DIRECTORIO DE PROFESORES

PROBABILIDAD Y ESTADISTICA. FUNDAMENTOS Y APLICACIONES

DR. OCTAVIO A. RASCON CHAVEZ
INVESTIGADOR
INSTITUTO DE INGENIERIA, UNAM
CIUDAD UNIVERSITARIA
MEXICO 20, D.F.
TEL: 560. 26. 75

M. EN I. AUGUSTO VILLARREAL ARANDA
SECRETARIO ACADEMICO Y
JEFE DE LA SECCION DE MATEMATICAS
DIVISION DE ESTUDIOS SUPERIORES
FACULTAD DE INGENIERIA, UNAM
CIUDAD UNIVERSITARIA
MEXICO 20, D.F.
TEL: 548. 09. 50

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

PROBABILIDAD Y ESTADISTICA:
FUNDAMENTOS Y APLICACIONES

TEORIA DE PROBABILIDADES

DR. OCTAVIO A. RASCON CHAVEZ

JUNIO, 1978

TEORIA DE PROBABILIDADES

POR DR. OCTAVIO A. RASCON CH.

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

SIMBOLOS DE DESIGUALDADES:

$<$ menor que

$\leq$ menor o igual que

$>$ mayor que

$\geq$ mayor o igual que

$\neq$ diferente de

TEORIA DE CONJUNTOS

UN *CONJUNTO* ES UNA COLECCION BIEN DEFINIDA DE OBJETOS.

NOTACION: LOS CONJUNTOS SE DENOTAN USUALMENTE CON LETRAS MAYUSCULAS, Y SUS ELEMENTOS SE ANOTAN DENTRO DE UN PAR DE LLAVES.

EJEMPLOS

A) EL CONJUNTO DE NUMEROS ANOTADOS EN UN DADO ES

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

B) EL CONJUNTO DE LOS NUMEROS ENTEROS MENORES QUE 5 ES

$S = \{-\infty, \ldots, -3, -2, -1, 0, 1, 2, 3, 4\}$

ó $S = \{x: x \text{ ES ENTERO Y } x \leq 4\}$

C) EL CONJUNTO DE LOS NUMEROS ENTEROS POSITIVOS MENORES QUE 5 ES

$E = \{0, 1, 2, 3, 4\}$

$E = \{x: \text{ ES ENTERO Y } 0 \leq x \leq 4\}$

D) EL CONJUNTO DE LOS CONTINENTES ES

$C = \{$ASIA, EUROPA, AMERICA, AFRICA, OCEANIA$\}$

E) EL CONJUNTO DE MARCAS QUE TIENE UNA MONEDA ES

$M = \{$CARA, CRUZ$\}$

F) EL CONJUNTO DE NUMEROS MAYORES DE 5 PERO MENORES O IGUALES QUE 10

$S_1 = \{x: 5 < x \leq 10\}$

CONJUNTOS

- *FINITOS.-* CUANDO TIENEN UN NUMERO FINITO DE ELEMENTOS
- *INFINITOS.-* CUANDO TIENEN UN NUMERO INFINITO DE ELEMENTOS

SUBCONJUNTOS

PARA EXPRESAR QUE UN ELEMENTO *PERTENECE* A UN CONJUNTO SE USA EL SIMBOLO ϵ. PARA EXPRESAR QUE *NO PERTENECE* SE USA EL SIMBOLO $\not\epsilon$.

EJEMPLO

SI $S_1 = \{X: 5<X\leq 10\}$, ENTONCES

$3 \not\epsilon S_1$; $5 \not\epsilon S_1$; $8 \epsilon S_1$; $10 \epsilon S_1$

PARA EXPRESAR QUE UN CONJUNTO *ESTA CONTENIDO* EN OTRO SE USA EL SIMBOLO $\subset$; SI *NO ESTA CONTENIDO* SE USA EL SIMBOLO $\not\subset$.

PARA QUE UN CONJUNTO ESTE CONTENIDO EN OTRO SE REQUIERE QUE *TODOS* SUS ELEMENTOS LO ESTEN, ES DECIR, QUE TODOS SUS ELEMENTOS PERTENEZCAN A AMBOS CONJUNTOS.

EJEMPLO

SEAN $E=\{3,5\}$; $F=\{3,8\}$; $G=\{7,9\}$. $E\not\subset S_1$; $F\not\subset S_1$; $G\subset S_1$

SI UN CONJUNTO, B, ESTA CONTENIDO EN OTRO, S, SE DICE QUE B ES *SUBCONJUNTO* DE S.

EJEMPLO

$B = \{X: 3\leq X\leq 8\}$ Y $S_1=\{X: 5<X\leq 10\}$

EN ESTE CASO:

$G\subset S_1 \Rightarrow G$ ES SUBCONJUNTO DE S_1

$B\not\subset S_1 \Rightarrow B$ NO ES SUBCONJUNTO DE S_1

SE DICE QUE DOS CONJUNTOS SON IGUALES CUANDO CONTIENEN LOS MISMOS ELEMENTOS (NO IMPORTA EL ORDEN EN QUE ESTOS SE ESCRIBAN)

EJEMPLO

SEAN $A=\{1,3,5,7\}$, $B=\{7,5,1,3\}$ Y $C=\{7,5,1\}$

EN TAL CASO, $A = B \neq C$

CONJUNTO VACIO

DE LA MISMA MANERA QUE EXISTE EL CERO EN LOS NUMEROS, EN LA TEORIA DE CONJUNTOS EXISTE EL *CONJUNTO VACIO*, EL CUAL *NO* TIENE ELEMENTOS. USUALMENTE SE DENOTA $\emptyset$.

EJEMPLO

¿CUAL ES EL CONJUNTO DE ELEMENTOS, X, TALES QUE $2X=7$ Y X ES ENTERO?

SOLUCION.- ES EL CONJUNTO VACIO, $\emptyset$.

A $\emptyset$ SE LE CONSIDERA COMO SUBCONJUNTO DE CUALQUIER CONJUNTO. ASI, POR EJEM, TODOS LOS SUBCONJUNTOS DEL CONJUNTO

$S=\{2,5,10\}$ SON: $\{2\};\{5\};\{10\};\{2,5\};\{2,10\};\{5,10\};\{2,5,10\}$ Y $\emptyset$.

ESPACIO DE EVENTOS

ASOCIADO A UN EXPERIMENTO SIEMPRE HAY UN CONJUNTO DE RESULTADOS POSIBLES; A DICHO CONJUNTO SE LE LLAMA *ESPACIO DE EVENTOS*.

EJEMPLOS

EL ESPACIO DE EVENTOS ASOCIADO AL EXPERIMENTO DE LANZAR UN DADO Y ANOTAR LA CARA QUE QUEDA HACIA ARRIBA ES

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

EL ESPACIO DE EVENTOS CORRESPONDIENTE AL EXPERIMENTO DE LANZAR DOS DADOS Y ANOTAR LOS NUMEROS QUE QUEDAN HACIA ARRIBA ES

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

SI EN ESTE EXPERIMENTO LA OBSERVACION DE INTERES FUESE LA SUMA DE LOS DOS NUMEROS OBSERVADOS, ENTONCES EL ESPACIO DE EVENTOS DEL EXPERIMENTO SERIA

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

A TODO SUBCONJUNTO DE UN ESPACIO DE EVENTOS SE LE LLAMA *EVENTO*. A LOS EVENTOS QUE TIENEN UN SOLO ELEMENTO DEL ESPACIO SE LES LLAMA *EVENTOS SIMPLES*.

SI AL REALIZAR UN EXPERIMENTO SE OBSERVA UN ELEMENTO DEL EVENTO A, ENTONCES SE DICE QUE *OCURRIO* O *SE VERIFICO* EL EVENTO A. POR EJEMPLO, SI $A=\{2,4\}$ Y AL LANZAR UN DADO SE OBSERVA EL 2 O 4, SE DICE QUE OCURRIO EL EVENTO A; SI SE OBSERVA CUALQUIER OTRO NUMERO, ENTONCES SE DICE QUE *NO* OCURRIO A.

ESPACIOS DE EVENTOS:

- *DISCRETOS*.- SI SUS ELEMENTOS PUEDEN NUMERARSE O CONTARSE. TIENEN UN NUMERO FINITO O INFINITO NUMERABLE DE ELEMENTOS.
- *CONTINUOS*.- SI SUS ELEMENTOS *NO* PUEDEN ENUMERARSE. TIENEN UN NUMERO INFINITO NO NUMERABLE DE ELEMENTOS

EJEMPLO

LOS ESPACIOS DE EVENTOS $S_1=\{CARA, CRUZ\}$; $S_2=\{1,2,3,4,5,6,\}$;
$S_3=\{VERDE, ROJO\}$ SON DISCRETOS. LOS ESPACIOS DE EVENTOS
$S_4=\{X: -\infty<X\leq 0\}$; $S_5\{Z: Z\geq 3\}$; $S_6=\{Y:3\leq Y\leq 80\}$
SON CONTINUOS.

EJEMPLO

¿QUE TIPOS DE ESPACIOS DE EVENTOS CORRESPONDEN A LOS SIGUIENTES EXPERIMENTOS?

A) CONTEO DEL NUMERO DE GRANOS DE UNA MAZORCA DE MAIZ

$S=\{0,1,2,3,...,\infty\}$, ES DISCRETO E INFINITO

B) MEDICION DE LA LONGITUD DE UNA ESPIGA DE TRIGO

$S=\{X: 0<X<\infty\}$, X EN CM, ES CONTINUO E INFINITO

C) MEDICION DEL EFECTO DE UNA VACUNA, EN TERMINOS DE "EXITO" O "FRACASO"

S={EXITO, FRACASO} ES DISCRETO Y FINITO.

D) MEDICION DEL NUMERO DE MILIGRAMOS DE UN ANTIBIOTICO CONTENIDO EN UNA CAPSULA

$S=\{Y:0\leq Y<\infty\}$ Y en mg, ES CONTINUO E INFINITO.

COMPLEMENTO DE UN EVENTO

EL *COMPLEMENTO* DE UN EVENTO A ES OTRO EVENTO QUE CONTIENE TODOS LOS ELEMENTOS DEL ESPACIO DE EVENTOS CORRESPONDIENTE QUE NO ESTAN EN A. USUALMENTE SE DENOTA CON UNA TILDE SOBRE EL SIMBOLO QUE CORRESPONDE AL EVENTO QUE COMPLEMENTA, $\bar{A}$.

EJEMPLOS

SI $S=\{1,2,3,4,5,6\}$ Y $A=\{1,3,5\}$ ENTONCES $\bar{A}=\{2,4,6\}$.

SI $S=\{X: 0\leq X\leq 58\}$ Y $A=\{X: 3<X\leq 17\}$, ENTONCES $\bar{A}=\{X: 0\leq X\leq 3,\ 17<X\leq 58\}$

EVENTOS MUTUAMENTE EXCLUSIVOS

CUANDO DOS O MAS EVENTOS NO PUEDEN OCURRIR SIMULTANEAMENTE AL REALIZAR UNA SOLA VEZ UN EXPERIMENTO, SE DICE QUE ESTOS SON *MUTUAMENTE EXCLUSIVOS*, ES DECIR, DOS EVENTOS SON MUTUAMENTE EXCLUSIVOS CUANDO NO TIENEN NI UN SOLO ELEMENTO EN COMUN.

EJEMPLO

A) CUALQUIER EVENTO Y SU COMPLEMENTO SON MUTUAMENTE EXCLUSIVOS.

B) ¿SON $E=\{Y: 0\leq Y\leq 25\}$ Y $A=\{2,50,100\}$ MUTUAMENTE EXCLUSIVOS?

NO, PORQUE TIENEN EL ELEMENTO 2 EN COMUN.

OPERACIONES CON EVENTOS

UNION

LA *UNION* DE DOS EVENTOS ES OTRO EVENTO CUYOS ELEMENTOS SON TODOS LOS DE AMBOS. LA OPERACION DE UNION SE DENOTA CON EL SIMBOLO U.

EJEMPLOS

A) SI $A=\{2,4,6\}$ Y $B=\{1,6,12\}$, ENTONCES

$C=A\cup B=\{1,4,6,12,2\}$

B) ¿SON A Y B MUTUAMENTE EXCLUSIVOS? NO PORQUE TIENEN EL 6 EN COMUN.

C) SI $D=\{Y: 0\leq Y\leq 13\}$ Y $E=\{Y: 20\leq Y\leq 50\}$,

ENTONCES

$D\cup E=\{Y: 0\leq Y\leq 13,\ 20\leq Y\leq 50\}$

D) SI $F=\{Y: 8\leq Y\leq 20\}$, ENTONCES

$D\cup F=\{Y;\ 0\leq Y\leq 20\}$.

E) SI $G=\{Y: 3\leq Y\leq 10\}$, ENTONCES

$D\cup G=\{Y: 0\leq Y\leq 13\}=D$; OBSERVESE QUE EN ESTE CASO $G\subset D$. EN GENERAL, SI $A\subset B$, ENTONCES $A\cup B=B$.

EN GENERAL, LA UNION DE VARIOS EVENTOS ES OTRO EVENTO CUYOS ELEMENTOS SON TODOS LOS DE LOS EVENTOS QUE SE UNEN.

EJEMPLO

AUBUF = K = {1,2,4,6, y: 8$\leq$ y $\leq$20}

INTERSECCION

LA *INTERSECCION* DE DOS EVENTOS ES EL CONJUNTO DE ELEMENTOS QUE PERTENECEN SIMULTANEAMENTE A AMBOS. PARA DENOTAR LA OPERACION DE INTERSECCION SE USA EL SIMBOLO $\cap$.

EJEMPLOS

A) A = {2,3,6} Y B = {2,6,10} ENTONCES A$\cap$B = C = {2,6}

B) D = {Y: 4$\leq$Y$\leq$5}, ENTONCES A$\cap$D = Ø.

OBSERVESE QUE EN ESTE EJEMPLO A Y D SON MUTUAMENTE EXCLUSIVOS, YA QUE *NO* TIENEN NINGUN ELEMENTO EN COMUN. SIEMPRE QUE DOS EVENTOS SON MUTUAMENTE EXCLUSIVOS, SU INTERSECCION ES EL CONJUNTO VACIO.

EN GENERAL, LA *INTERSECCION* DE VARIOS EVENTOS ES EL CONJUNTO DE ELEMENTOS QUE TODOS ELLOS TIENEN EN COMUN.

EJEMPLO

SI A = {2,3,6,8}; B={2,3,10,100}; C={Y: 0$\leq$Y$\leq$5} Y D={Y: 2$\leq$Y$\leq$4}, ENTONCES

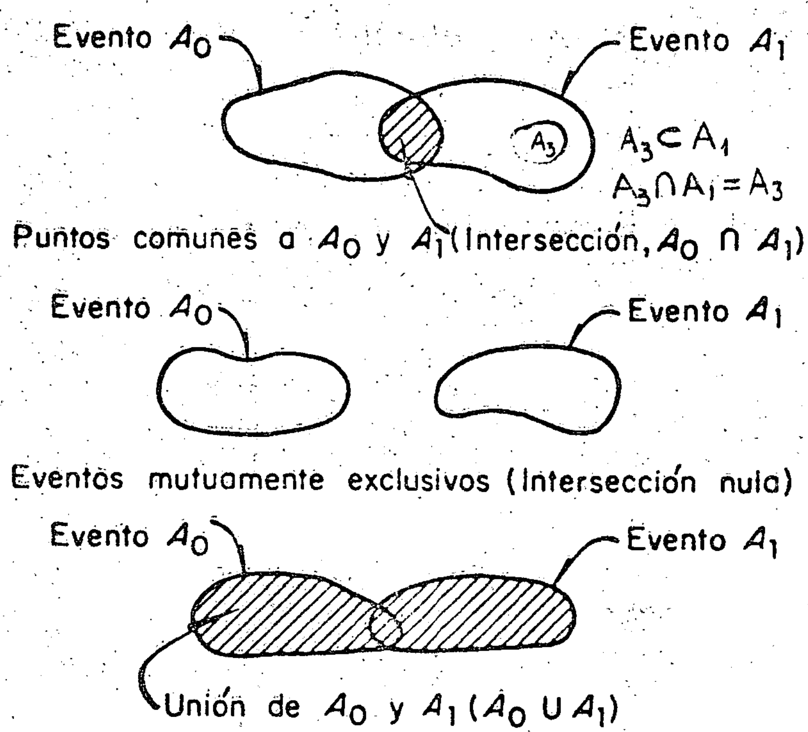
A$\cap$B$\cap$C$\cap$D = E ={2,3}

A$\cup$B$\cup$C$\cup$D = F ={Y: 0$\leq$Y$\leq$5, 6,8,10,100}

LA OCURRENCIA DE UN EVENTO "Y" OTRO IMPLICA LA OCURRENCIA DE AMBOS A LA VEZ, ES DECIR, QUE SE VERIFIQUE LA *INTERSECCION*. LA OCURRENCIA DE UN EVENTO "O" ALGUN OTRO, IMPLICA LA OCURRENCIA DE CUALQUIERA DE ELLOS, ES DECIR DE LA *UNION*.

DIAGRAMAS DE VENN

UNA MANERA DE ILUSTRAR GRAFICAMENTE LAS OPERACIONES CON CONJUNTOS ES MEDIANTE LOS *DIAGRAMAS DE VENN*. EN ESTOS, CADA CONJUNTO SE REPRESENTA POR UNA CURVA CERRADA QUE ENCIERRA LOS ELEMENTOS QUE LE CORRESPONDEN.



Diagramas de Venn (unión e intersección de eventos)

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 *FRECUENCIA RELATIVA* 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, $P(A) = \lim_{N\to\infty} \frac{N(A)}{N}$

EJEMPLO

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 TANTO

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

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.

AXIOMAS DE LA TEORIA DE PROBABILIDADES

LAS PROBABILIDADES QUE SE ASIGNAN A LOS DIFERENTES EVENTOS RELACIONADOS CON UN FENOMENO ALEATORIO DEBEN CUMPLIR CON LOS SIGUIENTES TRES AXIOMAS:

AXIOMA 1: LA PROBABILIDAD DE OCURRENCIA DE UN EVENTO A ES UN NUMERO, P(A), QUE SE LE ASIGNA A DICHO EVENTO, CUYO VALOR QUEDA EN EL INTERVALO

$$0 \leq P(A) \leq 1$$

AXIOMA 2: SI S ES UN ESPACIO DE EVENTOS, ENTONCES

$$P(S)=1$$

AXIOMA 3: LA PROBABILIDAD, P(C), DE LA UNION, C, DE DOS EVENTOS MUTUAMENTE EXCLUSIVOS, A Y B, ES IGUAL A LA SUMA DE LAS PROBABILIDADES DE ESTOS, ES DECIR,

$$P(A \cup B) = P(C) = P(A) + P(B)$$

EJEMPLOS

A) EN EL PROBLEMA DEL LANZAMIENTO DE UN DADO QUE NO ESTA CARGADO SE PUEDE ASIGNAR A CADA NUMERO (A CADA EVENTO SIMPLE) UNA PROBABILIDAD DE 1/6, SI $A=\{2,4\}$ Y $B=\{5,6\}$, ENTONCES, PUESTO QUE $A=\{2\}\cup\{4\}$ Y $B=\{5\}\cup\{6\}$, Y QUE LOS EVENTOS ELEMENTALES SON MUTUAMENTE EXCLUSIVOS ENTRE SI, APLICANDO EL AXIOMA 3 SE OBTIENEN:

$$P(A)=P(\{2\}) + P(\{4\}) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

$$P(B)=P(\{5\}) + P(\{6\}) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

SI $C=A\cup B$, Y DADO QUE A Y B SON EVENTOS MUTUAMENTE EXCLUSIVOS:

$$P(C) = P(A) + P(B) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

ADEMAS, OBSERVESE QUE SE CUMPLE CON LOS AXIOMAS 1 Y 2, YA QUE

$$P(S) = P(\{1\}) + P(\{2\}) + P(\{3\}) + P(\{4\}) + P(\{5\}) + P(\{6\})$$

$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{6}{6} = 1$$

EJEMPLO

EN EL PROBLEMA DEL LANZAMIENTO DE DOS DADOS, ¿CUAL ES LA PROBABILIDAD QUE AL REALIZAR UNA VEZ EL EXPERIMENTO LA SUMA DE LOS DOS NUMEROS QUE QUEDEN HACIA ARRIBA SEA 7 U 11? ESTO ES EQUIVALENTE A PREGUNTAR POR LA PROBABILIDAD DE QUE OCURRA EL EVENTO $C = \{7\} \cup \{11\}$. PUESTO QUE $\{7\}$ Y $\{11\}$ SON EVENTOS MUTUAMENTE EXCLUSIVOS:

$$P(C) = P(\{7\}) + P(\{11\}) = \frac{6}{36} + \frac{2}{36} = \frac{8}{36} = \frac{2}{9}$$

EJEMPLO

EN UN LABORATORIO SE PROBARON 100 VIGAS DE CONCRETO REFORZADO NOMINALMENTE IDENTICAS, Y SE ANOTARON LAS CARGAS CON LAS CUALES FALLO CADA UNA. DE ESTA SUCESION DE EXPERIMENTOS SE ASIGNARON, EN TERMINOS DE LAS FRECUENCIAS RELATIVAS CORRESPONDIENTES, LAS SIGUIENTES PROBABILIDADES:

SI $A = \{X: 0 \leq X \leq 20 \text{ ton}\}$; $P(A) = 0.17$

SI $B = \{X: 20 < X \leq 40 \text{ ton}\}$; $P(B) = 0.24$

SI $C = \{X: 40 < X \leq 60 \text{ ton}\}$; $P(C) = 0.27$

SI $D = \{X: 60 < X \leq 80\}$; $P(D) = 0.13$

SI $E = \{X: 80 < X \leq 100\}$; $P(E) = 0.11$

SI $F = \{X: 100 < X\}$; $P(F) = 0.08$

$\Sigma P(.) = 1.00$

SI SE REALIZA UNA VEZ MAS EL EXPERIMENTO, CALCULEMOS LAS SIGUIENTES PROBABILIDADES:

A) QUE LA RESISTENCIA SEA MENOR O IGUAL QUE 80 TON. PUESTO QUE $G = \{X: 0 \leq X \leq 80 \text{ ton}\}$ SE TIENE QUE $G = A \cup B \cup C \cup D$, POR LO QUE

$P(G) = P(A) + P(B) + P(C) + P(D) = 0.17 + 0.24 + 0.27 + 0.13 = 0.81$

B) LA PROBABILIDAD QUE RESISTA MAS DE 60 TONS. PUESTO QUE $H = \{X: 60 < X \leq \infty\}$ O $H = \{X: X > 60\}$ SE TIENE QUE $H = D \cup E \cup F$,

POR LO QUE $P(H) = P(D) + P(E) + P(F) = 0.13 + 0.11 + 0.08 = 0.32$

C) LA PROBABILIDAD QUE RESISTA MAS DE 40 TON, PERO CUANDO MUCHO 100 TON.

PUESTO QUE $I = \{X: 40 < X \leq 100\}$ SE TIENE QUE $I = C \cup D \cup E$

POR LO QUE $P(I) = P(C) + P(D) + P(E) = 0.27 + 0.13 + 0.11 = 0.51$

TEOREMAS

DOS TEOREMAS IMPORTANTES QUE SE DEDUCEN A PARTIR DE LOS AXIOMAS SON:

TEOREMA 1.

SI A ES UN EVENTO DEL ESPACIO S, ENTONCES $P(\bar{A})=1-P(A)$

DEMOSTRACION

PUESTO QUE A Y $\bar{A}$ SON MUTUAMENTE EXCLUSIVOS Y ADEMAS $A \cup \bar{A}=S$, ENTONCES, $P(S)=P(A)+P(\bar{A})=1$

$\Rightarrow P(\bar{A})=1-P(A)$

CASO PARTICULAR: PUESTO QUE $P(\bar{S})=1-P(S)=0$ Y $\bar{S}=\emptyset$, SE TIENE QUE $P(\emptyset)=0$

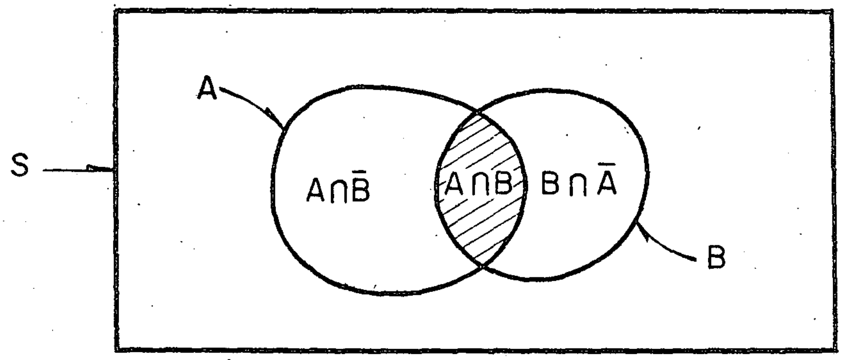
TEOREMA 2.

SI A Y B SON DOS EVENTOS CUALQUIERA DE S, ENTONCES

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

DEMOSTRACION

SEA EL DIAGRAMA DE VENN:



$A \cup B=(A \cap B) \cup (A \cap \bar{B}) \cup (B \cap \bar{A})$. PUESTO QUE $A \cap B$, $A \cap \bar{B}$ Y $B \cap \bar{A}$ SON MUTUAMENTE EXCLUSIVOS, SE TIENE QUE $P(A \cup B)=P(A \cap B)+P(A \cap \bar{B})+P(B \cap \bar{A})$.

SUMANDO Y RESTANDO $P(A \cap B)$ Y AGRUPANDO TERMINOS SE OBTIENE

$P(A \cup B) = [P(A \cap B)+P(A \cap \bar{B})]+[P(A \cap B)+P(B \cap \bar{A})]-P(A \cap B)$

PERO $A=(A \cap \bar{B}) \cup (A \cap B) \Rightarrow P(A \cap B)+P(A \cap \bar{B})=P(A)$

Y $B=(A \cap B) \cup (B \cap \bar{A}) \Rightarrow P(A \cap B)+P(B \cap \bar{A})=P(B)$, POR LO QUE

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

EJEMPLO

EN UNA URNA SE TIENEN 28 TIRAS DE PAPEL Y EN CADA UNA SE ENCUENTRA ANOTADA UNA LETRA DISTINTA DEL ALFABETO. CALCULE LA PROBABILIDAD DE QUE AL EXTRAER AL AZAR UNA TIRA:

A) SE OBTENGA UNA VOCAL

B) SE OBTENGA a O z

C) OCURRAN C Y D, DONDE $C=\{x,y,z\}$ y
$D=\{b,c,y,z\}$

D) OCURRA C O D

Respuestas

A) $A=\{a,e,i,o,u\} \Rightarrow P(A)=\frac{5}{28}$

B) $B=\{a,z\} \Rightarrow P(B) = \frac{2}{28}$

C) $F= C\cap D= \{y,z\} \Rightarrow P(F) = \frac{2}{28}$

D) $E= C\cup D= \{b,c,x,y,z\} \Rightarrow P(E) = \frac{5}{28}$

ó $P(E)=P(C)+P(D)-P(C\cap D)$

$P(C\cap D)=P(F)= \frac{2}{28} \Rightarrow P(E)= \frac{3}{28} + \frac{4}{28} - \frac{2}{28} = \frac{5}{28}$

REGLAS DE CONTEO

AL ASIGNAR PROBABILIDADES A LOS EVENTOS APLICANDO LA TEORIA CLASICA ES NECESARIO CALCULAR N(A) Y N PARA APLICAR LA FORMULA $P(A)=N(A)/N$.

SEAN, POR EJEMPLO, LOS EVENTOS $A=\{b,c\}$ Y $B=\{a,e,i,o,u\}$ CON LOS CUALES SE FORMAN PALABRAS DE DOS LETRAS, LA PRIMERA DE A Y LA SEGUNDA DE B. EL EVENTO QUE SE FORMA ASI ES

$$C=\{xy:\ x\varepsilon A;\ y\varepsilon B\}$$

SI ENUMERAMOS LOS ELEMENTOS:

$$\left.\begin{array}{l}\text{CON LA b: ba,be,bi,bo,bu}\\ \text{CON LA c: ca,ce,ci,co,cu}\end{array}\right\}\quad \text{10 ELEMENTOS}$$

SIN EMBARGO, LA SOLUCION SE PUEDE OBTENER RAPIDAMENTE SIN NECESIDAD DE ENUMERAR TODAS LAS POSIBILIDADES, OBSERVANDO QUE LA PRIMERA LETRA SOLO PUEDE SER DE DOS TIPOS b O c, MIENTRAS QUE LA SEGUNDA, DE CINCO TIPOS a,e,i,o,u, POR LO QUE EL TOTAL DE ELEMENTOS ES 2x5=10, ES DECIR, EL EVENTO C PUEDE OCURRIR DE 10 MANERAS DISTINTAS E IGUALMENTE PROBABLES.

