad pequeña se desea estimar el número medio de alumnos, por salón de clase, en las escuelas primarias. No se dispone de un listado de salones, pero se cuenta con un listado de 500 escuelas. La investigación se realiza considerando cada escuela un conglomerado y se selecciona una muestra de 25 escuelas y se obtienen los siguientes resultados:

ESCUELA (i)	No. DE SALONES POR ESCUELA(m_i)	No. TOTAL DE ALUMNOS por escuela (y_i).
1	12	520
2	6	310
.	.	.
.	.	.
25	8	380
	275	10500

Además $\sum y_i^2 = 2,100,000$ $\sum m_i^2 = 3050$ $\sum y_i m_i = 115,500$

Estimar:

a. El número medio de alumnos por salón de clase.

b. La varianza del número medio de alumnos por salón de clase.

c. Intervalos de confianza del 95% para el inciso a.

d. Si se sabe que el número total de salones es de 6000 en toda la población, estimar el número total de alumnos en la ciudad.

e. La varianza del inciso d.

f. El número medio de alumnos por escuela y su desviación estándar.

g. Si no se conoce el número total de salones en la comunidad, estimar el número total de alumnos en la ciudad y su error estándar.

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

FUNDAMENTOS DE LAS TECNICAS DE MUESTREO ESTADISTICO

CONGLOMERADOS
(Micas)

M. EN C. ADELA ABAD DE SERVIN

MARZO, 1979.

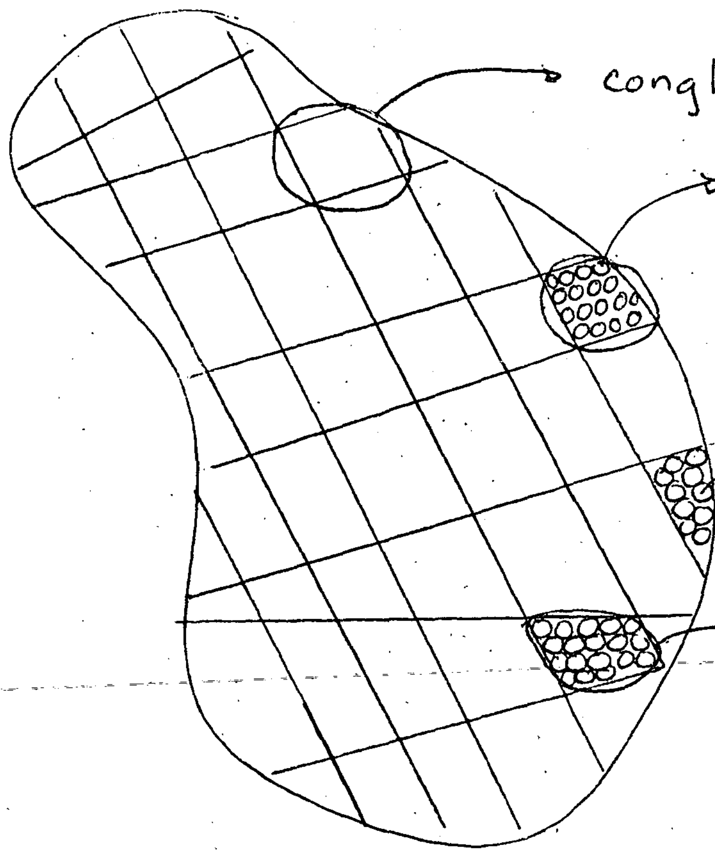
Muestreo por Conglomerados

Conceptos

- elemento
- unidad muestral
- conglomerado
- marco de referencia
- ventajas

Notación

Población : N conglomerados



conglomerado

M_i nº de elementos del conglomerado i-ésimo

$M = \sum_{i=1}^{N} M_i$ total de elem. en la pobl.