REGLA DE LA MULTIPLICACION

EN GENERAL, SI DOS EVENTOS, A Y B, PUEDEN OCURRIR DE N(A) Y N(B) MANERAS DISTINTAS, RESPECTIVAMENTE, ENTONCES EL TOTAL DE MANERAS EN QUE AMBOS PUEDEN OCURRIR, EN EL ORDEN INDICADO, ES N(A) X N(B). ESTA REGLA SE PUEDE GENERALIZAR A MAS DE DOS EVENTOS.

EJEMPLO

¿CUANTOS NUMEROS PARES DE TRES CIFRAS SE PUEDEN FORMAR UTILIZANDO LOS DIGITOS 5,6,7,8 y 9, SIN QUE SE USE EL MISMO DIGITO EN LAS DECENAS Y LAS CENTENAS?

SOLUCION.- SEAN LOS EVENTOS

$A = \{X: X$ ESTA EN LAS CENTENAS$\}$

$B = \{Y: Y$ ESTA EN LAS DECENAS$\}$

$C = \{Z: Z$ ESTA EN LAS UNIDADES Y ES PAR$\}$

$D = \{XYZ: X \varepsilon A;\ Y \varepsilon B;\ Z \varepsilon C\}$

PUESTO QUE NO SE PERMITE REPETICION DE DIGITOS, $N(A)=5$ Y $N(B)=4$. ADEMAS, PUESTO QUE EL NUMERO DEBE SER PAR, $N(C)=2$. POR LO TANTO

$N(D) = 5x4x2=40$

SI EL ULTIMO DIGITO NO TUVIESE QUE SER PAR: $S=\{XYZ: X \varepsilon A; Y \varepsilon B; Z \varepsilon F\}$

DONDE $F=\{Z: Z$ ESTA EN LAS UNIDADES$\}$

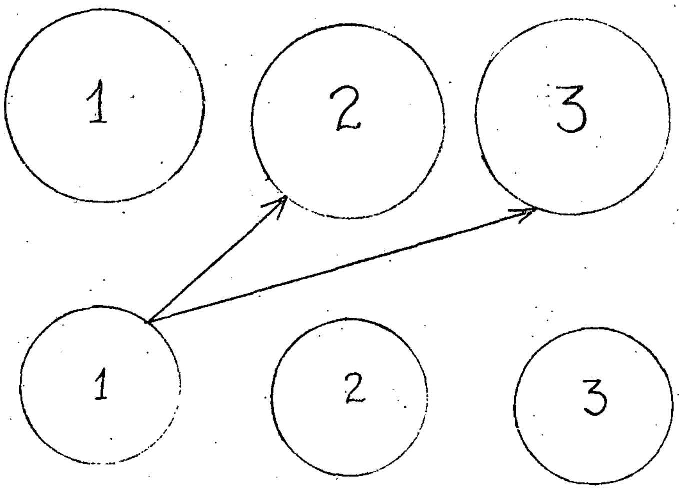
ENTONCES $N(F)=5$ Y $N(S)=5x4x5=100$.

CON ESTO, CALCULEMOS LA PROBABILIDAD DE QUE SI EL ESPACIO DE EVENTOS ES S Y SE ANOTAN TODOS LOS NUMEROS DEL MISMO EN UNA TIRA DE PAPEL, AL SACAR UNA AL AZAR DE UNA URNA, EL NUMERO SEA PAR:

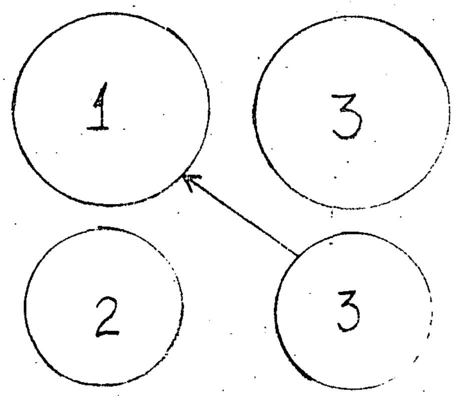
$$P(A) = \frac{N(A)}{N} = \frac{N(A)}{N(S)} = \frac{40}{100} = 0.4 = 40\%$$

EJEMPLO

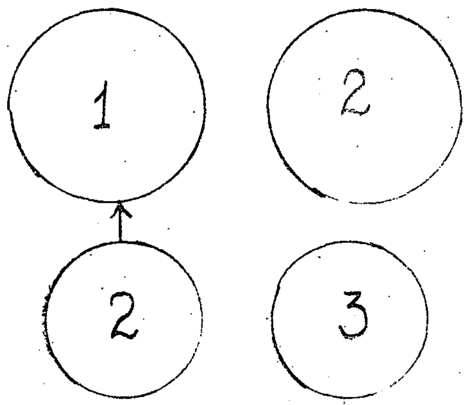
EN UNA CAJA SE TIENEN TRES PERFORACIONES NUMERADAS DEL UNO AL TRES. SI SE HECHAN EN ELLA TRES BOLAS TAMBIÉN NUMERADAS DEL 1 AL 3 Y SE AGITA LA CAJA, CALCULAR LA PROBABILIDAD DE QUE NINGUNA BOLA CAIGA EN LA PERFORACION QUE TIENE SU NUMERO (EVENTO A)



2 POSIBILIDADES



1 POSIBILIDAD



1 POSIBILIDAD

$$N(A) = 2x1x1=2$$

$$N = 3x2x1=6$$

$$P(A) = \frac{N(A)}{N} = \frac{2}{6} = \frac{1}{3}$$

EJEMPLO

SE DISPONE DE TRES BANDERAS: UNA BLANCA, UNA NEGRA Y UNA VERDE.

1. SI CADA PAREJA DE BANDERAS DE DISTINTO COLOR CONSTITUYE UNA SEÑAL, ¿CUANTAS SEÑALES SE PUEDEN HACER SI EL ORDEN DE COLOCACION DE LAS BANDERAS ES IMPORTANTE (EVENTO A)?

   $N(A) = 3x2=6$

2. SI TRES BANDERAS TAMBIEN CONSTITUYEN UNA SEÑAL CUANDO TODAS SON DE DIFERENTE COLOR ¿CUANTAS SEÑALES PODEMOS HACER CON LAS 3 BANDERAS A LA VEZ (EVENTO B)?

   $N(B) = 3x2x1=6$

3. ¿CUANTAS SEÑALES SE PUEDEN HACER CON DOS O TRES BANDERAS EN LAS CONDICIONES ANTERIORES (EVENTOS C)? $C = A \cup B$

   $N(C) = N(A) + N(B) = 6+6=12$

4. SI CADA SEÑAL DEL EVENTO C SE DIBUJA EN UNA TIRA DE PAPEL Y LUEGO SE COLOCAN EN UNA URNA, ¿CUAL ES LA PROBABILIDAD DE QUE SI SE TOMA UNA AL AZAR,

   A) SALGA UNA SEÑAL ESPECIFICA (EVENTO F): $(S=C)$

   $P(F) = N(F)/N(C) = 1/12$

   B) SALGA UNA SEÑAL CON DOS BANDERAS POR LO MENOS (EVENTO G):

   $G=C \Rightarrow N(G)=12 \Rightarrow P(G) = \frac{12}{12} = 1$

   C) SALGA UNA SEÑAL CON DOS BANDERAS, UNA DE ELLAS VERDE (EVENTO H):

   $N(H) = 1x2+2x1=4 \Rightarrow P(H) = \frac{4}{12} = \frac{1}{3}$

   D) SALGA UNA SEÑAL CON TRES BANDERAS, UNA DE ELLAS VERDE (EVENTO I)

   $N(I)= 1x2x1+ 1x1x2+ 2x1x1 =6$

   $P(I)= 6/12 = 1/2 = 50\%$

E) SALGA UNA SEÑAL CON DOS O TRES BANDERAS EN QUE SE USE UNA VERDE (EVENTO J)

$J = H \cup I \Rightarrow N(J) = N(H)+N(I) = 4+6=10$

$P(J) = 10/12 = 5/6$

PERMUTACIONES

EL ARREGLO DE N OBJETOS EN CIERTO ORDEN SE DENOMINA *PERMUTACION.* POR EJEMPLO, TODAS LAS PERMUTACIONES QUE PUEDEN HACERSE CON LAS LETRAS A,B,C SON: ABC, ACB, BAC, BCA, CAB, CBA. EL TOTAL ES 3x2x1=6 PERMUTACIONES (N=3).

EN GENERAL, EL NUMERO DE PERMUTACIONES ES $N(N-1)(N-2)(N-3)\times\ldots\times 1=N!$

EJEMPLO

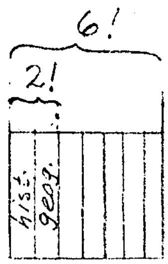
¿CUANTAS PERMUTACIONES SE PUEDEN HACER CON 5 OBJETOS?

$5! = 5\times4\times3\times2\times1=120$

EJEMPLO

EN UN LIBRERO SE COLOCARAN AL AZAR 7 LIBROS. CALCULEMOS LA PROBABILIDAD DE QUE EL DE HISTORIA Y EL DE GEOGRAFIA QUEDEN JUNTOS (EVENTO A).

$P(A) = N(A)/N$

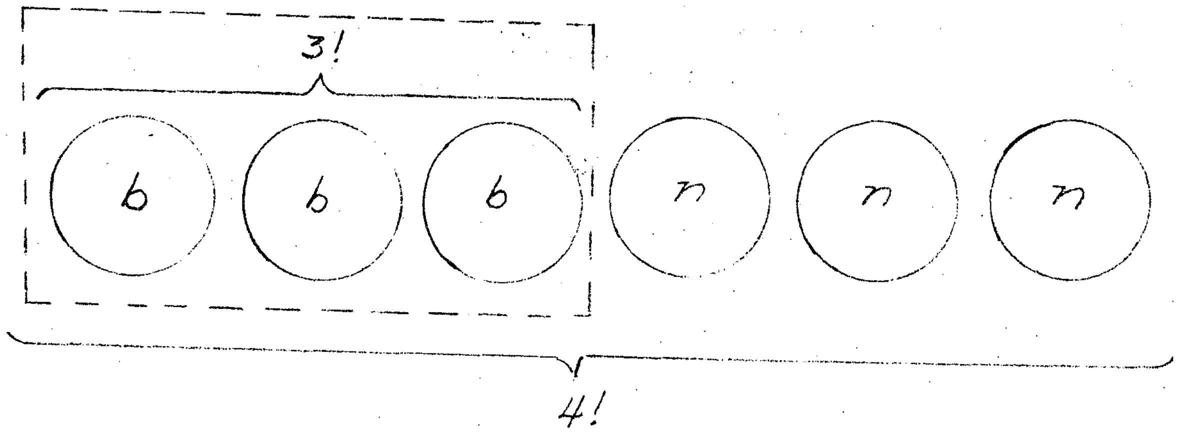


$N=7! = 7\times(6\times5\times4\times3\times2\times1) = 7\times6!$

$N(A) = 2!\times6!$; $P(A) = \dfrac{2!\times6!}{7!} = \dfrac{2!\times6!}{7\times6!} = \dfrac{2}{7}$

EJEMPLO

EN UNA URNA SE TIENEN 6 ESFERAS, DE LAS CUALES 3 SON BLANCAS Y 3 SON NEGRAS. SI LAS SEIS SE EXTRAEN AL AZAR, UNA TRAS OTRA, SIN REMPLAZO LA PROBABILIDAD DE QUE LAS 3 BLANCAS SALGAN EN FORMA CONSECUTIVA (EVENTO F) ES:



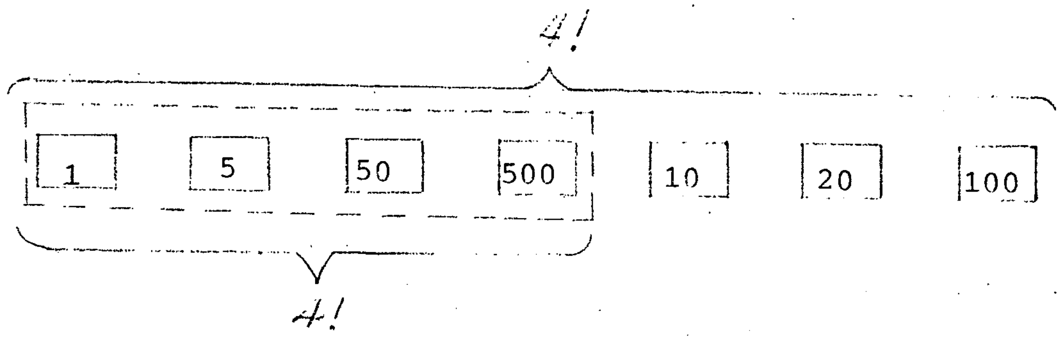
$$N(F) = 3! \times 4! \quad ; \quad N = 6!$$

$$P(F) = N(F)/N = \frac{3! \times 4!}{6!} = \frac{3! \times 4!}{6 \times 5 \times 4!}$$

$$P(F) = \frac{3 \times 2 \times 1}{6 \times 5} = \frac{1}{5}$$

EJEMPLO

EN UNA URNA SE TIENEN 7 SOBRES IDENTICOS Y CADA UNO CONTIENE UN BILLETE DE DIFERENTE DENOMINACION (1,5,10,20,50,100 y 500 PESOS). ¿CUAL ES LA PROBABILIDAD DE 1,5,50 y 500 PESOS SALGAN CONSECUTIVAMENTE EN CUALQUIER ORDEN, SI SE SACAN LOS SIETE AL AZAR, UNO TRAS OTRO (EVENTO A)?.



$N(A) = 4!x4!;\ N=7!$

$$P(A) = \frac{4!x4!}{7!} = \frac{4!x4!}{7x6x5x4!} = \frac{4x3x2x1}{7x6x5} = \frac{4}{35}$$

PERMUTACIONES PARCIALES

EL NUMERO DE MANERAS EN QUE SE PUEDEN ORDENAR N OBJETOS TOMANDO DE r EN r ES:

$$_{N}P_{r} = \frac{N!}{(N-r)!}$$

ESTO ES EQUIVALENTE A DECIR QUE $_{N}P_{r}$ ES EL NUMERO DE DIFERENTES MANERAS EN QUE r OBJETOS PUEDEN SER SELECCIONADOS DE N OBJETOS ($r \leq N$) *SIN REEMPLAZAR* NINGUNO DE ELLOS AL LOTE ANTES DE SACAR EL SIGUIENTE.

OBSERVESE QUE SI r=N:

$$_{N}P_{N} = \frac{N!}{(N-N)!} = \frac{N!}{0!} = N!$$

EJEMPLO

SI SE TIENEN LAS LETRAS A,B,C,D, EL NUMERO DE MANERAS EN QUE SE PUEDEN ORDENAR TOMANDO DE 2 EN 2 ES

$$_{4}P_{2} = \frac{4!}{(4-2)!} = \frac{4x3x2!}{2!} = 12$$

EL CONJUNTO DE ESTAS POSIBILIDADES ES:

S ={AB,AC,AD,BA,BC,BD,CA,CB,CD,DA,DB,DC}

OBSERVESE QUE CUANDO EL ORDEN ES IMPORTANTE, AC NO ES LO MISMO QUE CA, ETC.

COMBINACIONES

CUANDO EL ORDEN NO ES IMPORTANTE, ES DECIR, SI EL AGRUPAMIENTO GA ES EL MISMO QUE EL AG, A LOS AGRUPAMIENTOS SE LES DENOMINA *COMBINACIONES.* POR EJEMPLO, SI SE FORMARA UNA COMISION DE 2 INDIVIDUOS DE UN GRUPO DE 8 TOMANDO SUS NOMBRES AL AZAR DE UNA URNA, Y DESEAMOS SABER CUANTOS COMITES DE 2 MIEMBROS PODRIAN FORMARSE COMO RESULTADO DEL PROCESO, ENTONCES LOS RESULTADOS (PEDRO, JOSE) Y (JOSE, PEDRO) CONSTITUIRIAN EL MISMO COMITE, ES DECIR, NO IMPORTARIA EN QUE ORDEN SE SACARAN SUS NOMBRES DE LA URNA.

ASI, SE PUEDE DEMOSTRAR QUE EL NUMERO DE COMBINACIONES POSIBLES DE FORMAR DE N OBJETOS TOMANDO DE r EN r ES:

$$_N C_r = \binom{N}{r} = \frac{N!}{(N-r)!r!}$$

ESTO EQUIVALE A DECIR QUE $_N C_r$ ES EL NUMERO DE MANERAS DISTINTAS EN QUE r OBJETOS PUEDEN SELECCIONARSE DE N $(r \leq N)$ *SIN REEMPLAZO Y SIN IMPORTAR EL ORDEN* EN QUE APAREZCAN.

CUANDO UNO SE ENFRENTA A UN PROBLEMA QUE IMPLICA LA REPETICION DE UN EXPERIMENTO, ES NECESARIO DETERMINAR SI HAY O NO REEMPLAZO DE LAS OBSERBACIONES. POR EJEMPLO, EL REPETIR EL LANZAMIENTO DE UN DADO Y OBSERVAR CADA VEZ EL NUMERO QUE QUEDA HACIA ARRIBA LLEVA IMPLICITO QUE HAY REEMPLAZO.

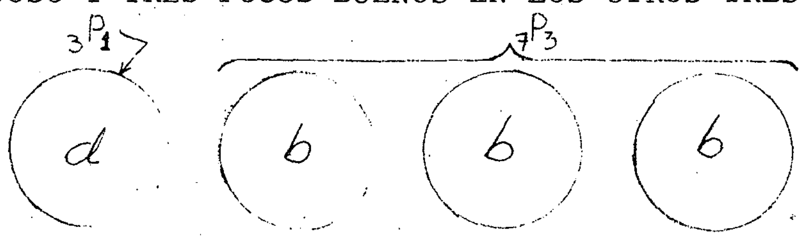
EJEMPLO

UNA CAJA CONTIENE 10 FOCOS, DE LOS CUALES 3 SON DEFECTUOSOS. SI SELECCIONAMOS 4 AL AZAR SIN REEMPLAZO

A) ¿CUANTOS SON LOS RESULTADOS POSIBLES, ES DECIR, CUANTOS ELEMENTOS TIENE EL ESPACIO DE EVENTOS?

$$N(S) = {}_{10}P_4 = \frac{10!}{(10-4)!} = 10x9x8x7 = 5040$$

B) ¿CUANTOS ELEMENTOS DE S TIENEN COMO PRIMER RESULTADO UN FOCO DEFECTUOSO Y TRES FOCOS BUENOS EN LOS OTROS TRES?

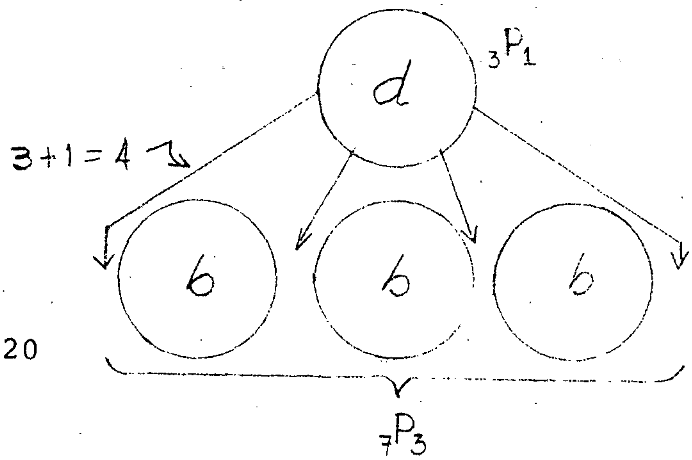


$$N(A) = {}_3P_1 \times {}_7P_3 = \frac{3!}{(3-1)!} \times \frac{7!}{(7-3)!} = \frac{3!}{2!} \times \frac{7!}{4!}$$

$$N(A) = 3 \times (7x6x5) = 630$$

$$\Rightarrow P(A) = 630/5040 = 63/504$$

C) ¿CUANTOS ELEMENTOS DE S TIENEN UN FOCO DEFECTUOSO Y 3 BUENOS?



$$N(B) = 4 \times {}_3P_1 \times {}_7P_3 = 4x630 = 2520$$

$$\Rightarrow P(B) = \frac{2520}{5040} = \frac{1}{2}$$

PERMUTACIONES DE GRUPOS DE OBJETOS

"EL NUMERO DE PERMUTACIONES POSIBLES DE N OBJETOS DE LOS CUALES SE TIENEN N_1, IGUALES ENTRE SI EN EL PRIMER GRUPO, N_2 IGUALES ENTRE SI EN EL SEGUNDO GRUPO, ETC, HASTA N_K IGUALES EN EL K-ESIMO GRUPO (LOS GRUPOS SON DISTINGUIBLES ENTRE SI), DE MANERA QUE $N_1 + N_2 + \ldots + N_K = N$ QUEDA DADO POR LA FORMULA:

$$N\ P_{N_1, N_2, \ldots, N_K} = \frac{N!}{N_1!N_2!\ldots N_K!}$$

EJEMPLO

EN EL INCISO C DEL EJEMPLO ANTERIOR SE TIENEN DOS GRUPOS EN LA MUESTRA, EL DE DEFECTUOSOS CON UN SOLO ELEMENTO, ES DECIR $N_1=1$, Y EL DE BUENOS, CON TRES ELEMENTOS, $N_2=3$, QUE SE PERMUTAN POR GRUPO DE $4^{P}1,3 = \frac{4!}{1!3!} = 4$ MANERAS DISTINTAS.

EJEMPLO

ENUMERE LAS PERMUTACIONES QUE SE PUEDEN HACER CON DOS GRUPOS DE BOLAS, 2 NEGRAS Y 2 BLANCAS.

b b n n
b n b n
b n n b
n b b n
n n b b
n b n b

$n_1 = 2$

$n_2 = 2$

$$4^{P}2,2 = \frac{4!}{2!\ 2!} = 6$$

EJEMPLO

UNA CAJA CONTIENE 25 TRANSISTORES DE LAS CUALES 3 SON EFECTUOSOS. ¿CUAL ES LA PROBABILIDAD DE QUE, SI SE EXTRAEN 5 AL AZAR SIN REMPLAZO,

A) SE OBTENGAN LOS 3 DEFECTUOSOS

B) SE OBTENGAN SOLO 2 DEFECTUOSOS

C) SE OBTENGA SOLO 1 DEFECTUOSO

D) NO SE OBTENGA NINGUNO DEFECTUOSO?

SOLUCION:

A) $N(S) = {}_{25}P_5 = \frac{25!}{(25-5)!} = \frac{25!}{20!}$

GRUPO DE 3, $N_1=3$ GRUPO DE 2, $N_2=2$

$\boxed{\underbrace{D \quad D \quad D}_{} \quad {}_3P_3}$ $\boxed{\underbrace{B \quad B}_{} \quad {}_{22}P_2}$

D=DEFECTUOSO

B=BUENO

${}_5P_{3,2}$

$N(A) = {}_3P_3 \cdot {}_{22}P_2 \cdot {}_5P_{3,2} = 3! \ \frac{22!}{(22-2)!} \ \frac{5!}{3! \ 2!}$

$N(A) = 60 \ \frac{22!}{20!}$

$$P(A) = \frac{60\frac{22!}{20!}}{\frac{25!}{20!}} = 60 \ \frac{22!}{25!} = \frac{60}{13800}$$

B) $\boxed{N_1=2 \quad \underbrace{D \quad D}_{} \quad {}_3P_2}$ $\boxed{N_2=3 \quad \underbrace{B \quad B \quad B}_{} \quad {}_{22}P_3}$

${}_5P_{2,3}$

$$N(B) = 3^{P}2 \times 22^{P}3 \times 5^{P}2,3 = \frac{3!}{(3-2)!} \quad \frac{22!}{(22-3)!} \quad \frac{5!}{2!x3!}$$

$$N(B) = 3! \frac{22!}{19!} \quad \frac{5x4}{2} = 60 \frac{22!}{19!}$$

$$P(B) = \frac{60\frac{22!}{19!}}{\frac{25!}{20!}} = 60 \frac{22!}{25!} \quad \frac{20x19!}{19!} = \frac{1200}{13800}$$

C)

$$\boxed{\begin{array}{c} N_1=1 \\ \underbrace{D} \\ 3 \end{array}} \quad \boxed{\begin{array}{c} N_2=4 \\ \underbrace{B \quad B \quad B \quad B} \\ 22^{P}4 \end{array}}$$

$$5^{P}1,4$$

$$N(C) = 3 \times 22^{P}4 \times 5^{P}1,4$$

$$N(C) = 3 \frac{22!}{(22-4)!} \times \frac{5!}{1!x4!}$$

$$N(C) = 3 \frac{22!}{18!} \quad \frac{5x4!}{1x4!} = 15 \frac{22!}{18!}$$

$$P(C) = \frac{15 \frac{22!}{18!}}{\frac{25!}{20!}} = \frac{15x20x19x18!}{\frac{25!}{22!} \times 18!} = \frac{5700}{13800}$$

D) $\underbrace{B \quad B \quad B \quad B \quad B}_{22^{P}5}$

$$N(D) = 22^{P}5 = \frac{22!}{(22-5)!} = \frac{22!}{17!} .$$

$$P(D) = \frac{22!/17!}{\frac{25!}{20!}} = \frac{20x19x18x17!}{\frac{25!}{20!} \times 17!}$$

$P(D)=6840/13800$

OBSERVESE QUE EN ESTE EJEMPLO HEMOS CALCULADO LAS PROBABILIDADES DE TODOS LOS ELEMENTOS DEL ESPACIO DE EVENTOS CORRESPONDIENTE AL "NUMERO DE DEFECTUOSOS QUE SE PUEDEN OBSERVAR EN UNA SELECCION AL AZAR DE 5 ELEMENTOS", EN LA CUAL SOLO SE PUEDEN TENER 0,1,2, Ó 3 DEFECTUOSOS, ES DECIR,

$$S = \{0,1,2,3\}$$

VERIFIQUEMOS QUE, EN EFECTO, P(S) = 1:

$$P(S) = P(\{0\}) + P(\{1\}) + P(\{2\}) + P(\{3\})$$

$$= \frac{6840}{13800} + \frac{5700}{13800} + \frac{1200}{13800} + \frac{60}{13800}$$

$$= \frac{13800}{13800} = 1$$

PROBABILIDAD CONDICIONAL

LA PROBABILIDAD CONDICIONAL, $P(A|B)$, DEL EVENTO A, DADO QUE EL B HA OCURRIDO SE CALCULA CON LA FORMULA

$$P(A|B) = \frac{P(A\cap B)}{P(B)} \quad ; \; P(B)>0 \qquad (1)$$

EVENTOS INDEPENDIENTES

SI DOS EVENTOS, A Y B, SON INDEPENDIENTES, LA PROBABILIDAD DE A NO SE ALTERA SI OCURRE EL EVENTO B; ES DECIR, DOS EVENTOS SON *INDEPENDIENTES* SI

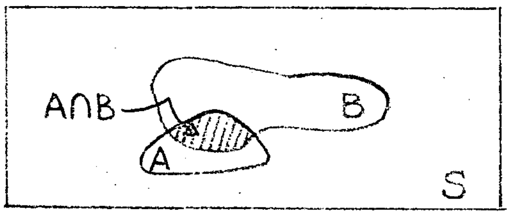
$$P(A|B) = P(A)$$

EN TAL CASO, DE LA ECUACION 1 :

$$P(A\cap B) = P(A) \times P(B) \qquad (1')$$

PUESTO QUE $P(A\cap B) = N(A\cap B)/N(S)$ Y $P(B) = N(B)/N(S)$ LA ECUACION 1 SE PUEDE ESCRIBIR COMO

$$P(A|B) = \frac{\frac{N(A\cap B)}{N(S)}}{\frac{N(B)}{N(S)}} = \frac{N(A\cap B)}{N(B)} \qquad (2)$$



EL TRABAJAR CON LA ECUACION 2 EQUIVALE A EMPLEAR UN ESPACIO DE EVENTOS REDUCIDO DE S A B.

EJEMPLO

EN UNA URNA HAY 10 TRANSISTORES BUENOS Y 10 DEFECTUOSOS. ¿CUAL ES LA PROBABILIDAD DE SACAR UNO BUENO Y UNO DEFECTUOSO (EN CUALQUIER ORDEN) AL REALIZAR DOS EXTRACCIONES AL AZAR, SI HAY REEMPLAZO DEL PRIMER TRANSISTOR OBSERVADO?