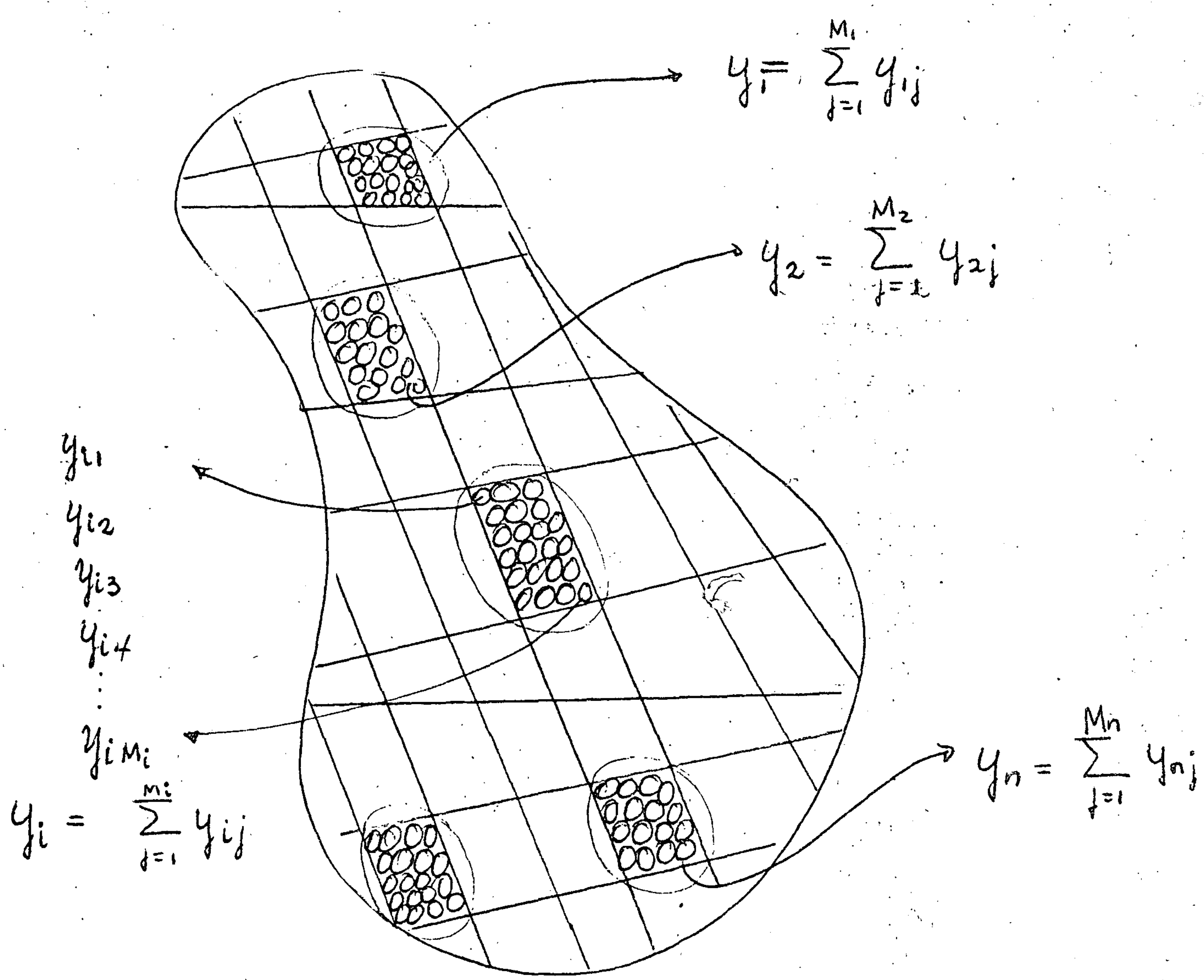
y_{ij} característica del elem. j-ésimo del conglomerado i-ésimo.

$y_i = \sum_{j=1}^{M_i} y_{ij}$ característica del congl. i-ésimo

$\bar{M} = \dfrac{M}{N}$ tamaño medio de los conglomerados en la población

... en conglomerados.

- Marco de referencia de conglomerados.
- Se selecciona una muestra de n conglomerados.



Los conglomerados que se seleccionan en la muestra se investigan totalmente.

- Parámetros a estimar:
  - Total: de la característica de los elementos en la población: Y
  - Media por unidad. $\bar{Y}$
  - Media por elemento $\bar{\bar{Y}}$
  - Proporción. P

Caso A: Conglomerados de tamaños desiguales.

- ESTIMADORES INSESGADOS y LAS ESTIMACIONES DE SUS VARIANZAS.

- MEDIA POR UNIDAD (conglomerado) $\hat{\bar{Y}} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ donde $y_i = \sum_{j=1}^{M_i} y_{ij}$

$$v(\hat{\bar{Y}}) = v(\bar{y}) = \frac{1-f}{n} \frac{\sum^{n} (y_i - \bar{y})^2}{n-1}$$

- TOTAL $\hat{Y} = N\bar{y}$ $\qquad v(\hat{Y}) = N^2 v(\bar{y})$

- MEDIA por elemento

$$\hat{\bar{\bar{Y}}} = \bar{\bar{y}} = \frac{\hat{Y}}{M} = \frac{1}{\bar{M}n}\sum_{i=1}^{n} y_i$$

$$v(\hat{\bar{\bar{Y}}}) = v(\bar{\bar{y}}) = \frac{1-f}{\bar{M}^2 n} \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$

- ESTIMADORES DE RAZÓN: La variable auxiliar es M_i tamaño del conglomerado.

- Media por elemento $\quad \hat{\bar{\bar{Y}}}_R = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} M_i}$, $v(\hat{\bar{\bar{Y}}}_R) \doteq \frac{1-f}{n\bar{M}^2} \frac{\sum_{i=1}^{n}(y_i - \bar{\bar{y}}_R M_i)^2}{n-1}$

- Total $\quad \hat{Y}_R = M\hat{\bar{\bar{Y}}}_R$, $v(\hat{Y}_R) \doteq M^2 v(\hat{\bar{\bar{Y}}}_R)$

- Proporción $\quad \hat{P}_R = p_R = \frac{\sum^{n} a_i}{\sum^{n} M_i}$, a_i nº de elementos del conglom. i-ésimo que pertenecen a la clase de interés.

$$v(p_R) \doteq \frac{1-f}{n\bar{M}^2} \frac{\sum a_i^2 - 2p_R \sum a_i M_i + p_R^2 \sum M_i^2}{n-1} , \quad \hat{\bar{M}} = \frac{\sum^{n} M_i}{n}$$

CASO B: Conglomerados de tamaños iguales

Población : N conglomerados

Muestra : n conglomerados

$M' = \bar{M}$ tamaños de los conglomerados

$M = NM'$ total de elem. en la población.

$$y_i = \sum_{j=1}^{M'} y_{ij}$$

Estimadores:

- Media por unidad (conglomerado) $\qquad \hat{\bar{Y}} = \bar{y} = \dfrac{\sum_{i=1}^{n} y_i}{n} \quad , \quad v(\bar{y}) = \dfrac{1-f}{n} \dfrac{\sum (y_i - \bar{y})^2}{n-1}$

- Total $\qquad \hat{Y} = N\bar{y} \quad , \quad v(\hat{Y}) = N^2 v(\bar{y})$

- Media por elemento $\qquad \hat{\bar{\bar{Y}}} = \dfrac{\sum^{n} y_i}{nM'} \quad , \quad v(\hat{\bar{\bar{Y}}}) = \dfrac{1}{(M')^2} v(\bar{y})$

- Proporción $\qquad \hat{P} = p = \dfrac{\sum^{n} a_i}{nM'}$

DIRECTORIO DE ASISTENTES AL CURSO: FUNDAMENTOS DE LAS TECNICAS DE MUESTREO ESTADISTICO

(DEL 13 DE FEBRERO AL 15 DE ABRIL DE 1979)