HAY VARIAS FORMAS DE RESOLVER ESTE PROBLEMA:

1. PUESTO QUE EL NUMERO DE DEFECTUOSOS ES IGUAL AL DE BUENOS, SE PUEDE FORMULAR EL SIGUIENTE ESPACIO DE EVENTOS, EN EL QUE TODOS LOS ELEMENTOS SON IGUALMENTE PROBABLES:

   $S = \{(d,d), (d,b), (b,b), (b,d)\}$

   EL EVENTO DE INTERES ES:

   $A = \{(d,b), (b,d)\}$

   POR LO QUE $N(S) = 4$, $N(A) = 2$

   Y $P(A) = 2/4 = 1/2$

2. HAY 10 x 10 MANERAS DISTINTAS DE QUE SALGA PRIMERO EL BUENO Y LUEGO EL DEFECTUOSO, Y OTRAS TANTAS DE QUE OCURRA DE MANERA INVERSA. POR LO TANTO:

   $N(A) = (10 \times 10) \times 2 = 200$

   $N(S) = 20 \times 20 = 400$

   $P(A) = 200/400 = 1/2$

3. SEAN LOS EVENTOS

   B = {SALE PRIMERO EL BUENO Y LUEGO EL DEFECTUOSO} $= \{(b,d)\}$

   F = {SALE PRIMERO EL DEFECTUOSO Y LUEGO EL BUENO} $= \{(d,b)\}$

   D = {SALE PRIMERO EL BUENO}

   E = {SALE SEGUNDO EL DEFECTUOSO}

   O = {SALE PRIMERO EL DEFECTUOSO}

   R = {SALE SEGUNDO EL BUENO}

POR LO TANTO, AL REALIZAR LAS DOS EXTRACCIONES CONSECUTIVAMENTE:

$B = D \cap E$ Y $F = O \cap R$

SI $A = \{$SALE UNO BUENO Y UNO MALO$\} = B \cup F$

SE TIENE QUE $P(A) = P(B)+P(F)$

YA QUE B Y F SON MUTUAMENTE EXCLUSIVOS, Y

$$P(B) = P(D \cap E) = \frac{10}{20} \times \frac{10}{20} = \frac{100}{400} = \frac{1}{4}$$

$$P(F) = P(O \cap R) = \frac{10}{20} \times \frac{10}{20} = \frac{1}{4}$$

YA QUE D Y E, Y O Y R SON INDEPENDIENTES. ESTO CONDUCE A

$$P(A) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

RESOLVAMOS AHORA ESTE PROBLEMA SI NO HAY REEMPLAZO:

$P(D \cap E) = P(E|D)P(D) = \frac{10}{19} \times \frac{10}{20} = \frac{10}{38}$, YA QUE

$P(D) = 10/20$, $P(E|D) = 10/19$ ANALOGAMENTE, $P(F) = \frac{10}{38}$,

POR LO QUE $P(A) = \frac{10}{38} + \frac{10}{38} = \frac{10}{19}$.

INDEPENDENCIA DE UN GRUPO DE EVENTOS

EN GENERAL, LOS EVENTOS $A_1, A_2, \ldots, A_M$

SON INDEPENDIENTES SI, Y SOLO SI,

$$P(A_{K_1} \cap A_{K_2} \cap \ldots \cap A_{K_R}) = P(A_{K_1}) \times P(A_{K_2}) \times \ldots \times P(A_{K_R})$$

PARA *CUALQUIER* GRUPO DE ENTEROS $K_1, K_2, \ldots, K_R$, CON $K_R \leq M$ (TODAS LAS PAREJAS, TERCIAS, ETC, DE EVENTOS POSIBLES DE FORMARSE DEBEN SER INDEPENDIENTES).

DICHO DE OTRA MANERA, LOS EVENTOS $A_1, A_2, \ldots, A_M$ SON INDEPENDIENTES SI LOS ELEMENTOS DE TODOS LOS SUBCONJUNTOS POSIBLES DE $R=\{A_1, A_2, \ldots A_M\}$ SON INDEPENDIENTES.

EJEMPLO

SI $M=3$, A_1, A_2 y A_3 SON INDEPENDIENTES SI, Y SOLO SI,

$$\left.\begin{aligned} P(A_1 \cap A_2) &= P(A_1)P(A_2) \\ P(A_1 \cap A_3) &= P(A_1)P(A_3) \\ P(A_2 \cap A_3) &= P(A_2)P(A_3) \\ P(A_1 \cap A_2 \cap A_3) &= P(A_1)P(A_2)P(A_3) \end{aligned}\right\}$$

TODAS LAS COMBINACIONES QUE PUEDAN FORMARSE CON DOS EVENTOS:

$\{A_1, A_2\}$, $\{A_1, A_3\}$, $\{A_2, A_3\}$

SI $M=4$, PARA QUE A_1, A_2, A_3 y A_4 SEAN INDEPENDIENTES SE REQUIERE QUE SE CUMPLA QUE

$$P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1)P(A_2)P(A_3)P(A_4)$$

$$\left.\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1)P(A_2)P(A_3) \\ P(A_1 \cap A_2 \cap A_4) &= P(A_1)P(A_2)P(A_4) \\ P(A_2 \cap A_3 \cap A_4) &= P(A_2)P(A_3)P(A_4) \\ P(A_1 \cap A_3 \cap A_4) &= P(A_1)P(A_3)P(A_4) \end{aligned}\right\}$$

TODAS LAS COMBINACIONES DE TRES EVENTOS QUE PUEDAN FORMARSE $= {}_4C_3 = \frac{4!}{3!\ 1!} = 4$

$$\left.\begin{aligned} P(A_1 \cap A_2) &= P(A_1)P(A_2) \\ P(A_1 \cap A_3) &= P(A_1)P(A_3) \\ P(A_1 \cap A_4) &= P(A_1)P(A_4) \\ P(A_2 \cap A_3) &= P(A_2)P(A_3) \\ P(A_2 \cap A_4) &= P(A_2)P(A_4) \\ P(A_3 \cap A_4) &= P(A_3)P(A_4) \end{aligned}\right\}$$

TODAS LAS COMBINACIONES DE DOS EVENTOS QUE PUEDAN FORMARSE $= {}_4C_2 = \frac{4!}{2!\ 2!} = 6$

EJEMPLO

EN UN ESTUDIO SOCIOLOGICO SE INTERROGARON 1200 PERSONAS DE UNA COLONIA RESIDENCIAL, Y SE OBTUVIERON LOS SIGUIENTES DATOS:

GUSTO POR LA MUSICA CLASICA	TITULO UNIVERSITARIO		SIN TITULO UNIVERSITARIO		
	VARONES	DAMAS	VARONES	DAMAS	Σ
ALTO	100	50	200	250	600
BAJO	150	100	150	200	600
Σ	250	150	350	450	1200

SI $A = \{VARON\}$, $B = \{CON\ TITULO\}$

$C = \{GUSTO\ ALTO\}$

¿CUAL ES LA PROBABILIDAD DE QUE SI SE SELECCIONA UN CIUDADANO AL AZAR DE LA MISMA COLONIA, ESTE SEA VARON, TENGA TITULO Y GUSTO ALTO POR LA MUSICA?

POR EL METODO FRECUENCIAL:

NUMERO DE VARONES = 250 + 350 = 600

NUMERO DE PERSONAS CON TITULO = 250 + 150 = 400

NUMERO DE PERSONAS CON ALTO GUSTO POR LA MUSICA CLASICA = 600

POR LO TANTO

$P(A) = 600/1200 = 1/2$, $P(B) = 400/1200 = \frac{1}{3}$

Y $P(C) = 600/1200 = 1/2$. PUESTO QUE

$D = A \cap B \cap C$ Y A, B Y C SON INDEPENDIENTES, SE TIENE QUE

$$P(D) = \frac{1}{2} \times \frac{1}{3} \times \frac{1}{2} = \frac{1}{12}$$

DE OTRA MANERA: $P(D) = 100/1200 = 1/12$

LEY GENERAL DE MULTIPLICACION

DE LA ECUACION (1):

$$P(A \cap B) = P(A|B)P(B)$$

ESTA ECUACION SE PUEDE GENERALIZAR A MAS DE DOS EVENTOS ASI:

$$P(E_1 \cap E_2 \cap \ldots \cap E_k) = P(E_1)P(E_2|E_1)\ldots P(E_k|E_1, E_2, \ldots, E_{k-1}) \quad (3)$$

POR EJEMPLO, SI K=4

$$P(E_1 \cap E_2 \cap E_3 \cap E_4) = P(E_1)P(E_2|E_1)P(E_3|E_1, E_2) \times$$
$$\times P(E_4|E_1, E_2, E_3)$$

<u>EJEMPLO</u>

¿CUAL ES LA PROBABILIDAD DE QUE AL EXTRAER SIN REEMPLAZO CUATRO CARTAS AL AZAR DE UN PAQUETE DE 52, LAS DOS PRIMERAS SEAN DIAMANTES Y LAS DOS ULTIMAS SEAN CORAZONES (EVENTO E)?

SEAN $A = \{$LA 1a. ES DIAMANTE$\}$, $B = \{$LA 2a. ES DIAMANTE$\}$,
$C = \{$LA 3a. ES CORAZON$\}$, $D = \{$LA 4a. ES CORAZON$\}$.

EN TAL CASO

$$E = A \cap B \cap C \cap D = \{(d, d, c, c)\}$$

$$P(A) = 13/52, \; P(B|A) = 12/51, \; P(C|A,B) = 13/50$$

$$P(D|A,B,C) = 12/49$$

APLICANDO LA ECUACION 3 SE OBTIENE

$$P(E) = \frac{13}{52}\,\frac{12}{51}\,\frac{13}{50}\,\frac{12}{49} = \frac{78}{20825}$$

SI LOS EVENTOS E_i QUE APARECEN EN LA ECUACION (3) SON INDEPENDIENTES, ENTONCES

$$P(E_1 \cap E_2 \cap \ldots \cap E_k) = P(E_1) \times P(E_2) \times \ldots \times P(E_k)$$

QUE ES LA <u>*LEY GENERAL DE MULTIPLICACION*</u>

EJEMPLO

SE TIENEN EN UNA URNA TRES BOLAS BLANCAS Y TRES NEGRAS. ¿CUAL ES LA PROBABILIDAD DE QUE APAREZCAN LAS TRES BLANCAS AL PRINCIPIO SI SE EXTRAEN SIN REEMPLAZO SUCESIVAMENTE LAS SEIS?

1. $\overbrace{\underbrace{b\ b\ b}_{3!}\ \underbrace{n\ n\ n}_{3!}}^{6!}$

CON PERMUTACIONES:

N(A) = 3! x 3!

N(S) = 6!

P(A) = 3! 3! / 6! = 1/20

2. $\boxed{b\ b\ b}\ n\ n\ n$

CON PERMUTACIONES POR GRUPOS:

N(A) = 1

$N(S) = {}_6P_{3,3} = \frac{6!}{3!3!} = 20$

P(A) = 1/20

3. $b\ b\ b\ \underbrace{n\ n\ n}$

PROBABILIDADES: $\frac{1}{2}\ \frac{2}{5}\ \frac{1}{4}\quad 1$

CON PROBABILIDADES CONDICIONALES:

$$P(A) = \frac{1}{2} \times \frac{2}{5} \times \frac{1}{4} = \frac{1}{20}$$

AHORA, ¿CUAL ES LA PROBABILIDAD DE QUE LAS TRES BLANCAS APAREZCAN CONSECUTIVAMENTE?

1. $\overbrace{\underbrace{b\ b\ b}_{3!}\ \boxed{n\ n\ n}}^{6!}$ $4!$

CON PERMUTACIONES:

$$P(A) = \frac{3!\,4!}{6!} = \frac{1}{5}$$

2. $\boxed{b\ b\ b}\ n\ n\ n$

CON PERMUTACIONES POR GRUPOS:

$$N(A) = {}_4P_{1,3} = \frac{4!}{1!\,3!} = 4, \quad N(S) = {}_6P_{3,3} = \frac{6!}{3!\,3!} = 20$$

$$P(A) = 4/20 = 1/5$$

3.

$$\begin{array}{lcccc} & b & b & b & \underbrace{n\ n\ n} \\ & \uparrow & \uparrow & \uparrow & \uparrow \\ \text{PROBABILIDADES:} & \frac{1}{2} & \frac{2}{5} & \frac{1}{4} & 1 \end{array}$$

CON PROBABILIDADES CONDICIONALES:

$$P(A) = \left(\frac{1}{2}\times\frac{2}{5}\times\frac{1}{4}\right)\,{}_4P_{1,3} = \frac{4}{20} = \frac{1}{5}$$

EJEMPLO

DE UN LOTE DE 100 EJES DE RELOJERIA SE EXTRAEN CUATRO AL AZAR SIN REEMPLAZO, ¿CUAL ES LA PROBABILIDAD DE QUE APAREZCAN DOS DEFECTUOSOS (EVENTO A) SI EN EL LOTE HAY 20 POR CIENTO DE DEFECTUOSOS?

1.

$$\overbrace{\begin{array}{cccc} d & d & b & b \\ \uparrow & \uparrow & \uparrow & \uparrow \\ \frac{20}{100} & \frac{19}{99} & \frac{80}{98} & \frac{79}{97} \end{array}}^{{}_4P_{2,2} = 6}$$

PROBABILIDADES: $\frac{20}{100} \quad \frac{19}{99} \quad \frac{80}{98} \quad \frac{79}{97}$

CON PROBABILIDAD CONDICIONAL:

$$P(A) = \frac{20}{100}\;\frac{19}{99}\;\frac{80}{98}\;\frac{79}{97}\;6 = 0.15$$

2.

$$\overbrace{\underbrace{d \quad d}_{{}_{20}P_2}\;\underbrace{b \quad b}_{{}_{80}P_2}}^{{}_{100}P_4}$$

$${}_{20}P_2 \quad {}_{80}P_2 \rightarrow {}_4P_{2,2}$$

CON PERMUTACIONES PARCIALES Y EN GRUPOS:

$$P(A) = \frac{\frac{20!}{18!}\;\frac{80!}{78!}\;\frac{4!}{2!2!}}{\frac{100!}{96!}}$$

$$= \frac{(20x19)(80x79)(6)}{100x99x98x97} = 0.15$$

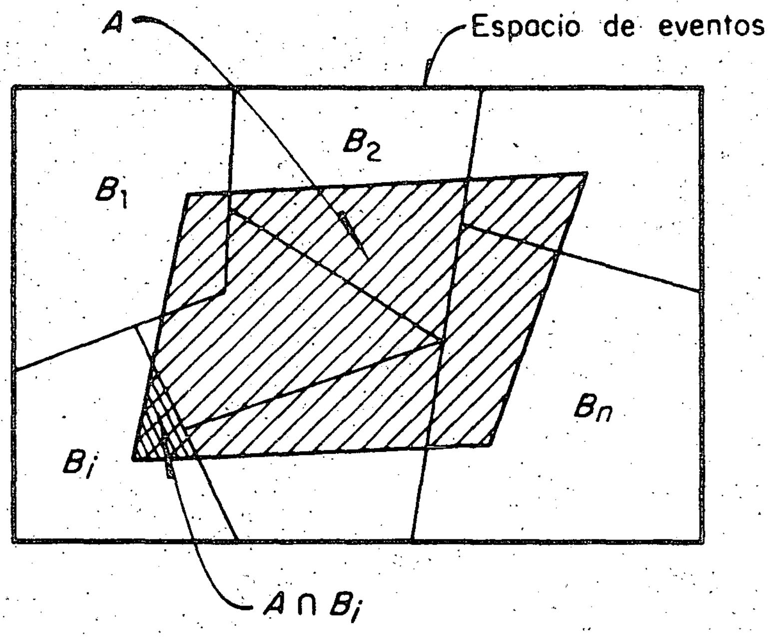
LEY GENERAL DE LA ADICION

SI TODOS LOS EVENTOS E_i SON MUTUAMENTE EXCLUSIVOS ENTRE SI, EL AXIOMA 3 TAMBIEN SE GENERALIZA A:

$$P(E_1 \cup E_2 \cup \ldots \cup E_k) = P(E_1)+P(E_2)+\ldots+P(E_k)$$

TEOREMA DE LA PROBABILIDAD TOTAL

SE DICE QUE UN GRUPO DE EVENTOS ES *COLECTIVAMENTE EXHAUSTIVO* SI LA UNION DE TODOS ELLOS ES EL ESPACIO DE EVENTOS CORRESPONDIENTE.



EN UN GRUPO DE EVENTOS COLECTIVAMENTE EXHAUSTIVOS Y MUTUAMENTE EXCLUSIVOS, $B_1, B_2, \ldots, B_n$, SI A ES UN EVENTO CUALQUIERA DEFINIDO EN EL MISMO ESPACIO, ENTONCES, APLICANDO EL AXIOMA 3, RESULTA

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

YA QUE LOS EVENTOS $A \cap B_i$ SON MUTUAMENTE EXCLUSIVOS, Y $A = (A \cap B_1) \cup (A \cap B_2)$

TOMANDO EN CUENTA QUE $P(A \cap B_i) = P(B_i)P(A|B_i)$, SE OBTIENE FINALMENTE LA ECUACION

$$P(A) = \sum_{i=1}^{i=n} P(B_i)P(A|B_i)$$

CON LA CUAL SE DEFINE EL LLAMADO *TEOREMA DE LA PROBABILIDAD TOTAL*.

TEOREMA DE BAYES

CONSIDERANDO QUE $P(B_j \cap A) = P(A \cap B_j)$, SE TIENE QUE

$$P(B_j|A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(A \cap B_j)}{P(A)}$$

DE DONDE

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^{i=n} P(B_i)P(A|B_i)}$$

ESTE RESULTADO SE CONOCE COMO *TEOREMA DE BAYES*. A LAS PROBABILIDADES $P(B_j)$ QUE SE ASIGNAN A LOS EVENTOS B_j ANTES DE OBSERVAR EL EVENTO A, SE LES DENOMINA *A PRIORI O PREVIAS*; A LAS PROBABILIDADES $P(B_j|A)$ QUE SE OBTIENEN DESPUES DE OBSERVAR EL EVENTO A, SE LES LLAMA *A POSTERIORI O POSTERIORES*.

EJEMPLO

EN UNA FABRICA SE RECIBEN REGULADORES DE VOLTAJE DE DOS PROVEEDORES, B_1 Y B_2, EN PROPORCION DE 3 A 1; ES DECIR, LA PROBABILIDAD DE QUE UN REGULADOR TOMADO AL AZAR PROVENGA DEL PROVEEDOR B_1 ES $P(B_1)=3/4$, Y DEL B_2 ES $P(B_2)=1/4$.

SUPONGAMOS ADEMAS QUE EL CONTROL DE CALIDAD DEL PROVEEDOR B_1 ES MEJOR QUE EL DE B_2, DE MANERA QUE EL 95% DE LOS REGULADORES DE B_1 TRABAJAN BIEN, Y SOLO EL 80% DE LOS DE B_2 FUNCIONAN CORRECTAMENTE. CALCULEMOS LA PROBABILIDAD DE QUE UN REGULADOR TOMADO AL AZAR FUNCIONE BIEN (EVENTO A).

$$P(A|B_1) = 0.95; \; P(A|B_2) = 0.80$$

DEL TEOREMA DE LA PROBABILIDAD TOTAL:

$$P(A) = P(A|B_1)P(B_1)+P(A|B_2)P(B_2)$$

$$= 0.95 \times \frac{3}{4} + 0.80 \times \frac{1}{4} = 0.9125$$

EJEMPLO

SUPONGAMOS AHORA QUE LA PREGUNTA DEL PROBLEMA SE CAMBIA A: ¿CUAL ES LA PROBABILIDAD DE QUE UN REGULADOR TOMADO AL AZAR PROVENGA DEL PROVEEDOR B_1, SI SE HIZO UNA PRUEBA DEL REGULADOR Y SE OBSERVO QUE FUNCIONA CORRECTAMENTE?

APLICANDO EL TEOREMA DE BAYES:

$$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1)+P(B_2)P(A|B_2)} \quad ;\ P(B_1) = \frac{3}{4}\ ;\ P(B_2) = \frac{1}{4}$$

$$= \frac{\frac{3}{4} \times 0.95}{\frac{3}{4} \times 0.95 + \frac{1}{4} \times 0.80} = \frac{2.85}{3.65} = 0.78$$

ADEMAS

$$P(B_2|A) = \frac{P(B_2)P(A|B_2)}{\frac{3.65}{4}} = \frac{\frac{1}{4} \times 0.80}{\frac{3.65}{4}} = 0.22$$

OBSERVESE QUE

$$P(B_1|A) + P(B_2|A) = 0.78 + 0.22 = 1.00$$

EJEMPLO

SUPONGASE QUE UNA PRUEBA PARA DETECTAR DIABETES TIENE UNA EFICIENCIA DEL 95%, ES DECIR, SOLO EN EL 95% DE LOS CASOS SE DETECTA CON ELLA LA DIABETES EN UNA PERSONA QUE LA PADECE. SUPONGASE TAMBIEN QUE EL 2% DE LAS PRUEBAS QUE RESULTAN POSITIVAS SON DE GENTE SANA, Y QUE EL 3% DE LA POBLACION DE UNA REGION DE MEXICO PADECE ESTA ENFERMEDAD.

a) ¿CUAL ES LA PROBABILIDAD DE QUE UNA PERSONA SELECCIONADA AL AZAR PUEDA SER DECLARADA DIABETICA POR LA PRUEBA?

b) SI LA PRUEBA DICE QUE SI ES DIABETICA, ¿CUAL ES LA PROBABILIDAD DE QUE REALMENTE LO SEA?

SOLUCION

$B_1 = \{\text{TIENE DIABETES}\}$; $B_2 = \{\text{NO TIENE DIABETES}\}$

$E = \{\text{LA PRUEBA DETECTA DIABETES}\}$

$P(B_1) = 0.03$, $P(B_2) = 0.97$

$P(E|B_1) = 0.95$, $P(E|B_2) = 0.02$

a) $$P(E) = P(E|B_1)P(B_1) + P(E|B_2)P(B_2)$$
$$= 0.95 \times 0.03 + 0.02 \times 0.97 = 0.0479$$

b) $$P(B_1|E) = \frac{P(B_1)P(E|B_1)}{P(B_1)P(E|B_1)+P(B_2)P(E|B_2)}$$
$$= \frac{0.03 \times 0.95}{0.03 \times 0.95 + 0.97 \times 0.02} = 0.59$$

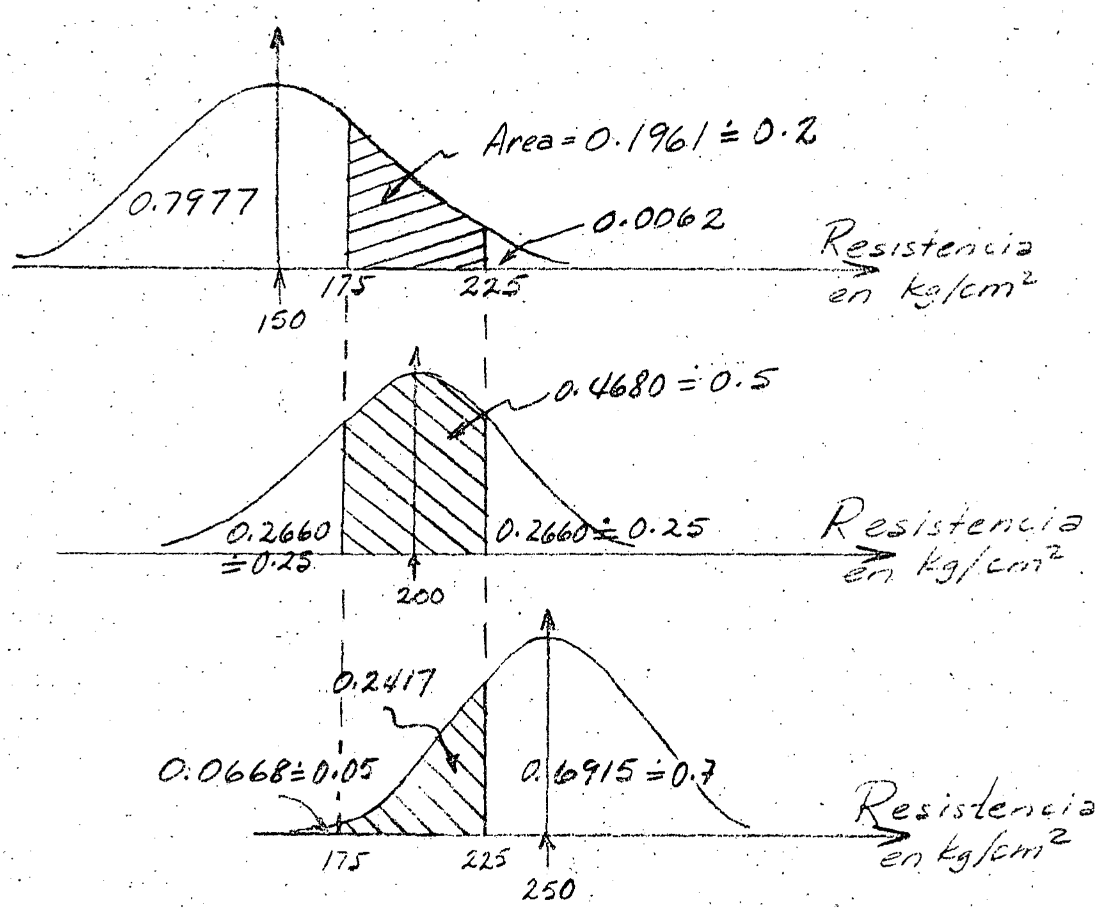
EJEMPLO

EXISTE UN EDIFICIO DE CONCRETO REFORZADO PARA EL CUAL SE INVESTIGA SU CAPACIDAD DE CARGA DE DISEÑO. UN INGENIERO, CON BASE EN SU EXPERIENCIA PERSONAL Y CON BASE EN LA EPOCA EN QUE FUE CONSTRUIDO, DECIDE QUE LA RESISTENCIA NOMINAL, f'_c DEL CONCRETO PUDO SER DE 150 KG/CM^2, 200 KG/CM^2 O 250 KG/CM^2, CON LAS SIGUIENTES PROBABILIDADES PREVIAS:

RESISTENCIA NOMINAL, EN KG/CM^2	PROBABILIDAD
$B_1 = \{150\}$	$P(B_1)=0.3$
$B_2 = \{200\}$	$P(B_2)=0.6$
$B_3 = \{250\}$	$P(B_3)=0.1$

PARA PREDECIR LA RESISTENCIA NOMINAL REAL, ES NECESARIO REALIZAR UN EXPERIMENTO QUE CONSISTE EN EXTRAER CORAZONES (MUESTRAS) DEL CONCRETO DE LA ESTRUCTURA Y PROBARLOS A COMPRESION SIMPLE. EL INGENIERO DECIDE QUE LA RESISTENCIA, S, DE UN SOLO CORAZON DARA UNA PREDICCION CONFIABLE, Y DEFINE LOS EVENTOS ANOTADOS EN LA PRIMERA COLUMNA DE LA TABLA DE LA SIGUIENTE HOJA, A LAS CUALES LES ASIGNA PROBABILIDADES CONDICIONALES, DADA LA RESISTENCIA NOMINAL, f'_c, DE ACUERDO CON LA SUPOSICION DE QUE LA RESISTENCIA TIENE DISTRIBUCION NORMAL CON LA MEDIA IGUAL A f'_c Y COEFICIENTE DE VARIACION CONSTANTE; DE ACUERDO CON LA TEXTURA DEL CONCRETO Y CON LA EPOCA DE CONSTRUCCION, LE ASIGNA UN VALOR DE 0.20 A DICHO COEFICIENTE (IMPLICA UN CONTROL DE CALIDAD MALO)

RESISTENCIA DEL CORAZON, S, EN KG/CM^2	$P(A_j \mid B_i)$		
	$(f'_c)_1=150$ $B_1=\{150\}$	$(f'_c)_2=200$ $B_2=\{200\}$	$(f'_c)_3=250$ $B_3=\{250\}$
EVENTO $A_1 = \{S<175\}$	0.80	0.25	0.05
EVENTO $A_2 = \{175 \leq S \leq 225\}$	0.20	0.50	0.25
EVENTO $A_3 = \{S > 225\}$	0	0.25	0.70



SUPONGASE QUE SE SACA UN CORAZON Y QUE SU RESISTENCIA RESULTA SER DE 164 KG/CM^2, ES DECIR, QUE OCURRE EL EVENTO A_1. LAS PROBABILIDADES A POSTERIORI DE LAS RESISTENCIAS NOMINALES SON, ENTONCES

$$P(\{150\}|A_1) = \frac{P(\{150\})P(A_1|\{150\})}{P(\{150\})P(A_1|\{150\})+P(\{200\})P(A_1|\{200\})+P(\{250\})P(A_1|\{250\})}$$

$$= \frac{0.3 \times 0.80}{0.3 \times 0.8 + 0.6 \times 0.25 + 0.1 \times 0.05} = \frac{0.24}{0.24 + 0.15 + 0.005}$$

$$= \frac{0.24}{0.395} = 0.6076$$

$$P(\{200\}|A_1) = \frac{P(\{200\})P(A_1|\{200\})}{0.395} = \frac{0.15}{0.395} = 0.3797$$

$$P(\{250\}|A_1) = \frac{P(\{250\})P(A_1|\{250\})}{0.395} = \frac{0.005}{0.395} = 0.0127$$

SUPONGASE AHORA QUE EN VEZ DE UN SOLO CORAZON EL INGENIERO HUBIESE DECIDIDO OBTENER DOS, SITUADOS EN DIFERENTES NIVELES DE LA ESTRUCTURA, Y QUE AL PROBARLOS EN UNO OCURRIO A_1 Y EN EL OTRO A_2. LA PROBABILIDAD DE QUE OCURRAN AMBOS EVENTOS (A_1, A_2) SI f'_c ES REALMENTE 150, 200 O 250 KG/CM^2, SERA EL PRODUCTO DE DOS PROBABILIDADES CONDICIONALES, PUESTO QUE A_1 y A_2 SON INDEPENDIENTES.