	NOMBRE Y DIRECCION	EMPRESA Y DIRECCION
1.	ACT. NANCY ARICEAGA RODRIGUEZ Calle 22 No. 15-4 Col. Sn. Pedro de los Pinos México 18, D.F. Tel. 271-24-12	ASEGURADORA NACIONAL AGRICOLA Y GANADERA, S.A. Benjamín Franklin No. 146 Col. Escandón Tel. 515-50-70 Ext. 133
2.	ING. GUSTAVO EDUARDO BARCENAS VILLA Av. Río San Javier No. 320-C-305 Col. Acueducto de Guadalupe México 14, D.F. Tel. 3-92-01-33 519-13-46	SECRETARIA DE ASENTAMIENTOS HUMANOS Y OBRAS PUBLICAS Av. Universidad y Xola Col. Narverte México, D.F. Tel. 519-13-46
3.	MAT. ROSA ELENA BARRIENTOS OLIVARES Privada Corina No. 47 Col. Carmen Coyoacán México 21, D.F. Tel. 554-48-54	BANCO DE MEXICO, S.A. Condesa No. 6-6o. Piso Col. Centro México 1, D.F. Tel. 512-50-58
4.	ING. ARMANDO CAMPOS PEREZ Neiva No. 1005-6 Col. Lindavista México 14, D.F. Tel. 754-19-40	NACIONAL DE COBRE Poniente 134 No. 719 Col. Ind. Vallejo México 16, D.F. Tel. 567-11-44
5.	LORENZO CARDENAS GUZMAN Av. Ermita Iztapalapa No. 920-81-304 Col. Barrio Sta. Bárbara México 13, D.F. Tel. 670-28-43	BANCO SOMEX, S.A. Paseo de la Reforma # 96-4o.Piso Col. Juárez México, D.F. Tel. 566-0233
6.	EDMUNDO CARMONA GARCIA Insurgentes Centro No. 60 Col. San Rafael México 4, D.F. Tel. 546-87-84	S.A.R.H. SUBDIRECCION DE PROGRAMAS Y PROMOCION Plaza de la República # 31-7o.Piso Col. Tabacalera México 1, D.F. Tel. 535-13-25

DIRECTORIO DE ASISTENTES AL CURSO: FUNDAMENTOS DE LAS TECNICAS DE MUESTREO ESTADISTICO

(DEL 13 DE FEBRERO AL 15 DE ABRIL DE 1979)

	NOMBRE Y DIRECCION	EMPRESA Y DIRECCION
7.	ACT. APOLINAR CHAVELAS CORTES Edif. Edo. de Gro. Ent. C-Dpto 11 Col. Tlatelolco México 3, D.F. Tel. 583-08-85	SECRETARIA DE ASENTAMIENTOS HUMANOS Y OBRAS PUBLICAS Col. Narvarte México, D.F. Tel. 530-99-85
8.	LIC. FRANCISCO J. CHAVEZ ZAMORA Apaninos No. 27 IV Sección Col. Lomas Verdes Naucalpan, Edo. de Méx. Tel. 393-05-61	BANCO DE MEXICO, S.A. Condesa No. 5-3o. Piso Mezzanine Col. Centro México 1, D.F. Tel. 518-05-00 Ext. 821-822
9.	ALFONSO DE LA ROSA SANCHEZ Fracc. Los Volcanes lote No. 56 Col. Volcanes Cuernavaca, Mor. Tel. 4-14-64	UNIVERSIDAD AUTONOMA DE MORELOS Chamilpa S/N Col. Buena Vista Cuernavaca, Mor. Tel. 3-10-90
10.	Q.F.B. MA. CARMEN ESTRADA O. Insurgentes Sur No. 1763-304 Col. Guadalupe Inn México 20, D.F.	LABORATORIO VALDECASAS Insurgentes Sur 4058 Col. Tlalpan México 22, D.F.
11.	ING. JESUS J. FUENTES CORDOVA Calz. de las Brujas No. 192 Col. Real del Sur Coapa México 22, D.F. Tel. 594-99-99	CIA. DE LUZ Y FUERZA DEL CENTRO Plaza Pie de la Cuesta No. 273 Col. Sn. Andrés Tetepilco México 13, D.F. Tel. 539-78-51
12.	GRACIELA E. GONZALEZ RODRIGUEZ Tokio No. 304 Altos Col. Portales México 13, D.F. Tel. 539-74-84	SECRETARIA DE ASENTAMIENTOS HUMANOS Y OBRAS PUBLICAS Constituyentes No. 947 Col. Belem de las flores México, D.F. Tel. 271-30-00

DIRECTORIO DE ASISTENTES AL CURSO: FUNDAMENTOS DE LAS TECNICAS DE MUESTREO ESTADISTICO

(DEL 13 DE FEBRERO AL 15 DE ABRIL DE 1979)