$$P(A_1,A_2|\{150\}) = P(A_1|\{150\})P(A_2|\{150\}) = 0.80 \times 0.20 = 0.16$$

$$P(A_1,A_2|\{200\}) = P(A_1|\{200\})P(A_2|\{200\}) = 0.25 \times 0.50 = 0.125$$

$$P(A_1,A_2|\{250\}) = P(A_1|\{250\})\ P(A_2|\{250\}) = 0.05 \times 0.25 = 0.0125$$

ESTAS PROBABILIDADES SON INDISPENSABLES, YA QUE EL TEOREMA DE BAYES EN ESTE CASO NOS DARIA:

$$P(\{150\}|A_1,A_2) = \frac{P(\{150\})P(A_1,A_2|\{150\})}{P(\{150\})P(A_1,A_2|\{150\})+P(\{200\})P(A_1,A_2|\{200\})+P(\{250\})\times P(A_1,A_2|\{250\})}$$

EN ESTE CASO LAS PROBABILIDADES A POSTERIORI SON:

$$P(\{150\}|A_1,A_2) = \frac{0.3 \times 0.16}{3.0 \times 0.16 + 0.6 \times 0.125 + 0.1 \times 0.0125}$$

$$= \frac{0.048}{0.048 + 0.075 + 0.00125} = \frac{0.048}{0.12425} = 0.386$$

$$P(\{200\}|A_1,A_2) = \frac{P(\{200\})P(A_1,A_2|\{200\})}{0.12425} = \frac{0.075}{0.12425} = 0.604$$

$$P(\{250\}|A_1,A_2) = \frac{P(\{250\})P(A_1,A_2|\{250\})}{0.12425} = \frac{0.00125}{0.12425} = 0.010$$

LOS MISMOS RESULTADOS SE HABRIAN OBTENIDO SI EL INGENIERO, DESPUES DE EXTRAER EL PRIMER CORAZON Y DE CALCULAR LAS PROBABILIDADES POSTERIORES CORRESPONDIENTES, HUBIERA DECIDIDO SACAR EL SEGUNDO Y RECALCULAR DICHAS PROBABILIDADES CON BASE EN LAS OBTENIDAS PARA EL PRIMERO; ES DECIR, LAS PROBABILIDADES PREVIAS EN EL SEGUNDO CALCULO SERIAN $P(\{150\}) = 0.6076$, $P(\{200\}) = 0.3797$ Y $P(\{250\}) = 0.0127$. EN TAL CASO, LAS PROBABILIDADES POSTERIORES, DADO QUE OCURRIO A_2, SON

$$P(\{150\}|A_2) = \frac{P(\{150\})P(A_2|\{150\})}{0.2 \times 0.6076 + 0.5 \times 0.3797 + 0.25 \times 0.0127}$$

$$= \frac{0.12152}{0.12152 + 0.18975 + 0.00318} = \frac{0.12152}{0.31445} = 0.386$$

$$P(\{200\}|A_2) = \frac{P(\{200\})P(A_2|\{200\})}{0.31445} = \frac{0.18975}{0.31445} = 0.604$$

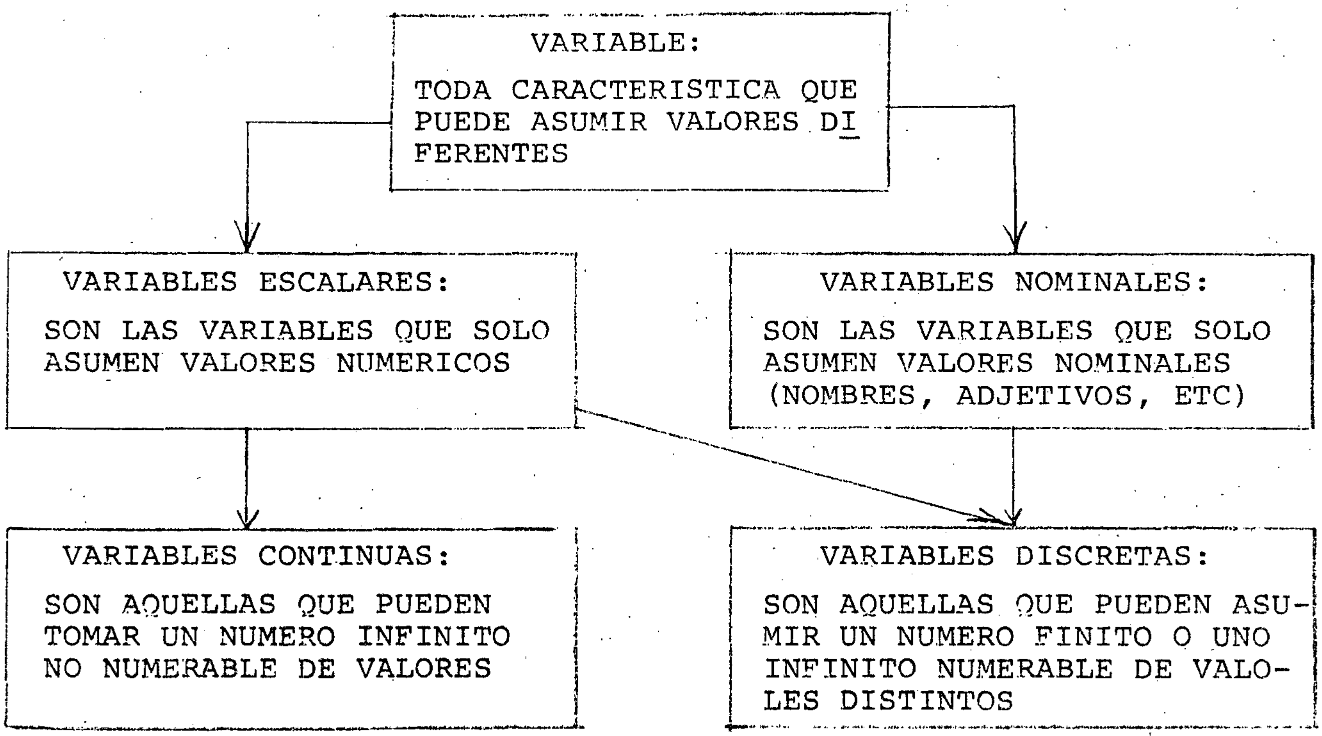
$$P(\{250\}|A_2) = \frac{P(\{250\})P(A_2|\{250\})}{0.31445} = \frac{0.00318}{0.31445} = 0.010$$

QUE SON IGUALES A LAS ANTERIORES.

CON LO ANTERIOR SE DEMUESTRA QUE LAS PROBABILIDADES SE PUEDEN ACTUALIZAR CONFORME SE VA OBTENIENDO NUEVA INFORMACION EXPERIMENTAL.

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

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

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

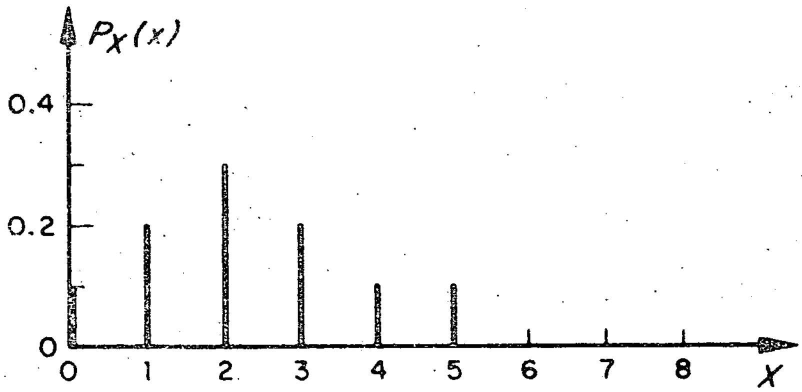
LA DISTRIBUCION DE PROBABILIDADES Y LA DE PROBABILIDADES ACUMULADAS CORRESPONDIENTES SERAN

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

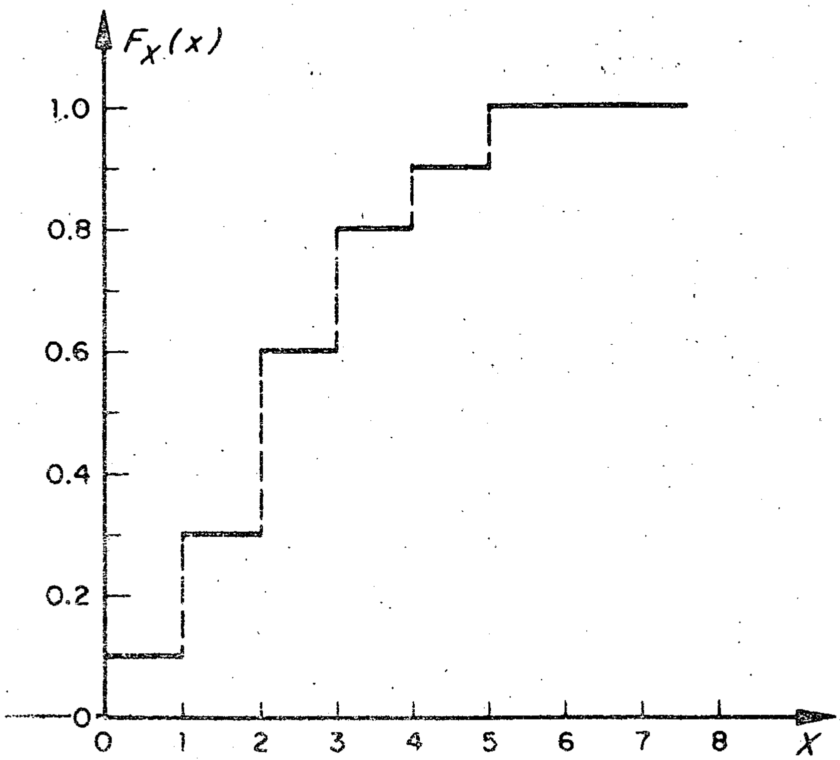
O SEA

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

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

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

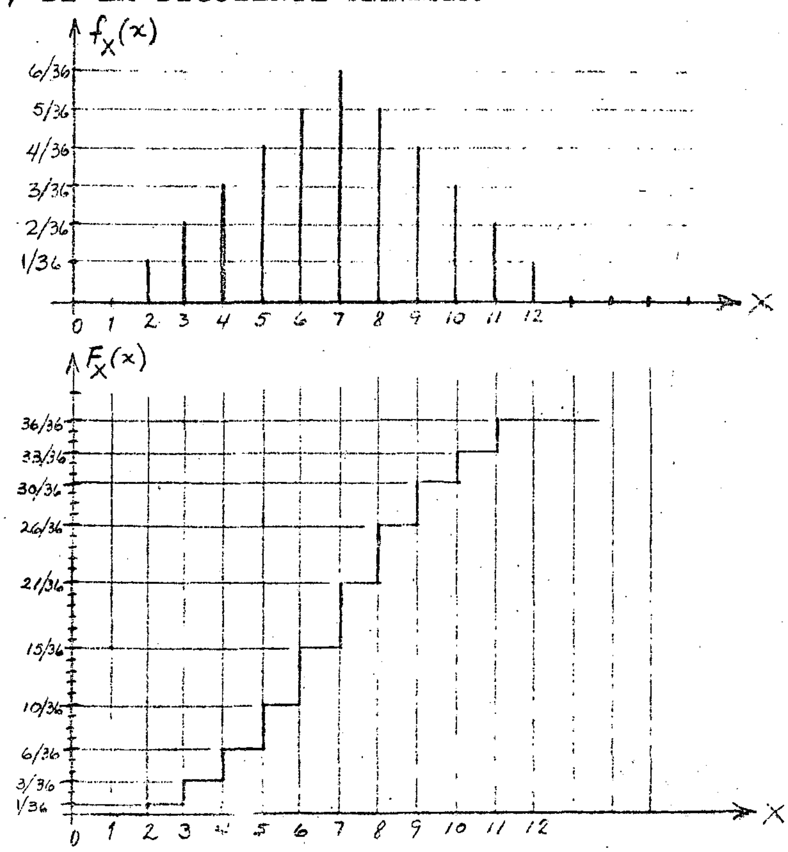
EN ESTE CASO $X_1=2,\ X_2=3,\ldots,X_{11}=12$

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

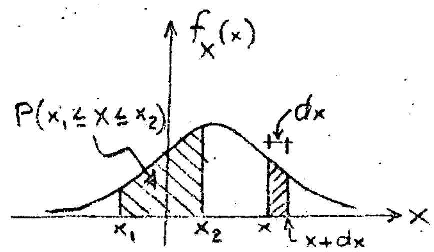
ESTAS PROBABILIDADES FUERON CALCULADAS EN UN EJEMPLO PREVIO SOBRE PROBABILIDADES DE EVENTOS .

CON ESTAS PROBABILIDADES SE PUEDE OBTENER LA FUNCION 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
	$\Sigma=1$	



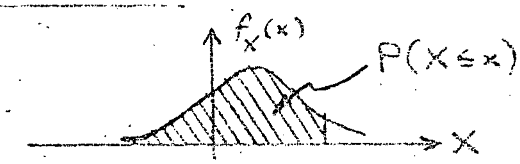
EN EL CASO DE UNA VARIABLE ALEATORIA CONTINUA, X, LA PROBABILIDAD DE QUE ESTA TOME UN VALOR COMPRENDIDO ENTRE x Y x + dx ESTA DADA POR $f_X(x)dx$, DONDE $f_X(x)$ ES LA DENSIDAD DE PROBABILIDADES DE X. POR LO TANTO, LA PROBABILIDAD DE QUE X ASUMA VALORES COMPRENDIDOS EN EL INTERVALO $x_1 \leq X \leq x_2$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

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

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

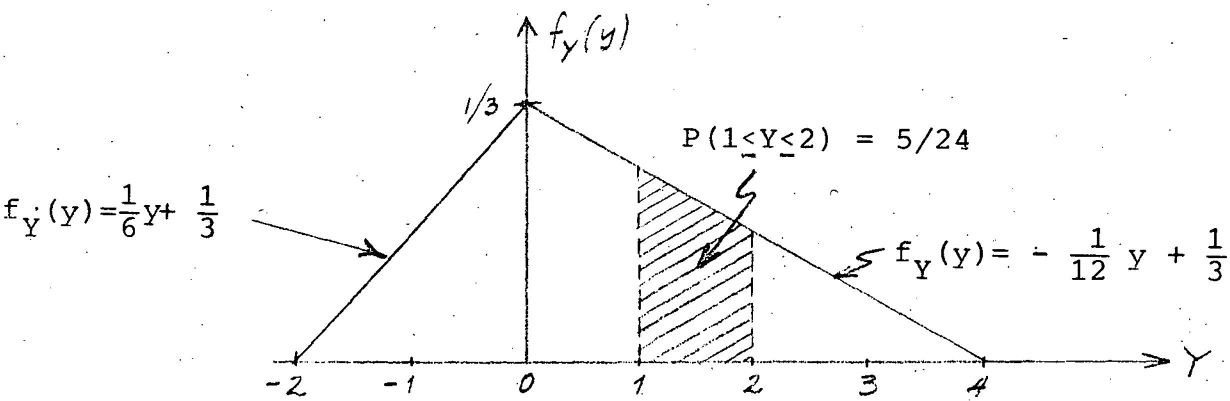
EJEMPLO

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

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

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

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



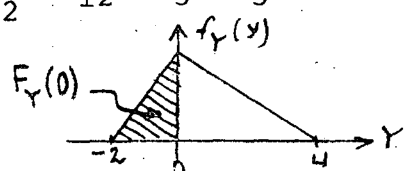
LA DISTRIBUCION DE PROBABILIDADES ACUMULADAS ES, ENTONCES:

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 $0 \le Y \le 4$

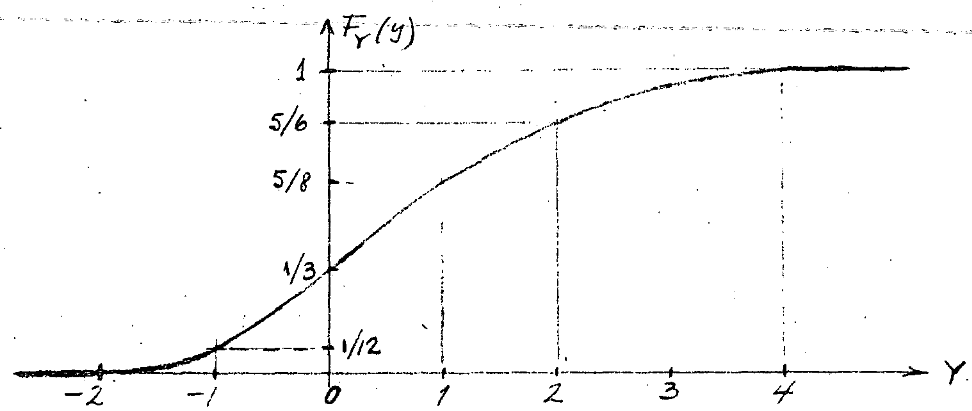


$$F_Y(y) = F_Y(0) + \int_0^Y \left(-\frac{1}{12}\,U + \frac{1}{3}\right) dU = \frac{1}{3} + \left[\frac{-U^2}{24} + \frac{U}{3}\right]_0^Y =$$

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

$F_Y(y) = 0 \qquad \text{SI} \qquad y \leq -2$

$F_Y(y) = 1 \qquad \text{SI} \qquad y \geq 4$



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

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

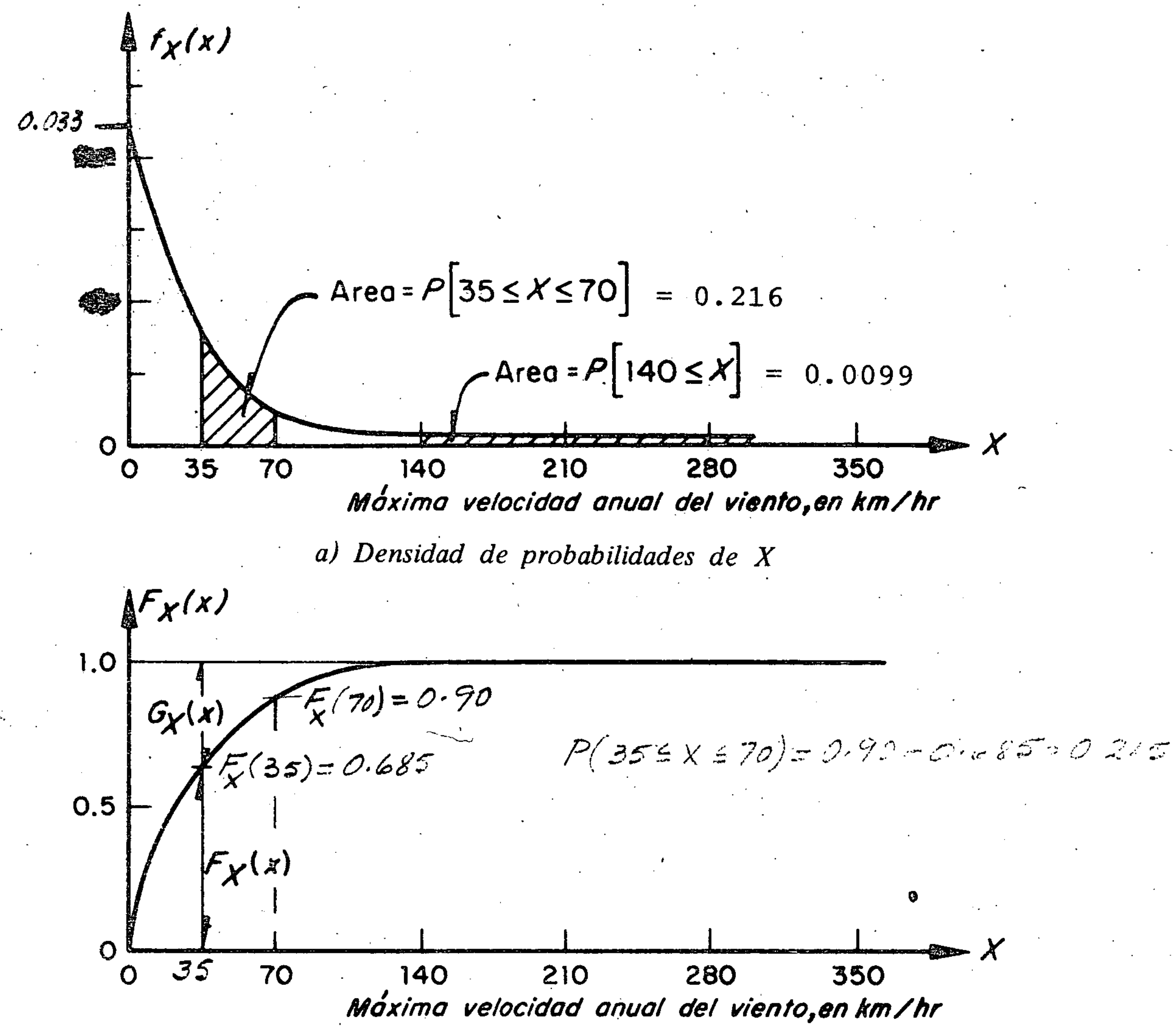
EL VALOR DE λ SE PUEDE TOMAR, POR EJEMPLO, DE MANERA QUE $F_X(x)$ SE AJUSTE PARA QUE COINCIDA CON UN VALOR EMPIRICO. ASI, SI LA FRECUENCIA RELATIVA DEL EVENTO $A = \{X \leq 70$ KM/H$\}$ ES 0.9, ENTONCES

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

DE DONDE

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

POR LO CUAL $\lambda = 0.033$.



a) Densidad de probabilidades de X

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 <u>VARIABLE ALEATORIA DISCRETA</u>, X, ES, POR DEFINICION

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

O PARA UNA <u>VARIABLE CONTINUA</u>

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

<u>EJEMPLOS</u>

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.*

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 NIVELACION, 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 CALCULOS 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 $\approx$ 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

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

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

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.

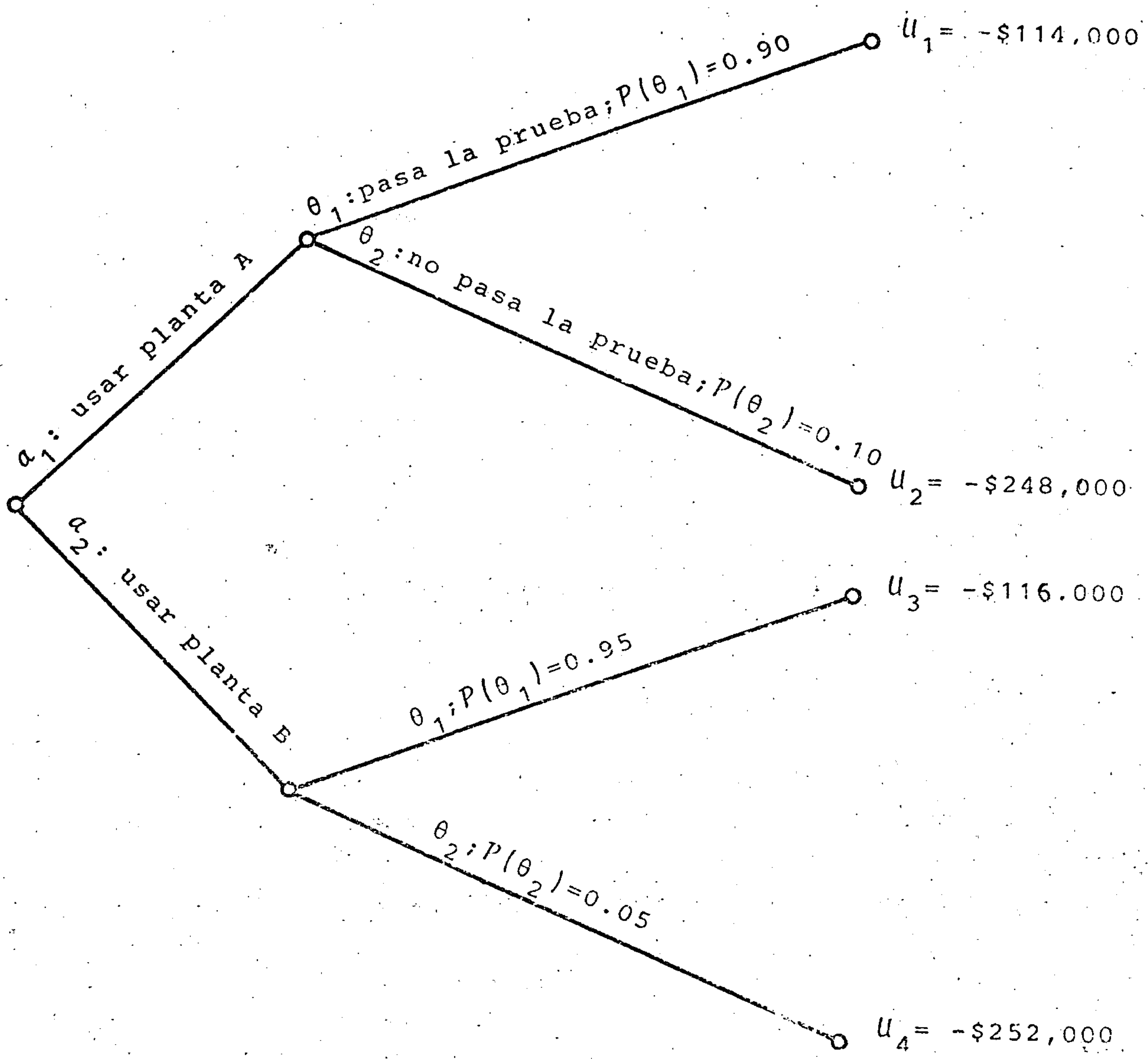


Fig 6.4 Arbol de decisiones del ejemplo 6.2

MEDIDAS DE 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:

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

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

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

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

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

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

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

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

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

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

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

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

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,

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

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

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

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

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

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

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

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

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

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

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

		p				
n	x	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	0.10	0.20	*p* 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

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

n	x	p 0.10	0.20	0.30	0.40	0.50
16	0	.1853	.0281	.0033	.0003	.0000
	1	.5147	.1407	.0261	.0033	.0003
	2	.7892	.3518	.0994	.0183	.0021
	3	.9316	.5981	.2459	.0651	.0106
	4	.9830	.7982	.4499	.1666	.0384
	5	.9967	.9183	.6598	.3288	.1051
	6	.9995	.9733	.8247	.5272	.2272
	7	.9999	.9930	.9256	.7161	.4018
	8	1.0000	.9985	.9743	.8577	.5982
	9	1.0000	.9998	.9929	.9417	.7728
	10	1.0000	1.0000	.9984	.9809	.8949
	11	1.0000	1.0000	.9997	.9951	.9616
	12	1.0000	1.0000	1.0000	.9991	.9894
	13	1.0000	1.0000	1.0000	.9999	.9979
	14	1.0000	1.0000	1.0000	1.0000	.9997
	15	1.0000	1.0000	1.0000	1.0000	1.0000
17	0	.1668	.0225	.0023	.0002	.0000
	1	.4818	.1182	.0193	.0021	.0001
	2	.7618	.3096	.0774	.0123	.0012
	3	.9174	.5489	.2019	.0464	.0064
	4	.9779	.7582	.3887	.1260	.0245
	5	.9953	.8943	.5968	.2639	.0717
	6	.9992	.9623	.7752	.4478	.1662
	7	.9999	.9891	.8954	.6405	.3145
	8	1.0000	.9974	.9597	.8011	.5000
	9	1.0000	.9995	.9873	.9081	.6855
	10	1.0000	.9999	.9968	.9652	.8338
	11	1.0000	1.0000	.9993	.9894	.9283
	12	1.0000	1.0000	.9999	.9975	.9755
	13	1.0000	1.0000	1.0000	.9995	.9936
	14	1.0000	1.0000	1.0000	.9999	.9988
	15	1.0000	1.0000	1.0000	1.0000	.9999
	16	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
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 ~~QUINTA~~ 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

$$\overbrace{\underbrace{\odot\odot\odot\cdots\odot\odot}_{\text{M EXITOS}}\Big|\underbrace{\odot\odot\odot\cdots\odot\odot}_{\text{(N-M) FRACASOS}}}^{\text{N ELEMENTOS}}$$

EN LA MUESTRA

$$\overbrace{\underbrace{\odot\odot\odot\cdot\cdot\odot\odot}_{\text{x EXITOS}}\Big|\underbrace{\odot\odot\cdot\cdot\odot}_{\text{(n-x) FRACASOS}}}^{\text{n ELEMENTOS}}$$

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

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

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?

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

$$= \frac{\frac{40!}{39! \times 1!} \times \frac{60!}{58! \times 2!}}{\frac{100!}{97! \times 3!}} = 0.438$$

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 $n=3$

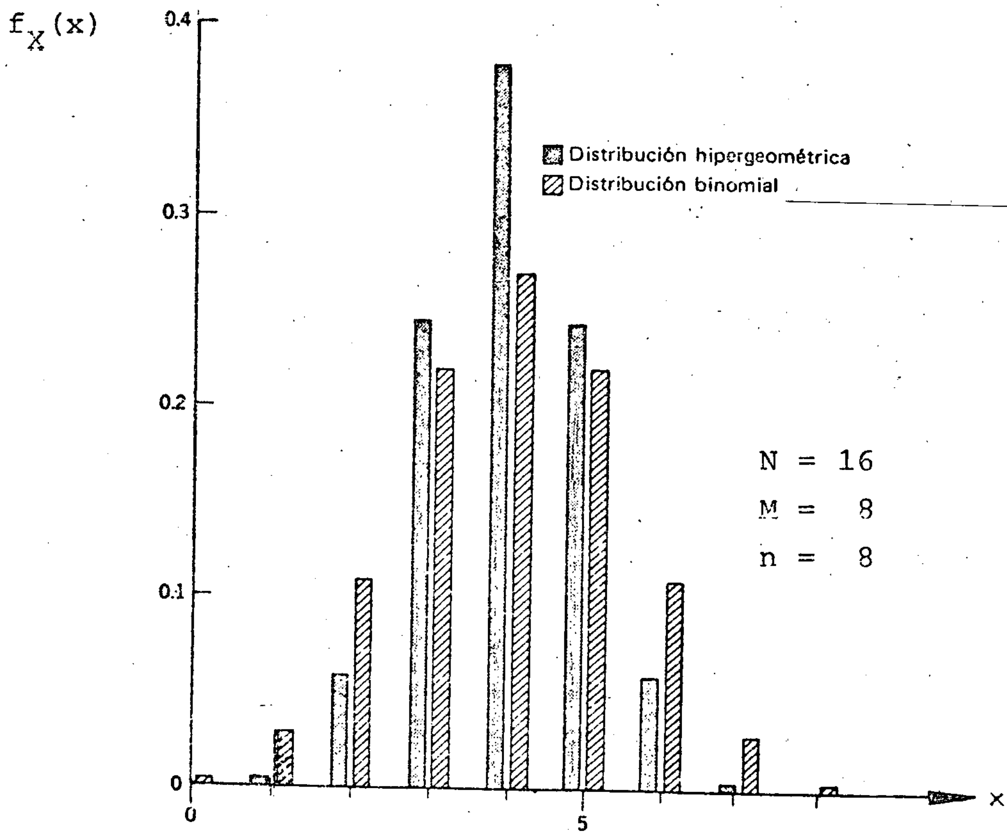
$$P[X=1] = \frac{3!}{1!\ 2!}(0.40)^1(0.60)^2 = 0.432$$

FORMULA DE STIRLING

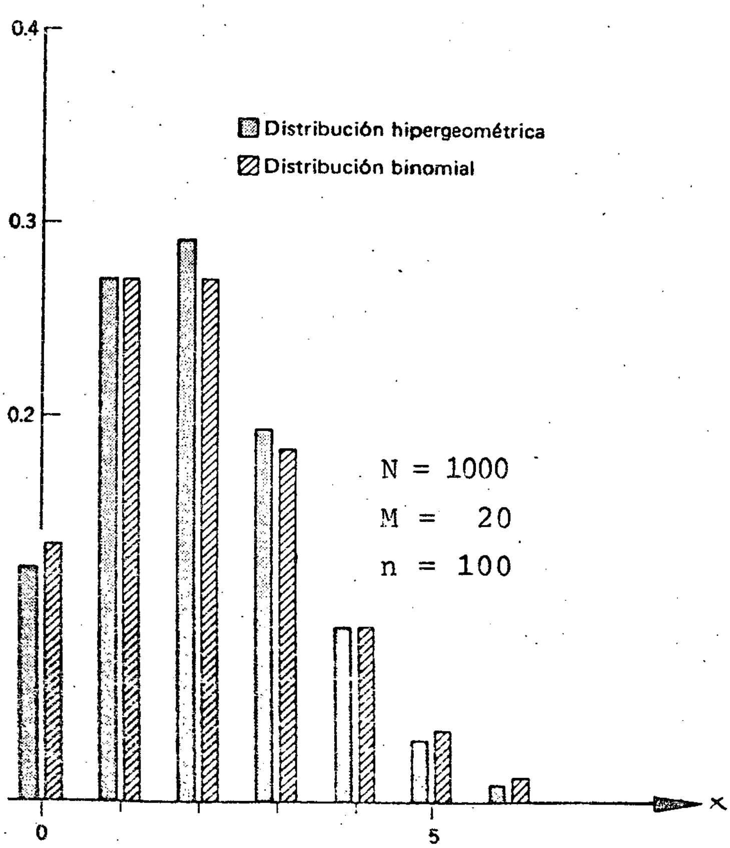
$$N! \simeq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \quad ; \quad (e = 2.718\ldots)$$

Error = 2% SI n = 4

Error = 0.8% SI n = 10



COMPARACION DE LAS DISTRIBUCIONES HIPERGEOMETRICA Y BINOMIAL



DISTRIBUCION DE POISSON

UNA DISTRIBUCION DE PROBABILIDADES PARA UNA VARIABLE ALEATORIA DISCRETA, X, DE LA FORMA

$$f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!} \; ; \; x = 0, 1, 2, \ldots$$

SE LLAMA *DISTRIBUCION DE POISSON*; EN ESTA ECUACION λ ES UNA CONSTANTE. SE PUEDE DEMOSTRAR QUE LA MEDIA Y LA VARIANCIA PARA ESTA DISTRIBUCION QUEDAN DADAS POR

$$E(X) = \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!} = \lambda$$

$$\sigma^2(X) = \sum_{x=0}^{\infty} (x-\lambda)^2 \frac{\lambda^x e^{-\lambda}}{x!} = \lambda$$

ES POSIBLE DEMOSTRAR QUE LA DISTRIBUCION DE POISSON PUEDE EMPLEARSE COMO UNA PROXIMACION DE LA DE BERNOULLI, TOMANDO $\lambda = np$ CUANDO n ES GRANDE Y p PEQUEÑA, PERO DE TAL MANERA QUE $npq > 1$. AL RESPECTO, SI $n=20$ Y $p=0.05$, ENTONCES EL ERROR QUE SE TIENE AL USAR DICHA APROXIMACION ES MENOR DE 3 POR CIENTO PARA VALORES DE X MENORES DE 3; PARA X=4 Y X=5 LOS ERRORES RESPECTIVOS SON 15 Y 41 POR CIENTO, DEBIDO A QUE NO SE CUMPLE CON LA CONDICION DE QUE npq SEA MAYOR DE UNO, YA QUE $npq = 20 \times 0.05 \times 0.95 = 0.95$.

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

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

x	λ 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	.462	.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

EJEMPLO

SI LA PROBABILIDAD DE QUE FALLE UNA VARILLA DE ACERO AL APLICARLE UNA DETERMINADA FUERZA DE TENSION ES DE 0.001, ¿CUAL ES LA PROBABILIDAD DE QUE DE 2000 VARILLAS PROBADAS FALLEN A) TRES, B) MAS DE DOS?

CON $\lambda = 2000 \times 0.001 = 2$ Y CONSIDERANDO QUE $npq = 1.9 > 1$, SE PUEDE USAR LA DISTRIBUCION DE POISSON COMO APROXIMACION DE LA BINOMIAL:

a) $$P[X = 3] = \frac{\lambda^3 \, e^{-\lambda}}{3!}$$

$$P[X = 3] = \frac{2^3 \, e^{-2}}{3!} = 0.18$$

EN ESTE CASO LA DISTRIBUCION BINOMIAL DA COMO RESULTADO

$$P[X=3] = \frac{2000!}{3! \times 1997!} (0.001)^3 (0.999)^{1997} = 0.184$$

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

$$+ P[X = 1] + P[X = 2] \Big\} = 1 - \frac{2^0 \, e^{-2}}{0!} - \frac{2^1 \, e^{-2}}{1!} - \frac{2^2 \, e^{-2}}{2!} =$$

$$= 1 - \frac{1}{e^2} - \frac{2}{e^2} - \frac{2}{e^2} = 1 - \frac{5}{e^2} = 0.323$$

EJEMPLO

UNA COMPAÑIA ASEGURADORA DESPUES DE MUCHOS AÑOS DE EXPERIENCIA HA ESTIMADO QUE EL 0.004% DE LA POBLACION FALLECE ANUALMENTE POR ACCIDENTE AUTOMOVILISTICO. SI ESTA COMPAÑIA TIENE 40,000 ASEGURADOS, ¿CUAL ES LA PROBABILIDAD DE QUE 2 DE ELLOS MUERAN EN UN AÑO POR ESE TIPO DE ACCIDENTE?

SEA X EL NUMERO DE PERSONAS QUE MUEREN ANUALMENTE DE ENTRE LOS ASEGURADOS, POR ACCIDENTE. LA MEDIA DE X ES

$$E[X] = 0.00004 \times 40{,}000 = 1.6 = \lambda$$

ADEMAS, TOMANDO EN CUENTA QUE npq>1, SE PUEDE USAR SIN GRAN ERROR LA DISTRIBUCION DE POISSON:

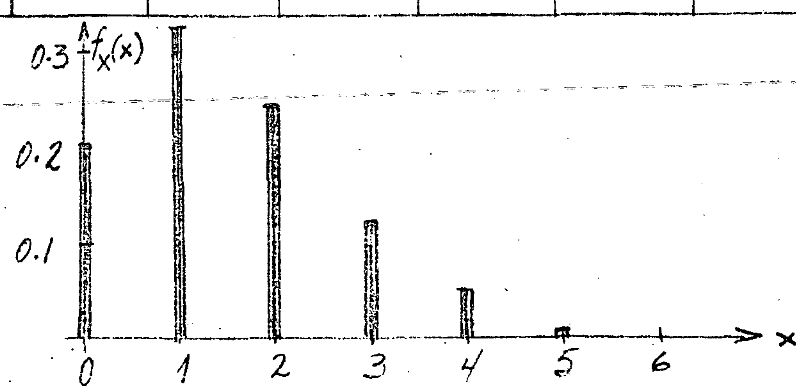
$$P[X=x] = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{(1.6)^x e^{-1.6}}{x!} \quad ; \; x=0,\ 1,\ 2,\ldots$$

POR LO QUE

$$P[X=2] = \frac{(1.6)^2 e^{-1.6}}{2!} = \frac{0.2019 \times 2.56}{2} = 0.26$$

LA DISTRIBUCION DE PROBABILIDADES PARA ESTA VARIABLE ALEATORIA ES:

x	0	1	2	3	4	5	6	...
$f_X(x)$	0.202	0.323	0.258	0.138	0.055	0.018	0.005	...



EJEMPLO

EN LA AMPLIACION DEL CARRIL PARA DAR VUELTA A LA IZQUIERDA EN UNA AVENIDA, SOLO HAY CAPACIDAD PARA 3 AUTOS COMO MAXIMO ESPERANDO LA FLECHA LUMINOSA DEL SEMAFORO. EN UN ESTUDIO ESTADISTICO DEL TRANSITO EN ESE LUGAR SE ENCONTRO QUE EN CADA CICLO DE LUCES DEL SEMAFORO HAY EN PROMEDIO 6 AUTOS QUE VAN A DAR VUELTA. ¿CUAL ES LA PROBABILIDAD DE QUE EN UN CICLO DEL SEMAFORO, TOMADO AL AZAR, SE CONGESTIONE EL TRANSITO POR EXCEDERSE LA CAPACIDAD DEL CARRIL?

$$P[X>3] = ?$$

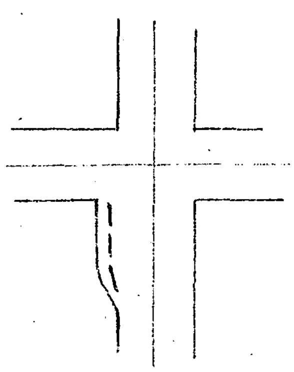
SI $A=\{X>3\}$, $\bar{A}=\{X\leq 3\}$

$P(\bar{A}) = 1-P(A)$ O $P(A)= 1-P(\bar{A})$. CON $\lambda=6$,

$$P(\bar{A}) = P[X\leq 3] = \sum_{x=0}^{x=3} f_X(x) = \sum_{x=0}^{x=3} \frac{e^{-6}6^x}{x!}$$

$$P(\bar{A}) = e^{-6}\left(1 + 6 + \frac{6^2}{2} + \frac{6^3}{6}\right) = 61e^{-6} = 0.152$$

$$P[A] = P[X>3] = 1-0.152 = 0.848$$



PROCESO ESTOCASTICO DE POISSON

CON BASE EN LA DISTRIBUCION DE POISSON SE PUEDE DEDUCIR QUE LA DISTRIBUCION DE PROBABILIDADES DEL NUMERO DE OCURRENCIAS DE UN EVENTO DURANTE UN PERIODO t QUEDA DADA POR

$$f_X(x) = P[X = x \text{ EN UN LAPSO } t]$$

$$f_X(x) = \frac{(\lambda t)^x \, e^{-\lambda t}}{x!} \; ; \; x = 0, 1, 2, \ldots$$

DONDE

λ = NUMERO MEDIO DE OCURRENCIAS POR UNIDAD DE TIEMPO.

LA <u>ESPERANZA</u> Y LA <u>VARIANCIA</u> DE ESTE PROCESO, PARA UN LAPSO t, SON

$$E(X) = \lambda t$$

$$\sigma^2(X) = \lambda t$$

PARA QUE ESTA DISTRIBUCION SE APLIQUE SE REQUIERE QUE EL EVENTO OCURRA CADA VEZ DE MANERA <u>INDEPENDIENTE</u> DE LAS OCURRENCIAS PREVIAS, Y QUE λ SEA CONSTANTE. A λ SE LE CONOCE COMO <u>INTENSIDAD DEL PROCESO</u>; A SU RECIPROCO, $1/\lambda$ SE LE DENOMINA *PERIODO DE RECURRENCIA*.

EJEMPLO

EN UNA CENTRAL DE COMUNICACIONES SE TIENE UNA DEMANDA MEDIA DEL SERVICIO DE 8 LLAMADAS CADA MINUTO. CALCULAR LAS PROBABILIDADES DE QUE EN 2 MINUTOS NO SE SOLICITE EL SERVICIO, DE QUE SE SOLICITE SOLO UNA VEZ, Y MAS DE UNA VEZ.

$$f_X(0) = P[X=0] = \frac{(\lambda t)^0 - e^{-8x2}}{0!} = e^{-16} = 0.00004$$

$$f_X(1) = \frac{16^1 e^{-16}}{1!} = 0.00064$$

$$P[X>1] = 1 - (0.00004 + 0.00064) = 0.99932$$

EJEMPLO

MEDIANTE UN ESTUDIO ESTADISTICO SOBRE LA OCURRENCIA DE MAREMOTOS EN LA COSTA MEXICANA DEL OCEANO PACIFICO SE ESTIMO QUE UNA OLA DE 4m DE ALTURA O MAYOR SOBRE EL NIVEL DE LA MAREA TIENE UN PERIODO DE RECURRENCIA DE 100 AÑOS. CALCULAR LAS PROBABILIDADES DE QUE EN LOS PROXIMOS 10, 50 y 100 AÑOS NO OCURRA NINGUN MAREMOTO EN DICHA REGION CUYA OLA MAXIMA EXCEDE DE 4m, SUPONIENDO QUE LA OCURRENCIA DE LOS MAREMOTOS SE PUEDE MODELAR MEDIANTE UN PROCESO ESTOCASTICO DE POISSON.

LA DISTRIBUCION DE PROBABILIDADES DE LA VARIABLE ALEATORIA X=NUMERO DE MAREMOTOS CUYA OLA MAXIMA ES MAYOR DE 4m, CON $\lambda=1/100=0.01$ ES

$$f_X(x) = \frac{\lambda t \; e^{-\lambda t}}{x!} = \frac{0.01t \; e^{-0.01t}}{x!}$$

POR LO TANTO, PARA t=10, 50 Y 100 AÑOS, SE TIENE, RESPECTIVAMENTE, QUE:

a) $$f_X(0) = \frac{(0.01x10)^0 e^{-0.01x10}}{0!} = e^{-0.1} = 0.905$$

b) $$f_X(0) = \frac{(0.01x50)^0 e^{-0.01x50}}{0!} = e^{-0.5} = 0.607$$

c) $$f_X(0) = \frac{(0.01x100)^0 e^{-0.01x100}}{0!} = e^{-1} = 0.368$$

PARA ESTE MISMO PROBLEMA, LAS PROBABILIDADES DE QUE OCURRA AL MENOS UN MAREMOTO CON OLA MAXIMA MAYOR DE 4m SON, RESPECTIVAMENTE,

a) $P[X \geq 1] = 1 - f_X(0) = 1 - 0.905 = 0.095$

b) $P[X \geq 1] = 1 - 0.607 = 0.393$

c) $P[X \geq 1] = 1 - 0.368 = 0.632$

VARIABLES ALEATORIAS CONTINUAS

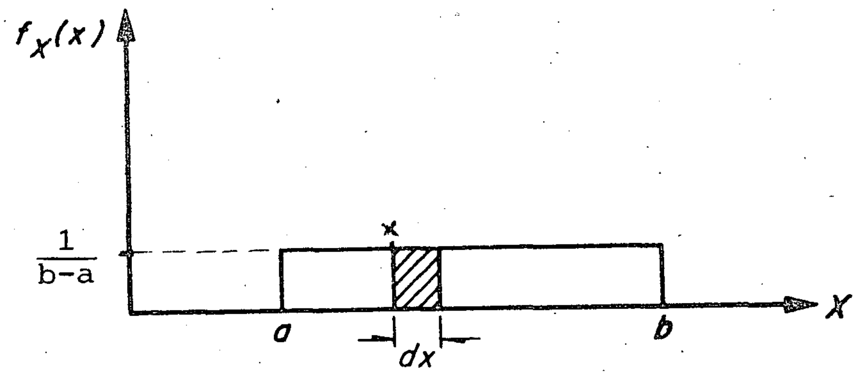
DISTRIBUCION UNIFORME

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

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

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

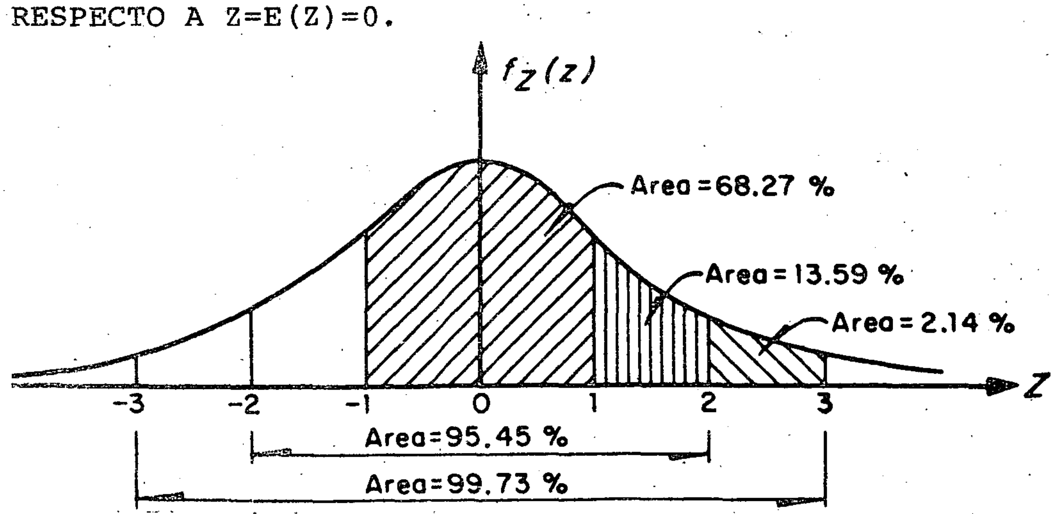
$$Z = (X-\mu)/\sigma$$

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

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

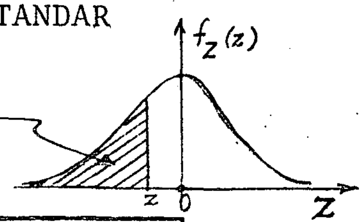
DONDE

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

TABLA 3

FUNCION DE DISTRIBUCION NORMAL ESTANDAR

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

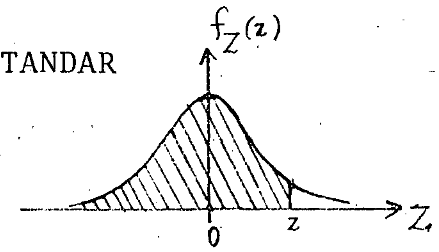


z	0	1	2	3	4	5	6	7	8	9
−3.	.0013									
−2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−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:

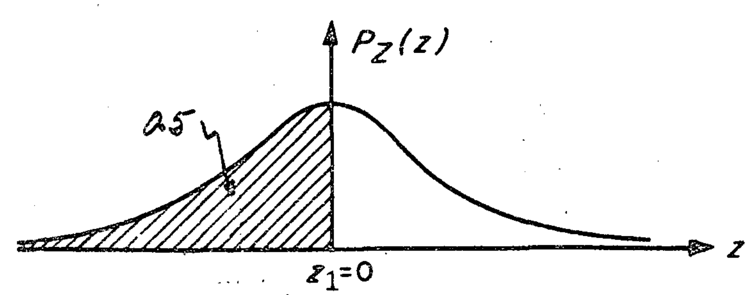


Fig 16. Distribución normal correspondiente . inciso c del ejemplo

TEOREMA CENTRAL DEL LIMITE

SEAN LAS VARIABLES ALEATORIAS $x_1, x_2, \ldots, x_k$, CON IDENTICAS 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 NÚMERO 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 $x_2=6.5$.

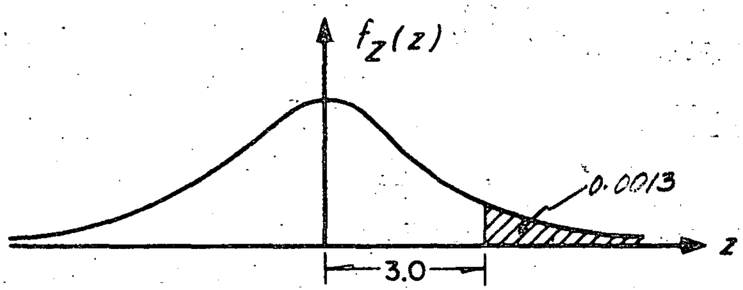
B) EL VALOR ESTANDARIZADO DE LA VARIABLE, PARA $x_1=330$ KG/CM^2, ES

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

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

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

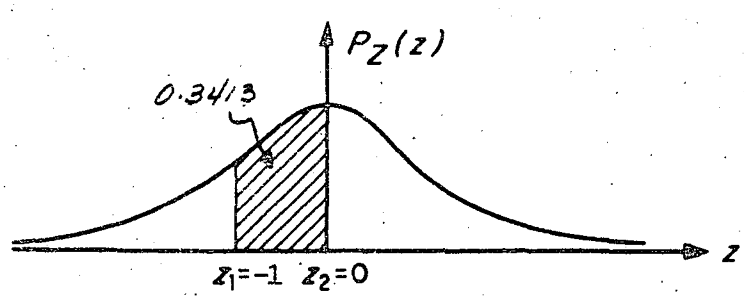


Fig 16. Distribución normal correspondiente a inciso c del ejemplo

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

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

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

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

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

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

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

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

DE DONDE

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

EJEMPLO

EN UNA SERIE DE 462 EXPERIMENTOS CON FINES ANTROPOLOGICOS, CONSISTENTES EN MEDIR EL TAMAÑO DE LA CABEZA DE LOS INDIGENAS RESIDENTES EN UNA ZONA TROPICAL, SE OBTUVIERON LOS RESULTADOS ANOTADOS EN LAS DOS PRIMERAS COLUMNAS DE LA SIGUIENTE TABLA. SI LA VARIABLE ALEATORIA "TAMAÑO DE LA CABEZA" SE CONSIDERA QUE TIENE DISTRIBUCION NORMAL, ¿QUE CANTIDAD DE RESULTADOS SE ESPERARIA OBTENER ANTES DE HACER LAS MEDICIONES, SI SE CONSIDERA QUE $\mu = \bar{x} = 191.8$MM y $\sigma = s = 6.48$MM, DONDE $\bar{x}$=PROMEDIO ARITMETICO Y s=DESVIACION ESTANDAR DE LOS DATOS?

$$z_1 = \frac{171.5 - 191.8}{6.48} = -3.13; \quad z_2 = \frac{175.5 - 191.8}{6.48} = -2.51; \quad x_3 = \frac{179.5 - 191.8}{6.48} =$$

$$= -1.90, \text{ ETC.}$$

$P(-3.13 \leq Z \leq -2.51) = 0.0051$; $P(-2.51 \leq Z \leq -1.90) = 0.0227$;

$P(-1.90 \leq Z \leq -1.28) = 0.0716$, ETC.

$462 \times 0.0051 = 2.4$; $462 \times 0.0227 = 10.5$; $462 \times 0.0716 = 33.1$, ETC.

INTERVALO DE VALORES DE X, EN MM	NUMERO DE OBSERVACIONES (frecuencia, f)	INTERVALO DE $Z=\frac{X-\mu}{\sigma}$	PROBABILIDAD $P(z_1 \leq z \leq z_2)=P$	FRECUENCIA ESPERADA = 462 P
171.5-175.5	3	(-3.13)-(-2.51)	0.0051	2.4
175.5-179.5	9	(-2.51)-(-1.90)	0.0227	10.5
179.5-183.5	29	(-1.90)-(-1.28)	0.0716	33.1
183.5-187.5	76	(-1.28)-(-0.66)	0.1543	71.3
187.5-191.5	104	(-0.66)-(-0.05)	0.2255	104.2
191.5-195.5	110	(-0.05)- 0.57	0.2356	108.8
195.5-199.5	88	0.57 - 1.19	0.1673	77.3
199.5-203.5	30	1.19 - 1.80	0.0811	37.5
203.5-207.5	6	1.80 - 2.42	0.0281	13.0
207.5-211.5	4	2.42 - 3.04	0.0066	3.0
211.5-215.5	2	3.04 - 3.66	0.0011	0.5
215.5-219.5	1	3.66 - 4.27	0.0001	0.0

TOTAL: 462

TOTAL: 461.6

BIBLIOGRAFIA DE CONTROL DE CALIDAD

1. Mathematical methods of statistical quality control, V. Sarkadi, Academic Press, 1974

2. Sensibilidad estadística de cartas de control de calidad, Universidad de La Habana, Centro de Información Científica y Técnica, 1971

3. Control de calidad, Seminario, Memoria del Instituto Mexicano del Petróleo, Publicación No. 69 JF/050, Ed. Instituto Mexicano del Petróleo, 1969

4. Control de calidad estadístico, E.L. Grant, Ed. CECSA, 1974

5. Control total de la calidad, A.V. Feigenbaum, Ed. CECSA, 1975

6. Practical Quality Control, Simmons A. David, Ed. Addison-Wesley Publishing Company, 1970

7. Quality Control Handbook, J.M. Juran, Gryna, F.M. y R.J. Bingham, 3a. edición, Ed. McGraw-Hill, 1974

8. Quality Planning and Analysis, J.M.Juran y F.M. Gryna, Ed. McGraw-Hill, 1970

9. Construction Specifications Writing: Principles and Procedures, H.J. Rosen, Ed. J. Wiley, 1974

10. Transactions of the American Society for Quality Control.

---

1. Kreyszig, E., "Introducción a la estadística matemática", Limusa-Wiley (1973)

2. Larson, H.J., "Introducción a la teoría de probabilidades e inferencia estadística", Limusa-Wiley (1978)

BIBLIOGRAFIA SELECCIONADA SOBRE TEORIA DE LA PROBABILIDAD, ESTADISTICA MATEMATICA, ESTADISTICA APLICADA, MUESTREO, Y PROCESOS ESTOCASTICOS, Y SUS APLICACIONES A LA INGENIERIA, DISPONIBLE EN LA BIBLIOTECA DE LAS DIVISIONES DE INVESTIGACION Y DE ESTUDIOS SUPERIORES.