	NOMBRE Y DIRECCION	EMPRESA Y DIRECCION
13.	ING. DANIEL GUTIERREZ CARBAJAL Morelos No. 171 Col. Del Carmen Coyoacán México 21, D.F. Tel. 554-49-94	LABORATORIO DE LA COMPAÑIA DE LUZ Av. Pie de la Cuesta y Av. de las Torres Col. San Andrés México 13, D.F. Tel. 539-52-31
14.	JORGE HAM TRUJILLO Calle Sánchez Colín No. 13-B Col. Providencia Atz. México 16, D.F. Tel.	INSTITUTO NACIONAL DE INVESTIGACIONES Av. Progreso No. 5 P. baja Col. Coyoacán México 21, D.F. Tel. 658-06-43
15.	JOSE ANTONIO HERNANDEZ GARCIA Calle "U" 14-2 Col. Coyoacán México 21, D.F. Tel. 677-32-97	INSTITUTO MEXICANO DEL SEGURO SOCIAL Río Mississipi No. 50-8o. Piso Col. Centro México 1, D.F. Tel. 511-70-55
16.	BIOL. [illegible]ICITAS IBARRA GARCIA Pto. Progreso No. 90 Col. Casas Alemán México 14, D.F. Tel. 781-97-75	DIRECCION DE ESTUDIOS DEL TERRITORIO NACIONAL San Antonio Abad No. 124 Col. Tránsito México, D.F. Tel. 578-62-00
17.	ING. J. PRIMO JUAREZ HUIDOBRO Ixtapalapa No. 100-B Col. C.F.E. Toluca, Edo. de Méx. Tel. 485-70	CIA. NESTLE, TOLUCA Carretera México-Toluca Km. 61.5 Toluca, Edo. de Méx. Tel. 6-17-00 Y 6-17-31 Ext. 28
18.	ING. MATIAS LOPEZ JIMENEZ Calle Sur 111-A Núm. 328 Col. Héroes de Churubusco México 13, D.F. Tel. 582-97-15	DIRECCION GENERAL DE AEROPUERTOS S.A.H.O.P. Xola Núm. 1755-4o. Piso Col. Narvarte México 13, D.F. Tel. 519-80-46

DIRECTORIO DE ASISTENTES AL CURSO: FUNDAMENTOS DE LAS TECNICAS DE MUESTREO ESTADISTICO

(DEL 13 DE FEBRERO AL 15 DE ABRIL DE 1979)

	NOMBRE Y DIRECCION	EMPRESA Y DIRECCION
19.	SERGIO LOPEZ SIERRA	U.N.A.M. DEPTO. DE ESTADISTICA Ciudad Universitaria Torre de Rectoria
20.	MARIO MEDINA ROSALES Av. Fabián Flores No. 24 Col. San. Pablo Oztotepec México 23, D.F. Tel.	S.A.R.H. SUBDIRECCION DE USOS DEL AGUA Reforma No. 107-1o. Piso Col. San Rafael Tel. 566-95-58
21.	ACT. GILBERTO MENA FLORES Mitla No. 262 Col. Narvarte México 12, D.F. Tel. 579-85-37	INSTITUTO MEXICANO DEL PETROLEO Av. de los 100 metros No. 152 Col. Nva. Vallejo México 14, D.F. Tel. 567-66-00
22.	ING. ARTURO MENDEZ LOPEZ Sup. Corte No. 203-7 Col. Federal México 9, D.F.	S.A.H.O.P. DEPTO. DE LABORATORIO D.G.S.T. Av. Xola y Universidad Col. Narvarte México, D.F.
23.	LIC. FRANCISCO JAVIER MIRAVETE NOVELO 3a. Privada de Amores No. 21 Col. Del Valle México 12, D.F. Tel. 523-23-45	U.N.A.M. DEPTO. DE ESTADISTICA Ciudad Universitaria Torre de Rectoria Tel. 550-52-15
24.	GABINO GASPAR MONTERROSA REYES Saratoga No. 1017-201 Col. Portales México 13, D.F. Tel. 532-46-59	S.A.R.H. SIECA Reforma No. 107-1o. Piso Col. San Rafael México 4, D.F. Tel. 566-95-58