1. Anderson, T., and D. Darling, Asymptotic theory of certain goodness of fit criteria based on stochastic processes. C-1097, A

2. Arrow, K., et al, Bayes and minimax solutions of sequential decision problems. C-1101, A

3. Bair, D., Experimentation: an introduction to measurement theory and experiment design, Prentice Hall, 1962. QC39 B2

4. Barucha-Reid, A., ed., Probabilistic methods in applied mathematics, Academic Press, 1968. QA273 B42

5. Bellman, R., Programmed statistics; with chapters on probability, computer theory, and programmed instruction, Holt, Rinehart and Winston, 1970. HA29 B45

6. Benjamin, J., Probability, statistics, and decision for civil engineers, McGraw-Hill, 1970. QA273 B46

7. Bhat, U., Elements of applied process, Wiley, 1972. QA274 B42

8. Blackman, R., Linear data-smoothing and prediction in theory and practice. Addison-Wesley, 1965. QA275 B55

9. Blackwell, D., Another countable Markow process with only instantaneous states. C-1589, B.

10. Box, G., A bayesian approach to some outlier problems. C-1586, B

11. Box, G., Time series analysis; forecasting and contro Holden-Day, 1970. QA276 B6

12. Box, G., Evolutionary operation; a statistical method for process improvement, Wiley, 1969. TP155.7 B66

13. Breiman, L., Probability and stochastic processes. Houghton Mifflin, Co., 1969. QA273 B73

14. Breipohl, A., Probabilistic system analysis; an introduction to probabilistic models, decisions, and applications of random processes, Wiley, 1970. QA273 B746

15. Brown, R., Smoothing, forecasting and prediction of discrete time series, Prentice-Hall, 1963. TA168 B6

16. Brown, W., and C. Palermo, Random processed, communications, and radar, McGraw-Hill, 1969. TK5101 B75

17. Brownlee, K., Statistical theory and methodology in science and engineering, Wiley, 1960. QA276 B77

18. Bruning, J. and B. Kintz, Computational handbook of statistics, Scott, Foreman and Co., 1968. HA29 B77

19. Bush, R., A stochastic model with applications to learning. C-1590, B

20. Castro, G. de, Introducacio ao curso sobre instrumento matematicos da estatística. F-2164, C

21. Clarke, B., Probability and random processes for engineers and scientists, Wiley, 1970. QA273 C48

22. Cochran, W., Experimental designs, Wiley, 1957. Q180 A1C6

23. Conover, W., Practical nonparametric statistics, Wiley, 1971. QA278.8 C65

. Cooper, G. Probabilistic methods of signal and systems analysis, Holt, Rinehart and Winstons, 1971. TK454.2 C664

. Cornell, C., A probability-based structural code, ACI, 1968.

. Derman, C., Finite state markovian decision processes, Academic Press, 1970.

. Deutsch, R., Nonlinear transformation of random processes. 517.7 D.

. Dubes, R., The theory of applied probability, Prentice-Hall, 1968. TK5101 D8

. Dynkin, E., Markow processes; theorems and problems, Plenum Press, 1969. QA273 D894

. Edwards, A., Experimental design in psichological research, Holt-Rinehart and Winsont, 1968. BF59 E37

1. Ehrenfeld, S. and S. Littauer, Introduction to statistical method, McGraw-Hill, 1964. 519.9 E

2. Esteva, L., Consideraciones práctias en la estimación bayesiana de riesgo sísmico, México, UNA. Instituto de Ingeniería, 1970. F-8843, E

33. Feller, W., Introducción a la teoría de probabilidades y sus aplicaciones, Limusa-Wilesy, 1973. QA273 F3714

34. Forcadas, J., Estadística aplicada a la ingeniería, 1969. QA276 F66

35. Freeman, H., Introducción a la inferencia estadística, Trillas, 1970. QA276 F6845

36. Freund, J., Mathematical statistics, Prentice Hall, 1971. QA276 F692

37. Freudenthal, H., Probability and statistics, Elsevier, 1965. QA273 F75

38. García, A., Elementos de método estadístico, México, UNA, 1966. HA29 G146

39. Gotkin, L., Estadística descriptiva; texto programado, Limusa-Wiley, 1967. HA29 G695

40. Greenwood, A. and H. Hartley, Guide to tables in mathematical statistics, Princenton University Press, 1962. 519.9 G

41. Grenander, U., Statistical spectral analysis of time series arising from stationary stochastic processes. C-1588, G

42. Hanman, H., Modern factor analysis, University of Chicago Press, 1967. QA276 H38

43. Hanna, E., Time series analysis, Methuen, 1960. QA76 H32

44. Hays, W., Statistics: probability inference and decisions, Holt-Rinehart and Winston, 1970. QA276 H39

45. Hoel, P., Elementary statistics, Wiley, 1971. HQ29 H66

46. Hoel, P., et al, Introduction to stochastic processes, Houghton, Mifflin, 1972. QA274 H63

47. Howard, R., Dynamic probabilistic systems, Wiley, 1971 T57.95 H66

48. Iosifoscu, M., Random process and learning, Springer Verlag, 1969. QA273 I64

49. Jazwinski, A., Stochastic process and filtering theory, Academic Press, 1970. QA276 J38

50. Kefer, J., and J. Wolfowitz, Sotchastic estimation of the maximum of a regression function. C-1481, K

51. Kemeny, J., Finite Markow chains, Van Nostrand, 1960.
QA273 K33

52. Kish, L., Muestreo de encuestas, Trillas, 1972.
HN29 K5

53. Kozin, F. and J. Bogdanoff, An introduction to random function for engineer, 1963. F-1930, K

54. Kyburg, H., Probability theory, Prentice-Hall, 1969.
QA273 K92

55. Lange, F., Correlation techniques: Foundation and applications of correlation analysis in modern communications, measurement and control, Iliffe, 1967.
TK7870 L34

56. Larson, H., Introduction to probability theory and statistical inference, Wiley, 1969. QA273 L37

57. La Valle, I., An introduction to probability decision, and inference, 1970. QA276 L36

58. Lecture notes in mathematics, probability and information theory, Springer Verlag, 1964. QA1 L42

59. Lee, T., et al, Estimating the parameters of the Markow probability model, North Holland, 1970.
HB74 M3L43

60. Lindgren, B., Statistical theory, Macmillan, 1968.
QA276 L55

61. Lin, Y., Random processes, F-8709, L.

62. Prelewicz, D., Range of validity of the method of averaging. F-9116, P

63. Prohorov, Y., Probability theory; basic concept, limit theorems, random processes, Springer Verlag, 1969. QA273 P75

64. Raiffa, H. and R. Schlaifer, Applied statistical decision theory, Harvard University Press, 1971.
QA279.4 R34

65. Raj, D., Sampling theory, McGraw-Hill, 1968.
QA276.6 R33

66. Raj, D., Design of sample survey, McGraw-Hill, 1972.
HA312 R33

67. Rascón, O., Introducción a la teoría de probabilidades. México, UNA, 1971. QA273 R37

68. Rascón, O., Introducción a la estadística descriptiva. 1970. HA31 R37

69. Ravindra, M., Probabilistic evaluation of safety factors, University of Waterloo, 1969. F-8680, R

70. Robinson, E., Multichanel time series analysis with digital computer programs, Holden-Day, 1967.
QA276 R633

71. Rosenblueth, E., Current research on probabilistic methods at the National University. C-1517, R

72. Ross, S., Applied probability models with optimization applications, Holden-Day, 1970. QA273 R67

73. Sage, A., Estimation theory with applications to communications and control, McGraw-Hill, 1971.
QA276.8 S33

74. Savage, L., The theory of static decision, C-1098, S

75. Sawaragi, Y., Statistical decision theory in adaptive control system, Academic Press, 1967. QA402.3 S37

76. Schäl, M., Markow renewal processes with auxiliary phats. C-1617, S

77. Schlaifer, R., Probability and statistics for business decisions and introduction to mangerial economics under uncertainty, McGraw-Hill, 1959. HD38 S35

78. Schwarts, J., Statistical methods in traffic engineering, The Ohio State University, 1967. HE336 S7S3

79. Searle, S., Linear models, Wiley, 1971. QA279 S4

80. Sengupta, J., Stochastic programming methods and applications, North Holland, 1973. T57.79 S44

81. Sengupta, S. and S. Jain, A representative theory for measurable random variable. F-5967, S

82. Sheppard, R., Multidimensional scaling, Seminar Press, 1972. CF39 M84

83. Spiegel, M., Theory and problems of statistics, Schaum's Publishing, Co., 1960. HA29 S65

84. Sterling, T., Introduction to statistical data processing. Prnetice-Hall, 1968. QA276.4 S82

85. Stratonovich, R., Conditional Markow processes and their application to the theory of optimal control, Elsevier, 1968. QA273 S76

86. Symposium on time series analysis proceedings, held at Brown University, June 11-14, 1962. Ed. by Murray Rosenblatt. New York, John Wiley and Sons, 1963. 519.9 S

87. Tannur, M., Statistics: a guide to the unknow, Holden Day, Inc., 1972. QA276 S82

88. Theil, H., Statistical decomposition analysis with applications, North Holland, 1972. H61 T43

89. Thomas J., An introduction to statistical communication theory, Wiley, 1969. TK5102.5 T45

90. Tintner, G., Stochastic economics, stochastic processes, control and programming. Academic Press, 1972. QA274 T54

91. Turkstra, C., Applications of bayesian decision theory, University of Waterloo, 1969. F-8619, T

92. Turkstra, C., Elements of probability theory. F-9105, T

93. Turkstra, C., Theoretical distribution functions. F-9104, T

94. Valdes, R., Nociones de cálculo de probabilidades y estadística, Limusa-Wiley, 1970. QA273 V34

95. Van der Gerr, J., Introduction to multivariate analysis for the social sciences, Freeman, 1971. QA278 V34

96. Vere-Jones, D., Stochastic models for earthquake occurence; discussion. F-9247, V

97. Waerden, B., Mathematical statistics, Springer-Verlag 1957, QA276 W3

98. Wold, H., ed., Bibliography on time series stochastic processes: an international team project. Edinburgh, Oliver and Boyd, 1965.

99. Yaglom, A., An introduction to the theory of stationa random function, Prentice-Hall, 1962. 517.5 Y

Recopilación elaborada p[illegible] el pr[illegible] [illegible] Kleiman

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

PROBABILIDAD Y ESTADISTICA: FUNDAMENTOS Y APLICACIONES

ESTADISTICA DESCRIPTIVA

DR. OCTAVIO A. RASCON CHAVEZ

JUNIO, 1978

ESTADISTICA DESCRIPTIVA

POR DR. OCTAVIO A. RASCON CH.

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

<u>EJEMPLOS</u>

1. EXPERIMENTO: LANZAMIENTO DE UNA MONEDA DIEZ VECES

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

   MUESTRA: GRUPO DE 10 OBSERVACIONES

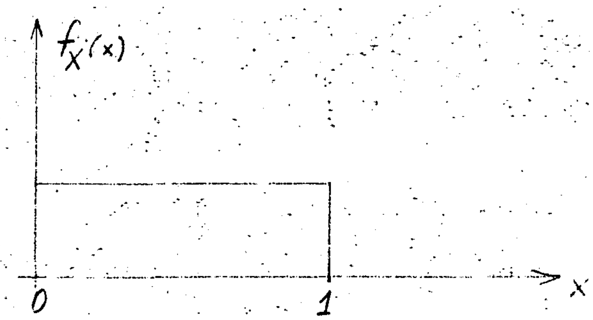
2. EXPERIMENTO: MEDICION DE LA PRECIPITACION PLUVIAL DIARIA EN LA CIUDAD DE MEXICO DURANTE DIEZ AÑOS

   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.

EJEMPLO

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 RENGLONES IMPARES 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
		$\Sigma=30$	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 relativa acumulada
\: 51-60	59,57	2	2/30=0.067(6.7%)	2	0.067
3: 61-70	67,65,69,67,67	5	5/30=0.166(16.6%)	2+5=7	0.067+0.166=0.233
.: 71-80	72,73,73,77,78,78,	6	6/30=0.200(20%)	7+6=13	0.233+0.200=0.433
): 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
3: 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		

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

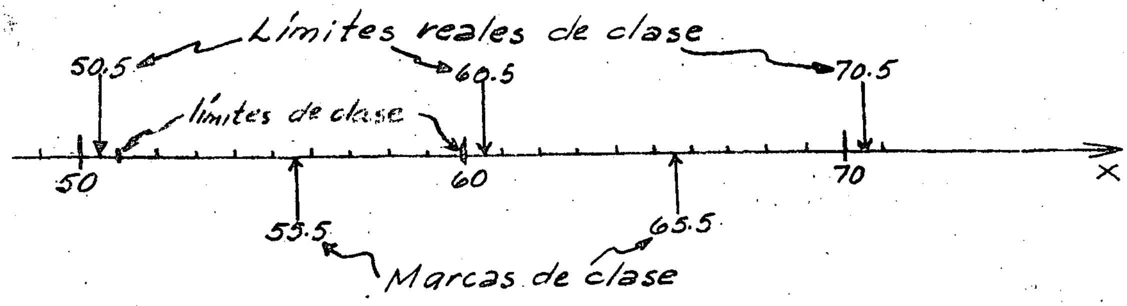
$$\text{NUMERO DE INTERVALOS} = \frac{30}{5} = 6$$

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



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

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

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

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

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

LIMITES REALES INFERIORES DE CLASE

LIMITES REALES SUPERIORES DE CLASE

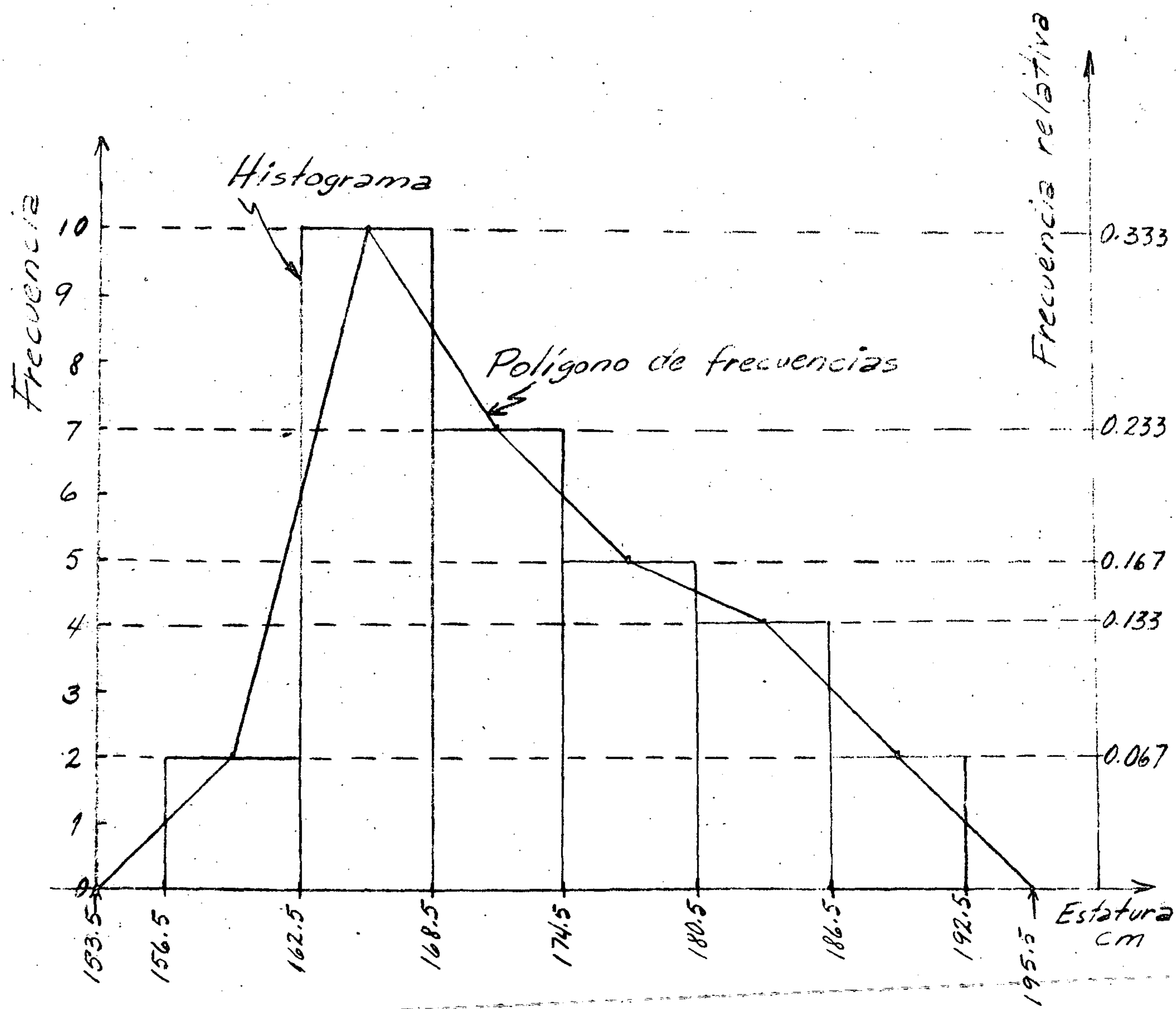
A MAYOR NUMERO DE DATOS SE REQUIERE MAYOR NUMERO DE INTERVALOS, PERO SE RECOMIENDA QUE ESTE NUMERO ESTE ENTRE 5 Y 20, SUPONIENDO QUE EN PROMEDIO CAIGAN 5 O MAS ELEMENTOS EN CADA INTERVALO. ASI, SI SE TIENEN 30 DATOS, SE RECOMIENDA USAR 30/5=6 INTERVALOS.

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

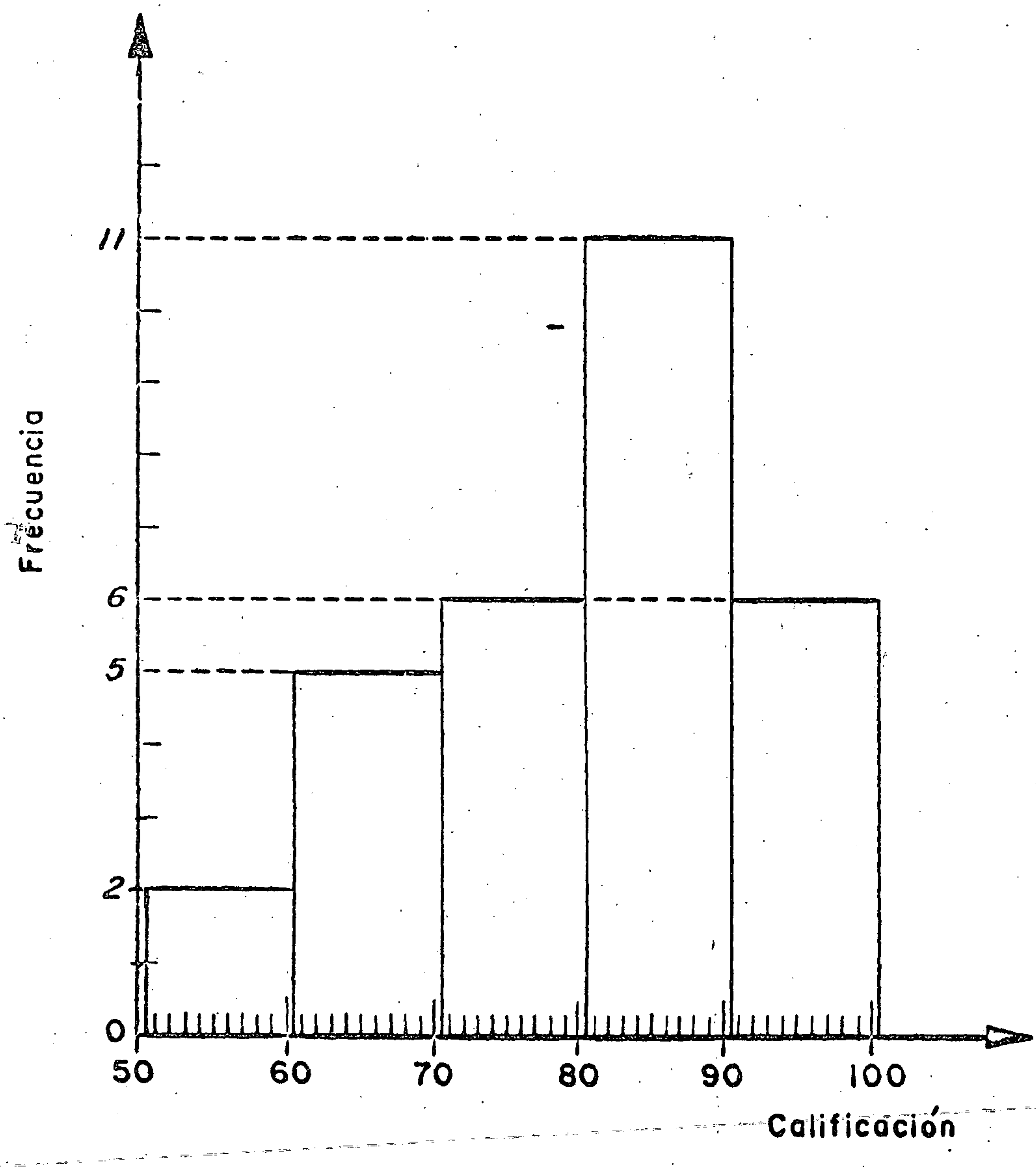
PRESENTACION GRAFICA DE LAS DISTRIBUCIONES DE FRECUENCIAS



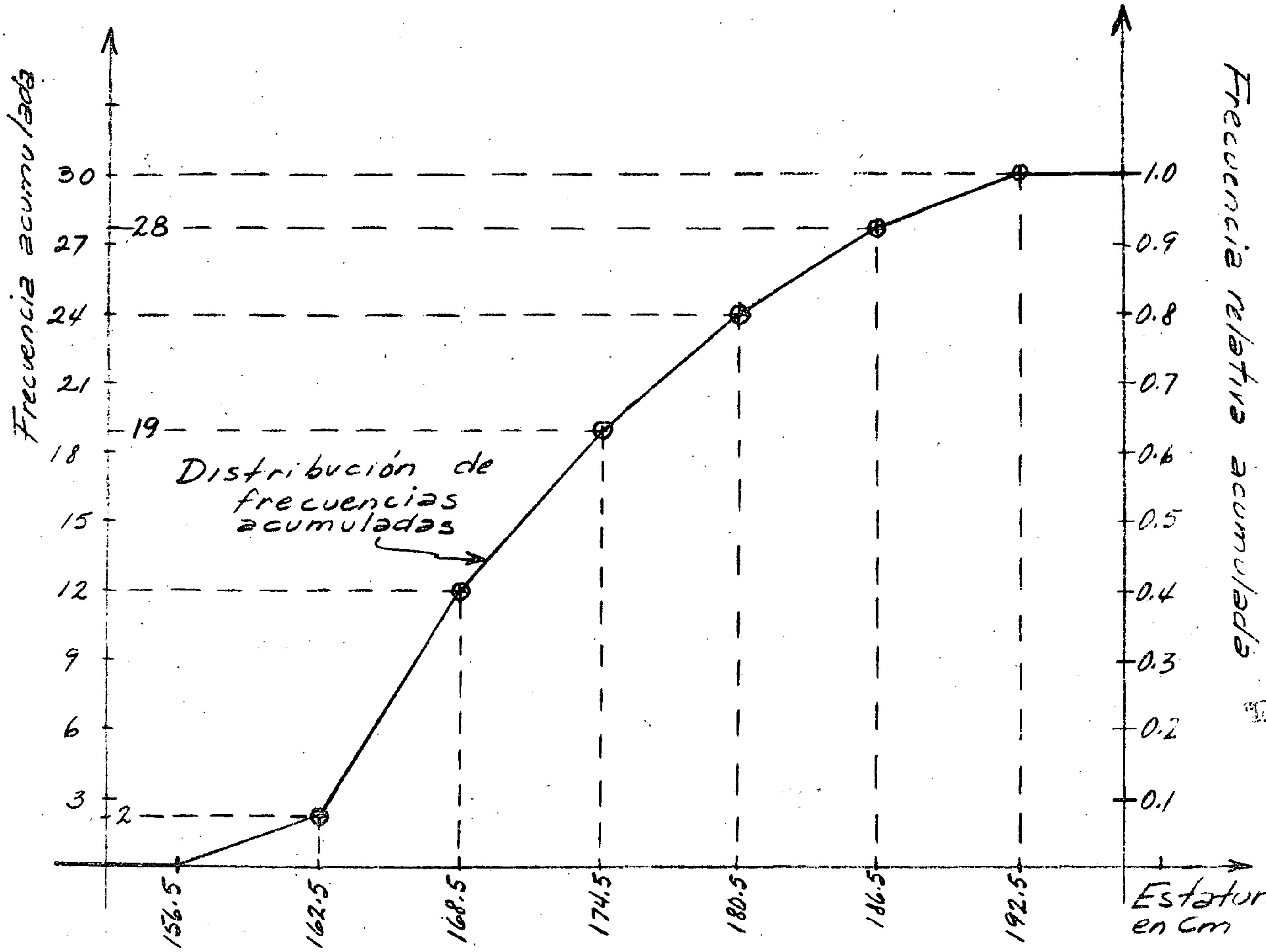
4. INTEGRACION DE LA TABLA:

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
			$\Sigma=30$	$\Sigma=1.000$		

HISTOGRAMA DEL PROBLEMA DE LAS CALIFICACIONES EN PEDAGOGIA



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



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

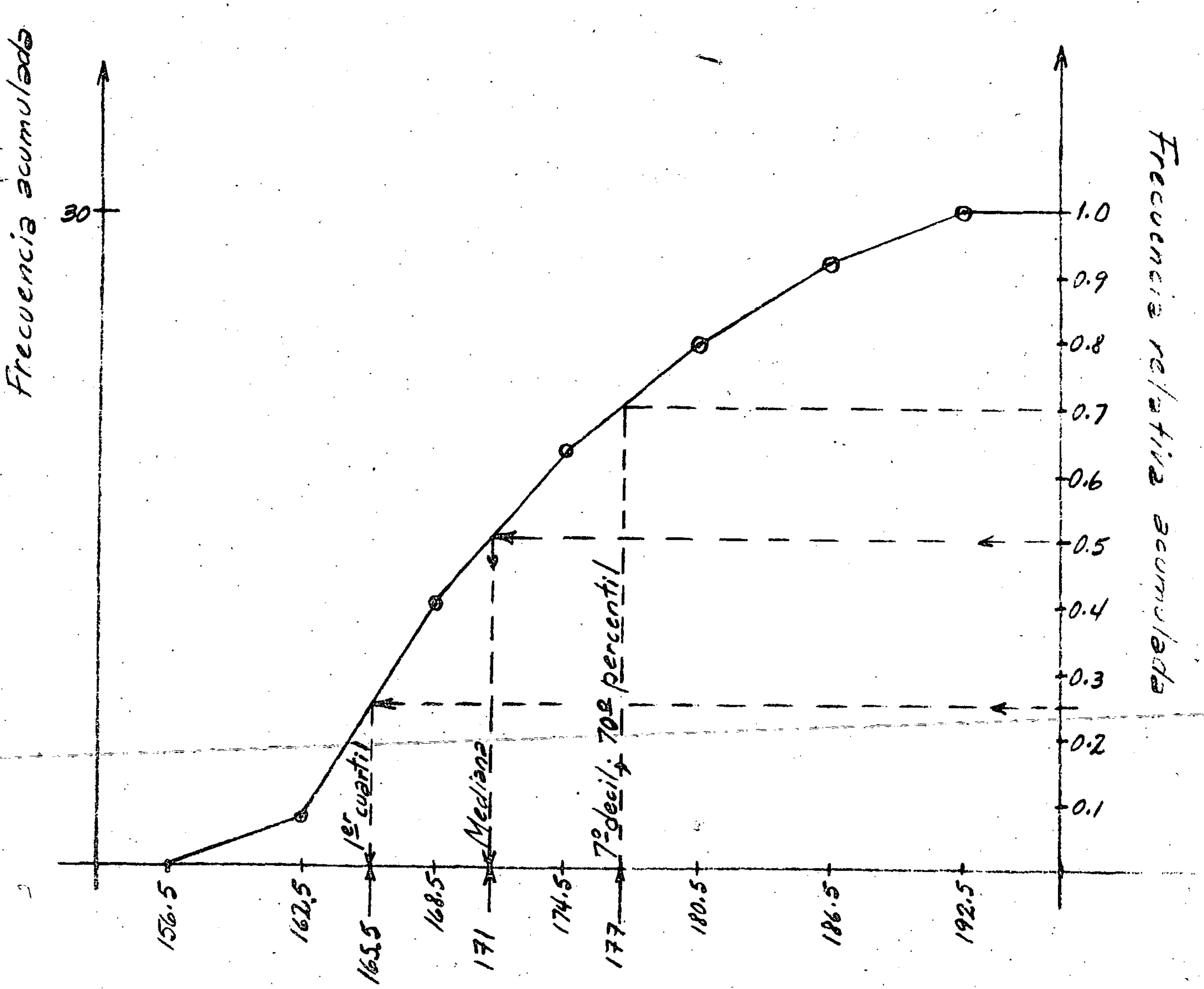
LA FRECUENCIA RELATIVA ACUMULADA COMPLEMENTARIA ES: 1-0.800=0.200 (20

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.

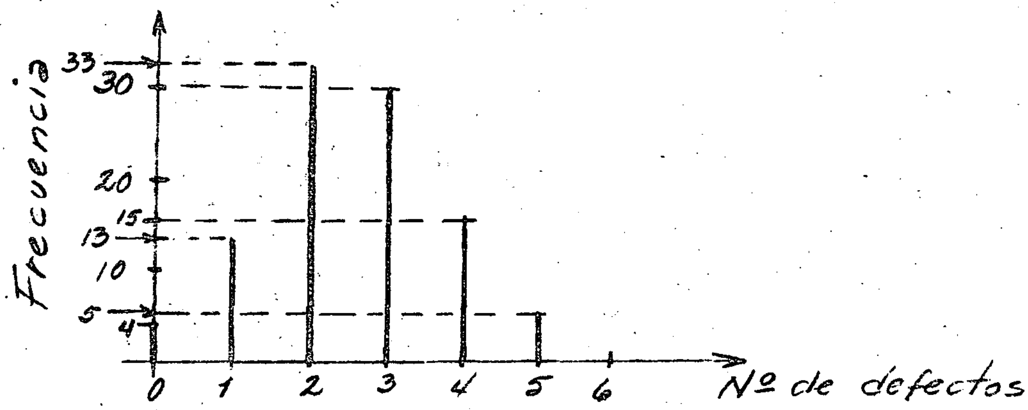
MEDIANA: VALOR DE LA VARIABLE CORRESPONDIENTE A LA FRECUENCIA RELATIVA ACUMULADA DE 50%.

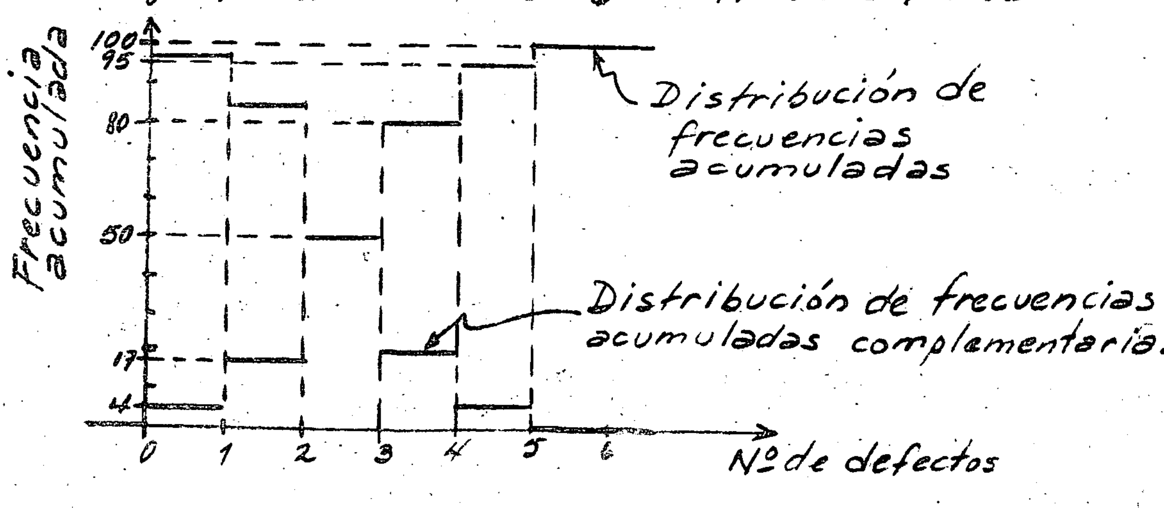


EJEMPLO

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		





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.

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

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

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

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

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

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

MEDIANA: 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):

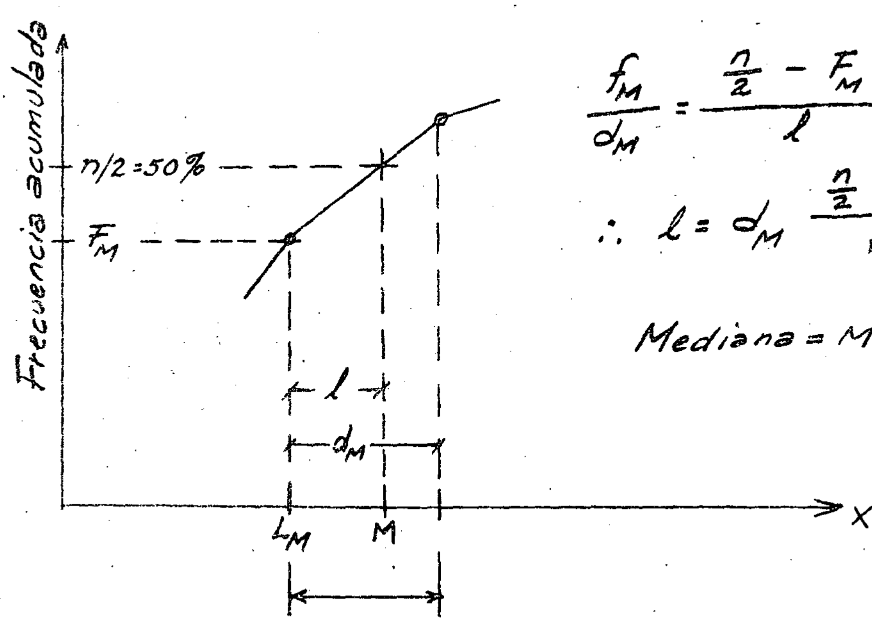
$$\text{MEDIANA} = M = L_M + \frac{\frac{n}{2} - F_M}{f_M} d_M$$

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



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

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

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

Intervalo que contiene a la mediana

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

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

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

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

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

TRANSFORMACION DE VARIABLES

SEA X UNA VARIABLE ALEATORIA CON DENSIDAD DE PROBABILIDADES $f_X(x)$, Y SEA LA TRANSFORMACION

$$Y = g(X) = a + bX$$

PARA EL CASO EN QUE X ES UNA VARIABLE CONTINUA, EN TEORIA DE PROBABILIDADES SE VIO QUE

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

$$= a + bE(X)$$

Y, ANALOGAMENTE QUE

$$\sigma^2(Y) = b^2\sigma^2(X)$$

(ESTOS RESULTADOS SON VALIDOS TAMBIEN PARA VARIABLES ALEATORIAS DISCRETAS.)

AHORA, SI SE TIENE UNA MUESTRA DE TAMAÑO n DE LA VARIABLE X, A CADA VALOR, x_i, DE DICHA MUESTRA LE CORRESPONDE UN VALOR, y_i, DE LA MUESTRA DE Y DADO POR

$$y_i = a + bx_i$$

POR LO TANTO, EL PROMEDIO ARITMETICO DE LAS y_i ES

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n} (a + bx_i) = \frac{1}{n}\sum_{i=1}^{n} a + \frac{b}{n}\sum_{i=1}^{n} x_i = a + b\bar{x}$$

ANALOGAMENTE, EL VALOR MEDIO CUADRATICO RESULTA SER

$$\overline{y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2 = \frac{1}{n}\sum_{i=1}^{n} (a+bx_i)^2 = \frac{1}{n}\sum_{i=1}^{n} a^2 + \frac{1}{n}\sum_{i=1}^{n} 2abx_i + \frac{1}{n}\sum_{i=1}^{n} b^2x_i^2$$

$$= a^2 + 2ab\bar{x} + b^2\overline{x^2}$$

TAREA

0.78	0.38	0.72	0.65	0.72	0.92	0.78	0.65	0.92	0.78
1.36	1.43	0.65	0.48	0.83	0.48	0.72	0.48	0.65	0.78
0.65	1.00	0.78	0.78	1.03	1.26	0.48	0.48	1.06	0.96
0.65	0.92	0.72	0.78	0.78	0.48	0.28	0.36	0.83	0.48
0.78	0.49	0.36	0.78	0.78	0.83	0.88	0.96	1.03	1.21
0.88	0.57	0.72	1.03	0.92	0.96	0.78	1.09	0.92	1.12
0.65	0.65	0.83	0.72	0.72	0.78	0.72	1.09	0.83	0.83
0.83	1.06	0.57	0.78	1.23	1.09	1.03	0.18	0.65	1.34
0.96	0.65	0.48	1.18	1.12	0.18	0.48	0.72	0.57	0.55
0.96	0.65	0.96	0.51	0.65	1.21	1.48	0.96	0.96	1.40

1. Agrupar datos por intervalos y elaborar tabla con frecuencias, frecuencias relativas, frecuencias acumuladas y frecuencias relativas acumuladas (anotar límites, límites reales y marcas de clase).

2. Dibujar:
   a) Histograma
   b) Polígono de frecuencias
   c) Curva de frecuencias acumuladas

3. Calcular todas las medidas de tendencia central y de dispersión que se han estudiado
   a. Sin agrupar datos
   b. Con datos agrupados

EJEMPLO

EN EL PROBLEMA DE LOS RESULTADOS, x_i, DE UN EXAMEN SOBRE PEDAGOGIA SE OBTUVO LA DISTRIBUCION DE FRECUENCIAS INDICADAS EN LAS DOS PRIMERAS COLUMNAS DE LA SIGUIENTE TABLA:

MARCAS DE CLASE x_i	FRECUENCIAS f_i	MARCAS DE CLASE TRANSFORMADA, y_i	$y_i f_i$	y_i^2	$y_i^2 f_i$
55.5	2	-2	- 4	4	8
65.5	5	-1	- 5	1	5
75.5	6	0	0	0	0
85.5	11	1	11	1	11
95.5	6	2	12	4	24
	Σ=30		Σ=14		Σ=48

$\bar{y} = 14/30 = 0.467$, $\overline{y^2} = 48/30 = 1.6$, $S^2(y) = 1.6 - (0.467)^2 = 1.382$

$\bar{x} = [0.467 - (-7.55)]/(1/10) = 80.17$, $S^2(x) = 1.382/(0.1)^2 = 138.2$

CALCULAREMOS EL PROMEDIO Y LA VARIANCIA DE ESTA MUESTRA, CALCULANDO PRIMERO $\bar{y}$ Y $S^2(y)$ DE LA TRANSFORMACION

$$y = a + bx = \frac{x - C_1}{C_2}$$

$(a = \frac{-C_1}{C_2}$; $b = \frac{1}{C_2}$, CON C_1 = MARCA DE CLASE CENTRAL Y

C_2 = ANCHO DE CLASE)

TOMANDO $a = -75.5/10$ y $b=1/10$ ($C_1=75.5$ y $C_2=10$), SE TIENE

$$y_i = (-75.5 + x_i)/10$$

POR LO QUE

Y, LA VARIANCIA,

$$S^2(y) = \overline{y^2} - \bar{y}^2 = a^2 + 2ab\bar{x} + b^2\overline{x^2} - (a+b\bar{x})^2 = b^2\overline{x^2} - b^2\bar{x}^2 = b^2S^2(x)$$

ESTAS TRANSFORMACIONES SE PUEDEN EMPLEAR PARA CALCULAR EL PROMEDIO $\bar{y}$, Y LA VARIANCIA $S^2(y)$ DE LA MUESTRA DE UNA VARIABLE QUE RESULTA DE UNA TRANSFORMACION Y, CON BASE EN ELLOS, CALCULAR $\bar{x}$ Y $S^2(x)$ DE LA MUESTRA ORIGINAL, MEDIANTE LAS ECUACIONES

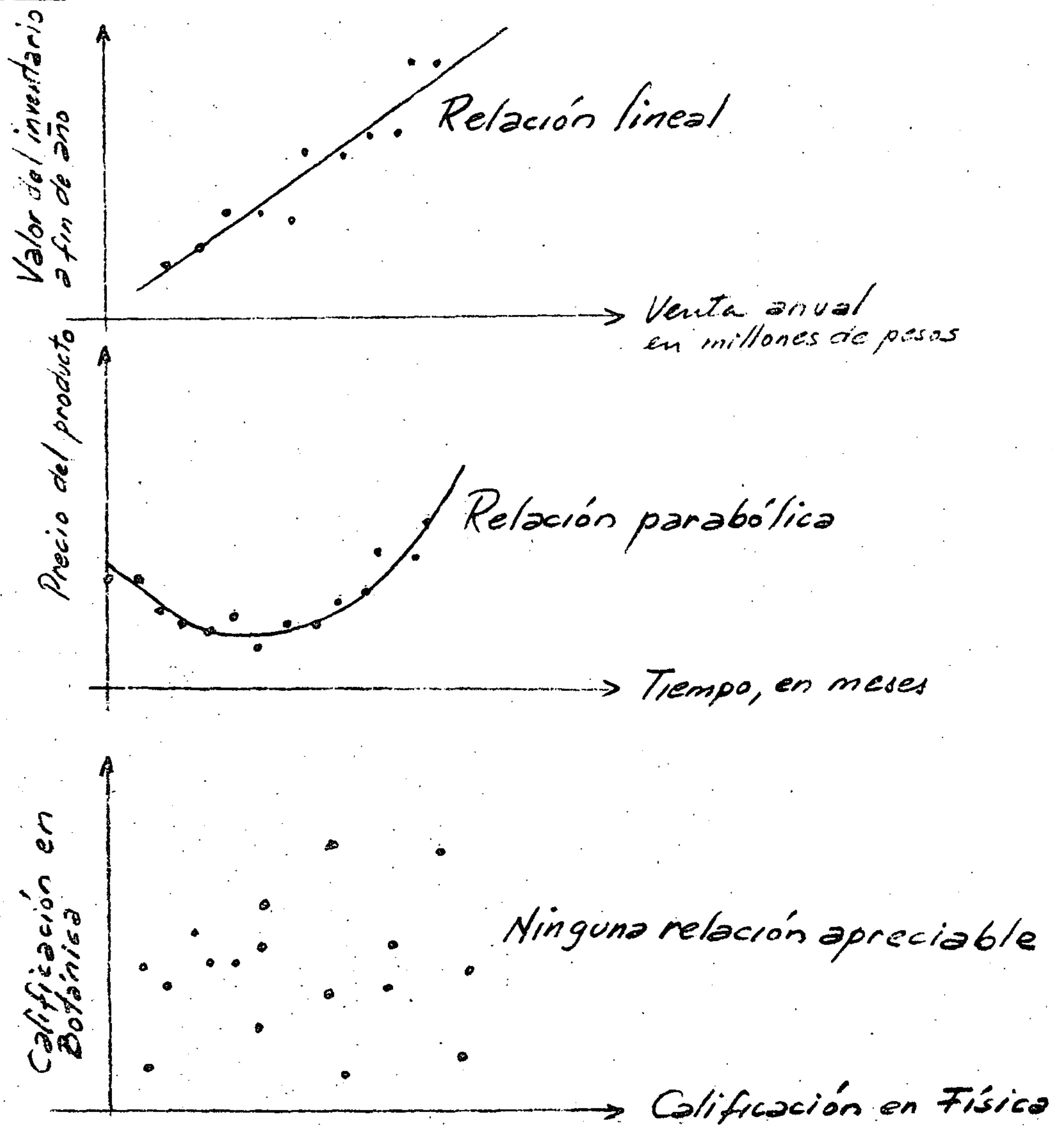
$$\bar{x} = (\bar{y} - a)/b$$
$$S^2(x) = S^2(y)/b^2$$

ESTE PROCEDIMIENTO AHORRA BASTANTE TIEMPO DE CALCULOS CUANDO LOS DATOS ESTAN AGRUPADOS, EN CUYO CASO LOS x_i SON LAS MARCAS DE CLASE.

REGRESION LINEAL

CON MUCHA FRECUENCIA SE PRESENTAN PROBLEMAS EN QUE INTERVIENEN DOS VARIABLES ALEATORIAS (O UNA ALEATORIA Y UNA DETERMINISTA) Y SE DESEA DETERMINAR UNA RELACION FUNCIONAL ENTRE ELLAS. SI SE OBTIENE UNA MUESTRA DE PAREJAS DE DATOS (x_i, y_i) Y SE ANOTAN EN UNA GRAFICA X-Y, VISUALMENTE SE PODRA PREVEER EL TIPO DE RELACION ENTRE AMBAS VARIABLES, Y LUEGO HACER UN AJUSTE MATEMATICO DE ALGUN TIPO DE CURVA.

EJEMPLOS (GRAFICAS DE CORRELACION)



$$y_1 = (-75.5 + 55.5)/10 = 2$$

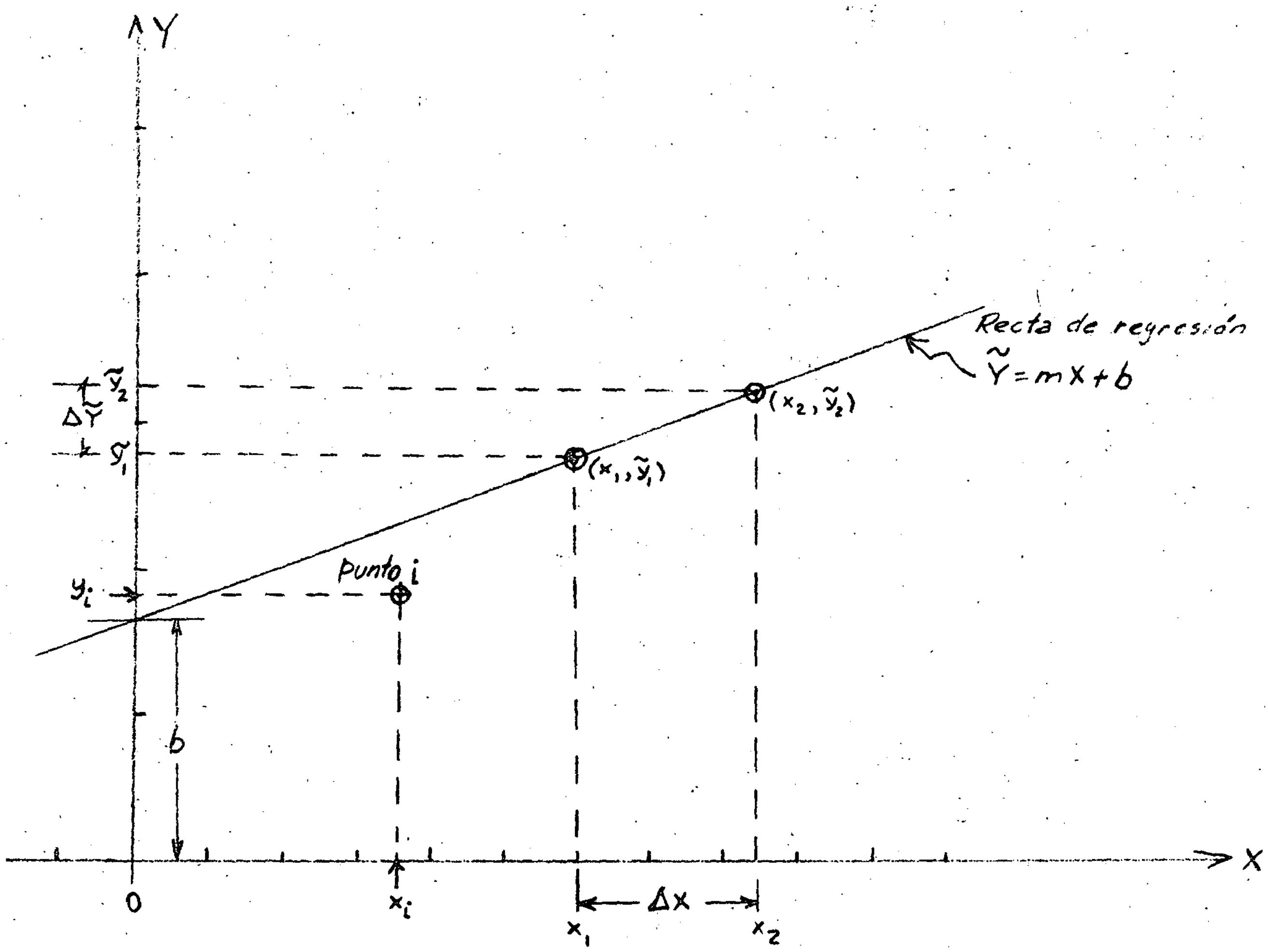
$$y_2 = (-75.5 + 65.5)/10 = -1$$

$$y_3 = (-75.5 + 75.5)/10 = 0$$

$$y_4 = (-75.5 + 85.5)/10 = 1$$

$$y_5 = (-75.5 + 95.5)/10 = 2$$

OBSERVESE QUE SE OBTIENE $y = 0$ PARA EL INTERVALO CORRESPONDIENTE A $X_i = C_1$, Y PARA LOS INTERVALOS CON VALORES MAYORES DE X BASTA CON IRLE SUMANDO UNA UNIDAD, MIENTRAS QUE A LOS DE VALORES MENORES, IRLE RESTANDO UNA UNIDAD.



$$m = \Delta\tilde{Y}/\Delta x$$

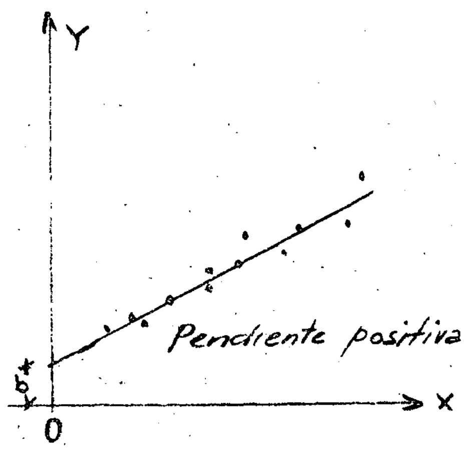
$$\Delta\tilde{Y} = \tilde{y}_2 - \tilde{y}_1$$

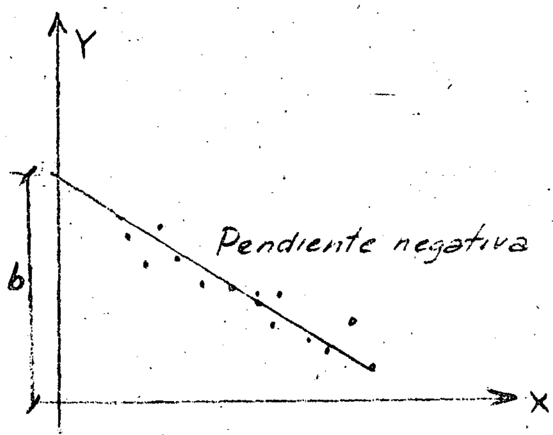
$$\Delta x = x_2 - x_1$$

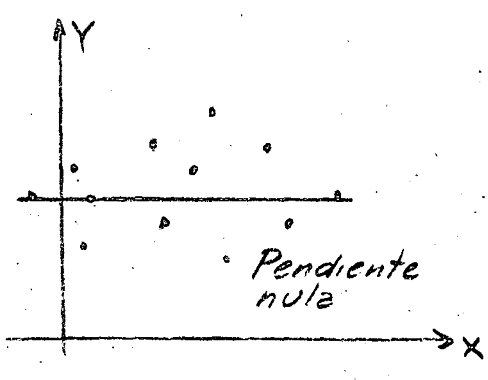
PARA AJUSTAR ALGUNA CURVA A UN GRUPO DE DATOS SE PUEDE PROCEDER DE DIFERENTES MANERAS, DE LAS CUALES LA MAS SENCILLA ES "A OJO", PERO TIENE LA DESVENTAJA DE QUE, POR NO SER SISTEMATICO, DIFERENTES PERSONAS PROPONEN DISTINTAS CURVAS. DE LOS METODOS ANALITICOS O MATEMATICOS, EL MAS COMUN ES EL DE *MINIMOS CUADRADOS*.

SI X ES LA VARIABLE INDEPENDIENTE Y Y LA DEPENDIENTE, SE DICE QUE LA REGRESION ES DE Y CON BASE EN X, Y VICEVERSA.

EN ESTE CURSO NOS CONCRETAREMOS AL CASO DE UN AJUSTE LINEAL, ES DECIR, MEDIANTE UNA LINEA RECTA, DE ECUACION $\tilde{Y}=mX + b$, EN DONDE m ES LA *PENDIENTE* Y b LA *ORDENADA AL ORIGEN*.







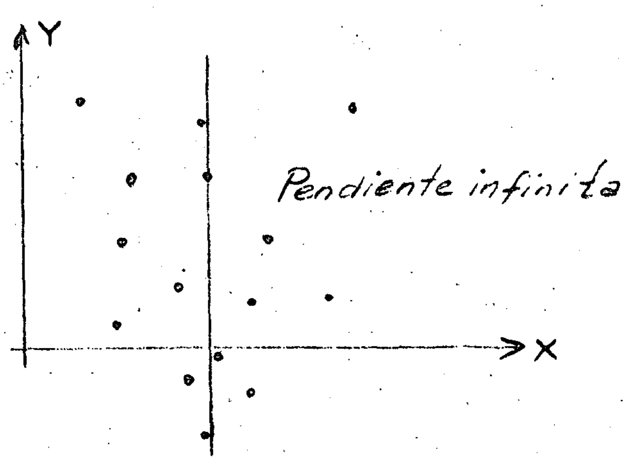


TABLA 1 , ESTATURA, x (EN CM), Y CIRCUNFERENCIA DE LA CABEZA, Y (EN CM), EN BEBES AL NACER (DATOS DEL PROF. E. NAVRATIL, UNIVERSITY HOSPITAL, GRAZ, 1962)

x	y	x	y	x	y	x	y	x	y
52	36	50	33	51	34	51	36	48	33
48	34	48	34	49	34	53	33	48	33
50	34	51	36	51	36	51	36	50	36
51	34	54	38	51	34	49	34	49	32
47	35	49	34	50	35	51	35	49	35
51	35	49	33	47	35	50	34	48	34
52	36	49	33	49	34	49	35	50	34
52	36	50	34	49	33	50	33	49	34
53	37	48	33	49	35	47	33	49	34
48	34	52	34	52	36	50	35	49	33
50	34	50	34	51	37	49	34	48	34
52	37	50	33	50	35	50	34	50	35
52	36	49	35	56	39	48	34	49	33
50	35	51	35	52	34	47	35	50	32
50	34	53	35	47	34	50	35	54	37
49	34	48	32	53	36	53	36	50	35
48	34	48	33	49	34	52	36	52	34
48	33	50	33	49	35	53	38	51	35
50	35	51	35	49	34	50	34	52	35
50	35	52	36	51	35	53	39	48	33

TABLA 2, GRAFICA DE CONTEO CORRESPONDIENTE A LA MUESTRA DE LA TABLA 1.

Circunferencia de la cabeza y (en cm)	Estatura x (en cm)									
	47	48	49	50	51	52	53	54	55	56
39							I			I
38							I	I		
37					I	I	I	I		
36				I	IIII	卌 II	II			
35	III		卌	卌 IIII	卌 I	I	I			
34	I	卌 II	卌 卌	卌 IIII	III	III				
33	I	卌 I	卌	IIII			I			
32		I	I	I						

AGRUPAMIENTO DE DATOS POR PAREJAS

CUANDO SE TIENE UNA MUESTRA CON MUCHOS DATOS TOMADOS POR PAREJAS, CORRESPONDIENTES A DOS VARIABLES ALEATORIAS, ES A MENUDO CONVENIENTE AGRUPARLOS POR VALORES O POR INTERVALOS Y LUEGO OBTENER LA *DISTRIBUCION CONJUNTA DE FRECUENCIAS*, DE LA MANERA QUE SE MUESTRA EN EL SIGUIENTE EJEMPLO.

EJEMPLO

EN UN ESTUDIO CON FINES ANTROPOLOGICOS REALIZADO EN UNA MATERNIDAD, SE OBTUVO LA MUESTRA POR PAREJAS, MOSTRADA EN LA TABLA 1, CORRESPONDIENTE A LAS VARIABLES ALEATORIAS

X = ESTATURA

Y = CIRCUNFERENCIA DE LA CABEZA

DE LOS NIÑOS AL NACER.

CALCULAR LA DISTRIBUCION CONJUNTA DE FRECUENCIAS Y DIBUJAR EL HISTOGRAMA CORRESPONDIENTE.

PROCEDIMIENTO:

1. Determinar los valores máximos y mínimos de los datos X y Y.
2. Elaborar la tabla de conteo
3. Elaborar la tabla con la distribución conjunta de frecuencias.

METODO DE MINIMOS CUADRADOS

EL METODO DE MINIMOS CUADRADOS TIENE COMO CRITERIO EL QUE LA SUMA DE LOS CUADRADOS DE LAS DESVIACIONES DE LAS ORDENADAS, y_i, RESPECTO A LA RECTA DE REGRESION, $\tilde{y}_i$, SEA MINIMA, ES DECIR, SE TIENE UN METODO DE OPTIMIZACION EN EL QUE SE PRETENDE QUE

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

$$D = \sum_{i=1}^{n} [y_i - (b + mx_i)]^2$$

$$\frac{\partial D}{\partial b} = 2 \sum_{i=1}^{n} (y_i - b - mx_i)(-1) = 0$$

$$\frac{\partial D}{\partial m} = 2 \sum_{i=1}^{n} (y_i - b - mx_i)(-x_i) = 0$$

CON ESTO SE TIENE UN SISTEMA DE ECUACIONES LINEALES CON DOS INCOGNITAS, b Y m, QUE CONDUCE A

$$m = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{\frac{1}{n}\Sigma x_i y_i - \bar{x}\bar{y}}{S^2(x)} = \frac{\frac{1}{n}\Sigma (x_i - \bar{x})(y_i - \bar{y})}{S^2(x)}$$

$$b = \frac{\Sigma x_i^2 \Sigma y_i - \Sigma x_i \Sigma x_i y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \bar{y} - m\bar{x}$$

ESTA ULTIMA ECUACION INDICA QUE LA RECTA PASA POR EL PUNTO $(\bar{x},\bar{y})$.