DIRECTORIO DE ASISTENTES AL CURSO: FUNDAMENTOS DE LAS TECNICAS DE MUESTREO ESTADISTICO
(DEL 13 DE FEBRERO AL 15 DE ABRIL DE 1979)

	NOMBRE Y DIRECCION	EMPRESA Y DIRECCION
25.	DR. HUGO JUAN PEDROZA ESCALERA Edificio E-15 Depto. 12 Col. Plateros México 19, D.F. Tel. 593-18-40	I.M.S.S. Marína Nacional y Mariano Escobedo Col. Anáhuac México, D.F. Tel. 5-27-61-74 Ext. 138
26.	FERNANDO RAMIREZ DELGADO Dibujantes No. 42 Col. Triunfo México 8, D.F. Tel. 670-18-34	ASEGURADORA NACIONAL AGRICOLA Y GANADERA, S.A. Benjamín Franklin No. 146 Col. Escandón México 8, D.F. Tel. 515-50-70 Ext. 133
27.	JORGE REYES CARMONA	SEGUROS LA PROVINCIAL, S.A. Miguel Angel de Quevedo No. 915 Tel. 549-30-20
28.	ARTURO RINCON VILLANUEVA Mar Egeo No. 179 Col. Popotla México 17, D.F. Tel. 527-75-05	CIA. NESTLE, S.A. Ejercito Nacional No. 453 Col. Granada México 17, D.F. Tel. 250-99-44
29.	ING. PEDRO RODARTE MIRELES Guanábana No. 310-3 Col. Nva. Sta. María México 16, D.F. Tel. 566-96-61	SECRETARIA DE AGRICULTURA Y RECURSOS HIDRAULICOS Reforma No. 107-1o. Piso Col. Sn. Rafael México 4, D.F. Tel. 566-96-61
30.	LIC. JOSE ROBLES TAMAYO Campanilla No. 45 Col. Jardines de Coyoacán México 22, D.F. Tel. 677-09-35	S.A.H.O.P. DIRECCION GENERAL DE AEROPUERTOS Xola No. 1755 Col. Narvarte México 12, D.F. Tel. 519-80-46

DIRECTORIO DE ASISTENTES AL CURSO: FUNDAMENTOS DE LAS TECNICAS DE MUESTREO ESTADISTICO

(DEL 13 DE FEBRERO AL 15 DE ABRIL DE 1979)

	NOMBRE Y DIRECCION	EMPRESA Y DIRECCION
31.	FIS. RAUL SIERRA OTERO Sgto. 2o. Miguel Angel Preciat No. 49 Col. Los Cipreses, Coy. México 21, D.F. Tel. 677-70-30	BANCO DE MEXICO, S.A. Condesa No. 6-6o. Piso Col. Centro México 1, D.F. Tel. 512-50-58
32.	LIC. PORFIRIO SILVA PEREZ 3 de Rifleros S/N L.P. Núm. 5A Col. Ejercito de Oriente México 13, D.F. Tel.	CENTRO DE EDUCACION CONTINUA Tacuba No. 5 Col. Centro México 1, D.F. Tel. 512-31-23
33.	LIC. FRIDA STAROPOLSKY N. Horacio No. 1605-16 Col. Polanco México 10, D.F. Tel. 540-28-86	DEPARTAMENTO DEL DISTRITO FEDERAL San Antonio Abad No. 231-4 Col. Cuauthémoc México 1, D.F. Tel. 761-17-48
34.	LIC. GUILLERMO TEJEIDA M. Hda. de Zotoluca No. 234 Col. Echegaray Naucalpan, Edo. de México Tel. 373-04-87	BANCO DE MEXICO, S.A. 5 de Mayo No. 2 Col. Centro México 1, D.F. Tel. 518-05-00 Ext. 821
35.	MAT. ERVIA YERENA O. Clemente Orozco No. 13-101 Col. Nápoles México 18, D.F. Tel. 598-62-18	BANCO DE MEXICO Condesa No. 6-5 Piso Col. Centro México 1, D.F. Tel. 585-42-99 Ext. 183