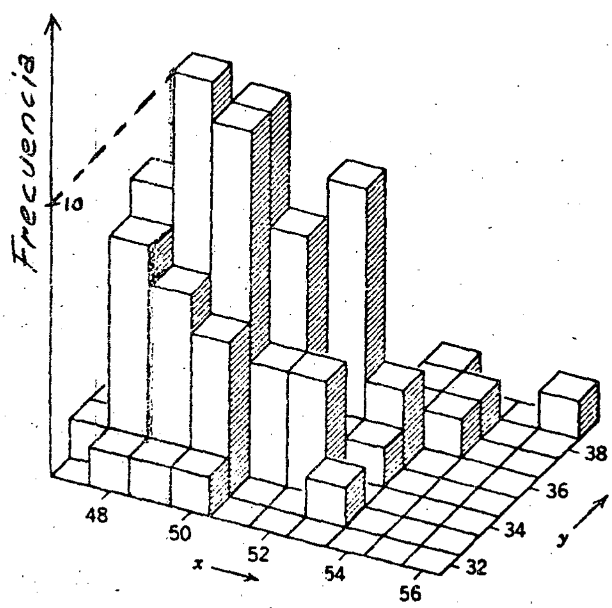
SI LAS PAREJAS DE DATOS ESTAN AGRUPADAS EN K CELDAS Y LA FRECUENCIA DE LA CELDA i ES f_{ixy}, Y x_j Y y_j SON SUS MARCAS DE CLASES, ENTONCES,

$$m = \frac{\frac{1}{n}\sum_{j=1}^{K} f_{jxy} x_j y_j - \bar{x}\bar{y}}{S^2(x)} = \frac{\frac{1}{n}\sum_{j=1}^{K} f_{jxy} (x_j - \bar{x})(y_j - \bar{y})}{S^2(x)}$$

TABLA 3. *DISTRIBUCION CONJUNTA DE FRECUENCIAS DE LA MUESTRA DE LA TABLA 11*

Circunferencia de la cabeza y (en cm)	Estatura x (en cm)									
	47	48	49	50	51	52	53	54	55	56
39							1			1
38							1	1		
37					1	1	1	1		
36				1	4	7	2			
35	3		5	9	6	1	1			
34	1	7	10	9	3	3				
33	1	6	5	4			1			
32		1	1	1						

HISTOGRAMA CORRESPONDIENTE A LA TABLA 1



ADEMAS, TOMANDO EN CUENTA QUE

$$S^2(x) = C_2^2 S^2(x')$$

LA FORMULA PARA CALCULAR LA PENDIENTE CAMBIA A

$$m = \frac{C_4 C_2 \sum_{j=1}^{K} \frac{1}{n} f x_j' y_j' - C_2 C_4 \bar{x}' \bar{y}'}{C_2^2 S^2(x')} = \frac{C_4}{C_2} \frac{(\frac{1}{n}) \sum_{j=1}^{K} f_{jxy} x_j' y_j' - \bar{x}' \bar{y}'}{S^2(x')}$$

EN ESTAS TRANSFORMACIONES C_1 Y C_3 DEBEN SER IGUALES A ALGUNA DE LAS MARCAS DE CLASE CENTRALES DE x Y y, RESPECTIVAMENTE, Y C_2 Y C_4 DEBEN SER IGUALES A LOS ANCHOS DE LOS INTERVALOS DE LOS DATOS DE x Y de y, RESPECTIVAMENTE.

METODO CORTO PARA CALCULAR LA RECTA DE REGRESION

A MENUDO SE PRESENTAN PROBLEMAS DE REGRESION LINEAL EN LOS QUE SE MANEJAN GRANDES CANTIDADES DE DATOS Y, ADEMAS, SUS VALORES SON DE VARIAS CIFRAS. PARA REDUCIR LA LABOR NUMERICA SE RECURRE A AGRUPAR LOS DATOS Y A TRANSFORMAR LAS VARIABLES DE LA MANERA SIGUIENTE:

$$x' = \frac{x - C_1}{C_2} \qquad y' = \frac{y - C_3}{C_4} \; ; \; C_2 > 0, \; C_4 > 0$$

DE DONDE $\quad x = C_2x' + C_1 \qquad y = C_4y' + C_3$

EN TAL CASO, EL PRIMER TERMINO DEL NUMERADOR DE LA FORMULA PARA CALCULAR m SE TRANSFORMA A:

$$\frac{1}{n}\sum_{j=1}^{K} f_{jxy}x_jy_j = \frac{1}{n}\sum_{j=1}^{K} f_{jxy}(C_2x_j' + C_1)(C_4y_j' + C_3)$$

$$= \frac{1}{n}\sum_{j=1}^{K} f_{jxy}(C_2C_4x_j'y_j' + C_2C_3x_j' + C_1C_4y_j' + C_1C_3)$$

$$= C_2C_4\,\frac{1}{n}\sum_{j=1}^{K} f_{jxy}x_j'y_j' + C_2C_3\sum_{j=1}^{K}\frac{1}{n}f_{jxy}x_j' + C_1C_4\sum_{j=1}^{K}\frac{1}{n}f_{ixy}y_j' + C_1C_3$$

$$= C_2C_4\,\frac{1}{n}\sum_{j=1}^{K} f_{jxy}x'y' + C_2C_3\overline{x'} + C_1C_4\overline{y'} + C_1C_3$$

EL SEGUNDO TERMINO DE LA MISMA FORMULA QUEDA:

$$\bar{x}\bar{y} = (C_2\bar{x}' + C_1)(C_4\bar{y}' + C_3) = C_2C_4\bar{x}'\bar{y}' + C_2C_3\bar{x}' + C_1C_4\bar{y}' + C_1C_3$$

EJEMPLO

OBTENER LA RECTA DE REGRESION DE LAS CARGAS EN LOS PISOS 1 Y 9 DE UN EDIFICIO

CARGAS EN TON/M^2

ZONA	PISO 1 x	PISO 9 y
A	38	355
B	354	370
C	207	307
D	273	270
E	127	182
F	324	962
G	358	222
H	519	405
I	147	315
J	181	420
K	118	484
L	114	287
M	243	228
N	522	470
O	236	194
P	269	260
Q	268	679
R	321	366
S	305	358
T	335	317
U	577	368
V	271	284

RANGO DE X: 577-38

RANGO DE Y: 962-182

EJEMPLO

CALCULAR LA RECTA DE REGRESION DE LOS DATOS ANOTADOS EN LA SIGUIENTE TABLA, MEDIANTE EL METODO DE MINIMOS CUADRADOS.

x	y	xy	x^2
8	7	56	64
6	12	72	36
4	2	8	16
6	6	36	36
13	7	91	169
10	3	30	100
1	6	6	1
7	2	14	49
3	9	27	9
12	11	132	144
Σ= 70	65	472	624

$\bar{x} = 70/10=7$, $\bar{y} = 65/10=6.5$, $\overline{x^2}=624/10=62.4$

$$S^2(x) = 62.4 - 7^2 = 13.4$$

$$m = \frac{\frac{1}{10}\,472 - 7 \times 6.5}{13.4} = 0.13$$

$$b = 6{,}5 - 0.13 \times 7 = 5{,}59$$

CALCULO DE $\bar{x}$ y $\overline{x^2}$

INTERVALOS	MARCAS DE CLASE	f	xf	x^2	$x^2 f$
0.5 -100.5	50.5	1	50.5	2,550.25	2,550.25
100.5-200.5	150.5	5	752.50	22,650.25	113,251.25
200.5-300.5	250.5	7	1,753.50	62,750.25	439,251.25
300.5-400.5	350.5	6	2,103.00	122,850.25	737,101.50
400.5-500.5	450.5	0	0.00	164,430.25	0.00
500.5-600.5	550.5	3	1,651.50	303,050.25	909,150.75
		Σ=22	Σ=6,311.00		Σ=2,201.304.50

$$\bar{x} = \frac{6,311.00}{22} = 286.86, \quad \bar{x}^2 = 82,288.66$$

$$\overline{x^2} = \frac{2,201.304.50}{22} = 100,059.30$$

$$s^2(x) = 100,059.30 - 82,288.66 = 17,770.64$$

$$S(x) = \sqrt{17,770.64} = 133.31$$

TABLA 4. DISTRIBUCION CONJUNTA DE FRECUENCIAS DE LAS CARGAS EN LOS PISOS 1 Y 9.

X \ Y	0.5 a 100.5	100.5 a 200.5	200.5 a 300.5	300.5 a 400.5	400.5 a 500.5	500.5 a 600.5	600.5 a 700.5	700.5 a 800.5	800.5 a 900.5	900.5 a 1 000.5
0.5 a 100.5				X (1)						
100.5 a 200.5		X (1)	X (1)	X (1)	XX (2)					
200.5 a 300.5		X (1)	XXXX (4)	X (1)			X (1)			
300.5 a 400.5			X (1)	XXXX (4)						X (1)
400.5 a 500.5										
500.5 a 600.5				X (1)	XX (2)					

MARCAS DE CLASE x	y	FRECUENCIAS f_{xy}	xy	$f_{xy}xy$
50.5	350.5	1	17,700.25	17,700.25
150.5	150.5	1	22,650.25	22,650.25
150.5	250.5	1	37,700.25	37,700.25
150.5	350.5	1	52,750.25	52,750.25
150.5	450.5	2	67,800.25	135,600.50
250.5	150.5	1	37,700.25	37,700.25
250.5	250.5	4	62,750.25	251,001.00
250.5	350.5	1	87,800.25	87,800.25
250.5	650.5	1	162,950.25	162,950.25
350.5	250.5	1	87,800.25	87,800.25
350.5	350.5	4	122,850.25	491,401.00
350.5	950.5	1	333,150.25	333,150.25
550.5	350.5	1	192,950.25	192,950.25
550.5	450.5	2	248,000.25	496,000.50
		$\Sigma = 22$		$\Sigma = 2,407,155.50$

PUESTO QUE $\bar{x} = 286.86$, $\bar{y} = 364.14$ Y $S^2(x) = 17,770.64$

SE OBTIENE FINALMENTE QUE

$$m = \frac{\frac{2,407,155.50}{22} - (286.86)(364.14)}{17,770.64} = 0.28$$

$$b = \bar{y} - m\bar{x} = 364.14 - 0.28 \times 286.86 = 283.82$$

$$\tilde{Y} = 0.28\,x + 283.82$$

CALCULO DE $\bar{y}$ y $\overline{y^2}$

INTERVALOS	MARCAS DE CLASE, y	f	yf	y^2	y^2f
100.5-200.5	150.5	2	301.00	22,650.25	45,300.50
200.5-300.5	250.5	6	1,503.00	62,750.25	376,501.50
300.5-400.5	350.5	8	2,804.00	122,850.25	982,802.00
400.5-500.5	450.5	4	1,802.00	202,950.25	811,801.00
500.5-600.5	550.5	0	0.00	303,050.25	0.00
600.5-700.5	650.5	1	650.50	423,150.25	423,150.25
700.5-800.5	750.5	0	0.00	563,250.25	0.00
800.5-900.5	850.5	0	0.00	723,350.25	0.00
900.5-1000.5	950.5	1	950.50	903,450.25	903,450.25
		Σ=22	Σ=8,011.00		Σ=3,543,005.50

$$\bar{y} = \frac{8,011.00}{22} = 364.14 \;, \quad \bar{Y}^2 = 132,597.94$$

$$\overline{Y^2} = \frac{3,543,005.50}{22} = 161,045.70$$

$$S^2(y) = 161,045.70 - 132,597.94 = 28,447.76$$

$$S(y) = \sqrt{28,447.76} = 168.66$$

TAREA: CALCULAR $\bar{x}$, $S^2(x)$, $\bar{y}$ y $S^2(y)$ DE LOS DATOS AGRUPADOS ANTERIORES, MEDIANTE TRANSFORMACIONES APROPIADAS DE VARIABLES.

Y \ X	0.5-100.5 50.5			100.5-200.5 150.5			200.5-300.5 250.5			300.5-400.5 350.5			500.5-600.5 550.5			f_y	y'	$f_y y'$	y'^2	$f_y y'^2$	$\Sigma f_{jxy} x'y'$
100.5-200.5 150.5				2	1	2	0	1	0							2	-2	-4	4	8	2
200.5-300.5 250.5				1	1	1	0	4	0	-1	1	-1				6	-1	-6	1	6	0
300.5-400.5 350.5	0	1	0	0	1	0	0	1	0	0	4	0	0	1	0	8	0	0	0	0	0
400.5-500.5 450.5				-1	2	-2							3	2	6	4	1	4	1	4	4
600.5-700.5 650.5							0	1	0							1	3	3	9	9	0
900.5-1000.5 950.5										6	1	6				1	6	6	36	36	6
f_x		1			5			7			6			3		22		3		63	12
x'		-2			-1			0			1			3							
$f_x x'$		-2			-5			0			6			9		8					
x'^2		4			1			0			1			9							
$f_x x'^2$		4			5			0			6			27		42					
$\Sigma f_{jxy} x'y'$			0			1			0			5			6	12					

EJEMPLO

RESOLVER EL PROBLEMA ANTERIOR MEDIANTE EL METODO CORTO.

PARA APLICAR EL METODO CORTO SE EMPLEA UNA TABULACION COMO LA SIGUIENTE:

Y \ X	DISTRIBUCION CONJUNTA DE FRECUENCIAS	f_y	y'	$f_y y'$	y'^2	$f_y y'^2$	$\Sigma f_{jxy} x'y'$
	CELDA j						
f_x							
x'							
$f_x x'$							
x'^2							
$f_x x'^2$							
$\Sigma f_{jxy} x'y'$							

$x'y'$	f_{jxy}	$f_{jxy} x'y'$	← CELDA j

PARA LA TRANSFORMACION DE VARIABLES

$$X' = \frac{X - C_1}{C_2} \quad Y \quad Y' = \frac{Y - C_3}{C_4}$$

TOMAREMOS $C_1 = 250.5$, $C_2 = 100$, $C_3 = 350.5$ Y $C_4 = 100$.

VARIANCIA Y ERROR ESTANDAR DE LA ESTIMACION MEDIANTE LE RECTA DE REGRESION

COMO YA SE INDICO, EL TERMINO $y_i - \tilde{y}_i$ REPRESENTA LA DIFERENCIA ENTRE EL VALOR OBSERVADO DE LA VARIABLE Y Y EL VALOR PREDICHO (LA ORDENADA DE LA RECTA DE REGRESION) CORRESPONDIENTE A x_i. DICHO TERMINO SE LLAMA *ERROR DE PREDICCION O DE ESTIMACION*. POR EJEMPLO, SI PARA $x_3=50$ SE OBSERVA QUE $y_3=65$, Y LA ECUACION DE LA RECTA DE REGRESION ES $\tilde{y}=.70\ x + 21.9$, EL VALOR PREDICHO RESULTA $\tilde{y}_3 = 0.70 \times 50 + 21.9=56.9$, Y EL ERROR DE PREDICCION CORRESPONDIENTE ES $65 - 56.9 = 8.1$.

LA *VARIANCIA DE LA PREDICCION O DE LA ESTIMACION*, $s^2_{y|x}$, QUE ES UNA ESTIMACION GLOBAL DEL ERROR DE PREDICCION PARA TODOS LOS PUNTOS OBSERVADOS, SE DEFINE MEDIANTE LA FORMULA

$$s^2_{y|x} = \frac{\sum_{i=1}^{i=N} (y_i - \tilde{y}_i)^2}{N} \qquad \text{(I)}$$

EN DONDE N ES EL TOTAL DE DATOS DE Y. ADEMAS, SE PUEDE DEMOSTRAR QUE $s^2_{y|x}$ SE RELACIONA CON LA PENDIENTE DE LA RECTA DE REGRESION MEDIANTE LA ECUACION

$$s^2_{y|x} = s^2_y - m^2\ s^2(x)$$

PUESTO QUE LA ECUACION $y = mx + b$ SE OBTIENE MEDIANTE EL METODO DE MINIMOS CUADRADOS, EN EL CUAL $\sum_{i=1}^{i=N} (y_i - \tilde{y}_i)^2$ TIENE EL MINIMO VALOR POSIBLE, Y COMO LA VARIANCIA DE LA PREDICCION SE CALCULA CON LA EC (I), LAS PREDICCIONES BASADAS EN LA RECTA DE MINIMOS CUADRADOS SON TALES QUE LA VARIANCIA DE LA PREDICCION ES MINIMA.

$\overline{x'} = \frac{8}{22} = 0.3636$; $\overline{Y'} = \frac{3}{22} = 0.1364$; $\overline{x'^2} = \frac{42}{22} = 1.9091$, $\overline{y'^2} = \frac{63}{22} = 2.8636$

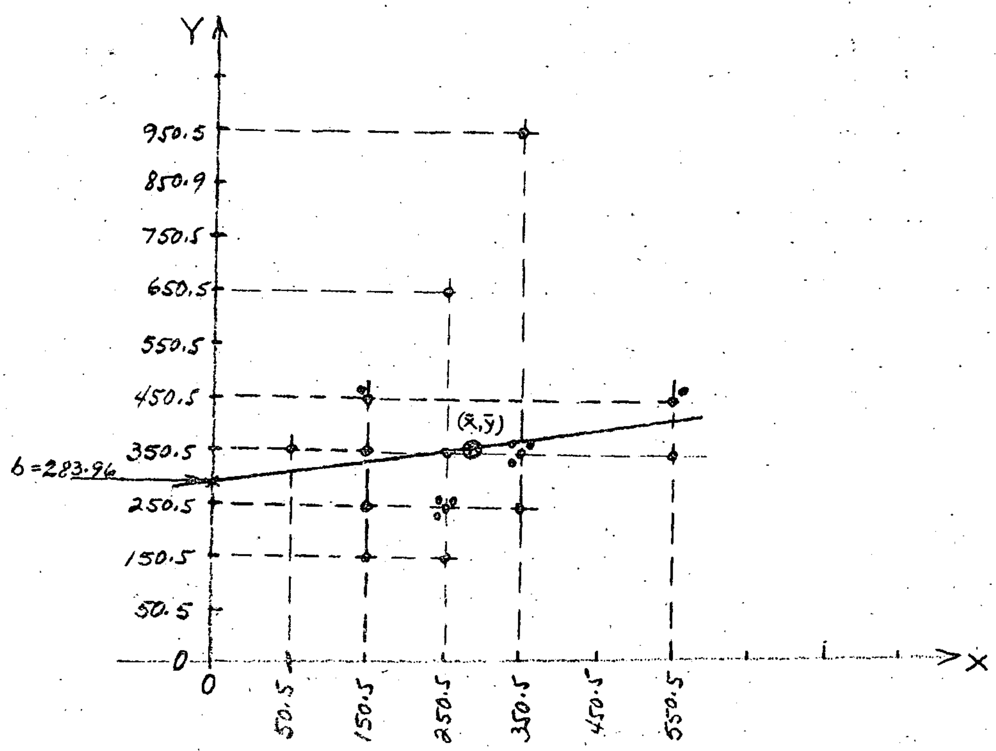
$s^2(x') = 1.9091-(0.3636)^2 = 1.7769$; $s^2(y') = 2.8636-(0.1364)^2 = 2.8450$

$\bar{x} = C_2\overline{x'} + C_1 = 36.36 + 250.5 = 286.86$

$\bar{y} = C_4\overline{y'} + C_3 = 13.64 + 350.5 = 364.14$

$$m = \frac{100}{100} \frac{\frac{1}{22} 12 - (0.3636)(0.1364)}{1.7769} = \frac{0.4959}{1.7769} = 0.28$$

$b = \bar{y} - m\bar{x} = 364.14 - 0.28 \times 286.36 = 283.96$



TAREA: CALCULAR LA RECTA DE REGRESION, Y TRAZAR LA GRAFICA CORRESPONDIENTE, DE LOS DATOS AGRUPADOS DE LA TABLA 3.

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

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

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

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

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

$$= NS^2(y) - 2mN\,S^2_{xy} + NS^2(x)m^2$$

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

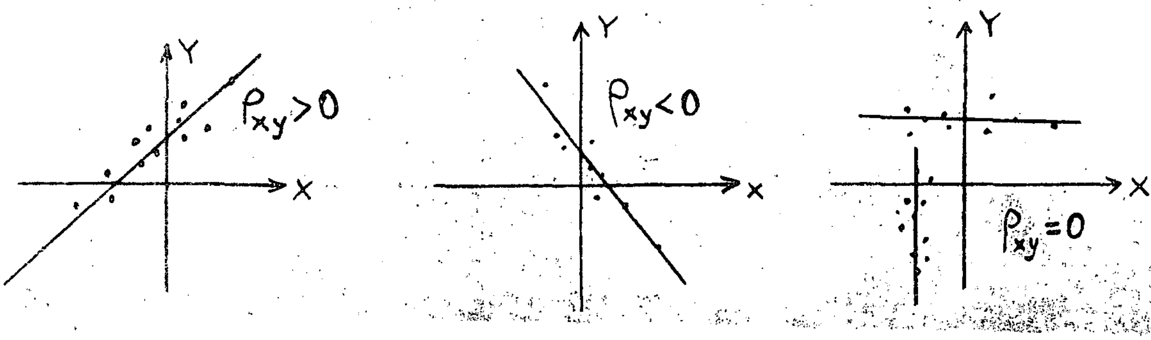
$$\rho_{xy} = \frac{S^2 xy}{S(x)\ S(y)}; \quad -1 \leq \rho_{xy} \leq 1$$

EN DONDE

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

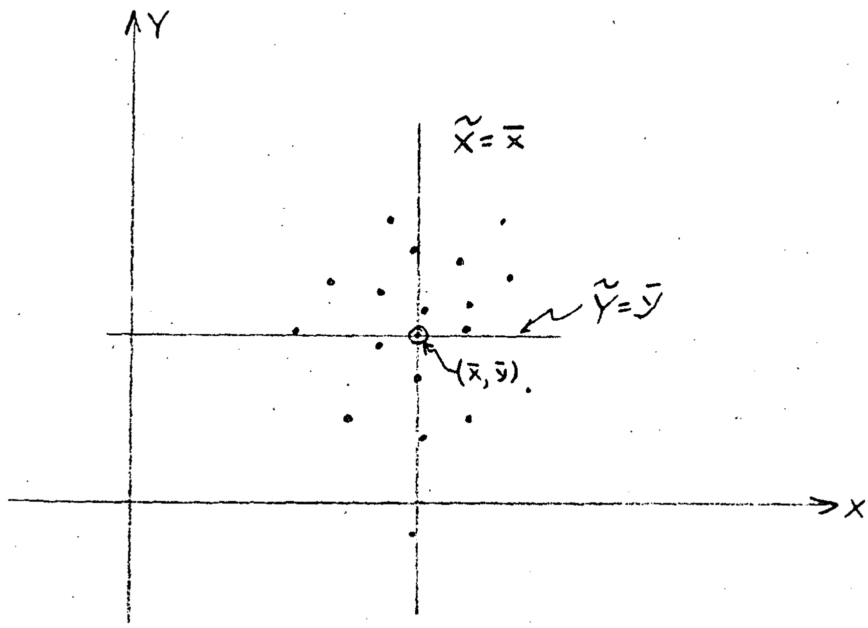
S(x) = DESVIACION ESTANDAR DE LOS DATOS DE X

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



CASO DE CORRELACION NULA

LA CORRELACION ENTRE LOS DATOS DE DOS VARIABLES ALEATORIAS RESULTA NULA SI $S_{xy} = 0$ (O $\rho_{xy} = 0$) LO CUAL SUCEDE CUANDO m=0 (VER EC (5)). EN TAL CASO, LA RECTA DE REGRESION DE Y CON BASE EN X TIENE COMO ECUACION A $\tilde{y} = \bar{y}$, Y LA DE X CON BASE EN Y, A $\tilde{X} = \bar{X}$ (m=∞).



EN EL CASO DE QUE TODOS LOS PUNTOS DE LA MUESTRA QUEDEN SOBRE LA RECTA DE REGRESION SE TIENE QUE

$$D = NS^2(y) = 2m\ NS^2_{xy} + m^2\ NS^2(x) = 0 \qquad (4)$$

POR OTRA PARTE, TOMANDO EN CUENTA QUE

$$m = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{S^2(x)} = \frac{S^2_{xy}}{S^2(x)} \qquad (5)$$

LA EC (4) QUEDA EN LA FORMA

$$S^2(y) = 2(S^2_{xy})^2/\ S^2(x) + (S^2_{xy})^2/\ S^2(x) = 0$$

DE DONDE, EN EL CASO DE REGRESION PERFECTA,

$$(S^2_{xy})^2 = S^2(x)\ S^2(y) \qquad (6)$$

Y, SI $S(x)>0$ Y $S(y)>0$,

$$\rho^2_{xy} = \frac{(S^2_{xy})^2}{S^2(x)S(y)^2} = \frac{S^2(x)S^2(y)}{S^2(x)S^2(y)} = 1, \text{ O SEA, } \rho_{xy} = \pm 1$$

CUANDO ESTO SUCEDE, ES DECIR, SI $\rho_{xy} = 1$ Ó $\rho_{xy} = -1$, SE TIENE EL CASO DE *CORRELACION PERFECTA*.

EN DONDE f_{ixy} ES LA FRECUENCIA DE LA CELDA i, x' Y y' SON LAS MARCAS DE CLASE DE LOS INTERVALOS, $\overline{x'}$ Y $\overline{y'}$ SON LOS PROMEDIOS ARITMETICOS, Y S(x') Y S(y') LAS DESVIACIONES ESTANDAR DE LOS DATOS DE X' Y Y' OBTENIDOS MEDIANTE LAS TRANSFORMACIONES

$$x' = \frac{X - C_1}{C_2} \quad \text{Y} \quad y' = \frac{y - C_3}{C_4}$$

EN DONDE

C_1 = MARCA DE CLASE CENTRAL DE LAS x

C_2 = ANCHO DE LOS INTERVALOS DE CLASE DE LAS x

C_3 = MARCA DE CLASE CENTRAL DE LAS y

C_4 = ANCHO DE LOS INTERVALOS DE CLASE DE LAS y

RELACION ENTRE EL COEFICIENTE DE CORRELACION Y LA PENDIENTE DE LA RECTA DE REGRESION

TOMANDO EN CUENTA QUE

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

Y HACIENDO SUSTITUCIONES EN LA ECUACION PARA CALCULAR LA PENDIENTE DE LA RECTA DE REGRESION SE OBTIENE

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

O SEA

$$m = \rho_{xy}\,\frac{S(y)}{S(x)} \qquad (8)$$

DE ESTA MANERA, SI CALCULAMOS m MEDIANTE EL METODO CORTO DESCRITO ANTERIORMENTE, PODEMOS CALCULAR ρ_{xy} DESPEJANDOLA DE LA EC (8), ES DECIR, EMPLEANDO LA ECUACION

$$\rho_{xy} = m\,\frac{S(x)}{S(y)} \qquad (9)$$

ALTERNATIVAMENTE, MEDIANTE EL METODO CORTO SE OBTIENE ρ_{xy} EN FORMA DIRECTA USANDO LA ECUACION

$$\rho_{xy} = \frac{\frac{1}{N}\sum_{i=1}^{N} f_{ixy}x'y' - \overline{x'}\;\overline{y'}}{S(x')\;S(y')} \qquad (10)$$

EJEMPLO

DIEZ VIGAS DE MADERA NOMINALMENTE IDENTICAS SE PROBARON CON UNA CARGA CONCENTRADA EN EL CENTRO DEL CLARO; LOS RESULTADOS SON LOS ANOTADOS EN LA TABLA SIGUIENTE. CALCULAR EL COEFICIENTE DE CORRELACION, LA RECTA DE REGRESION Y LAS VARIANCIAS EXPLICADA E INEXPLICADA.

CARGA DE FALLA x, EN KG	DE FLEXION MAXIMA, y, EN CM	$x - \bar{x}$	$y - \bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
950	0.33	140	0.017	2.38	19,600	0.000289
1050	0.37	240	0.057	13.68	57,600	0.003249
750	0.28	- 60	-0.033	1.98	3,600	0.001089
900	0.30	90	-0.013	- 1.17	8,100	0.000169
700	0.27	-110	-0.043	4.73	12,100	0.001849
650	0.28	-160	-0.033	5.28	25,600	0.001089
950	0.35	140	0.037	5.18	19,600	0.001369
850	0.40	40	0.087	3.48	1,600	0.007569
600	0.26	-210	-0.053	11.13	44,100	0.002809
700	0.29	-110	-0.023	2.53	12,100	0.000529
Σ= 8100	Σ = 3.13	Σ = 0	Σ= 0	Σ=49.20	Σ=204,000	0.020010

$\bar{x} = \frac{8100}{10} = 810$ KG; $\bar{y} = \frac{3.13}{10} = 0.313$ CM; $S^2_{xy} = \frac{49.20}{10} = 4.92$

$S^2(x) = \frac{204,000}{10} = 20,400; \quad S(x) = \sqrt{20,400} = 142,83$

$S^2(y) = \frac{0.020010}{10} = 0.002001; \quad S(y) = \sqrt{0.002001} = 0.0447$

$\rho_{xy} = \frac{4.92}{142.83 \times 0.0447} = 0.77$

$m = \rho_{xy}\ S(y)/S(x) = 0.77 \times 0.0447/142.83 = 0.000241$ CM/KG

RANGO DELCOEFICIENTE DE CORRELACION

DE LA ECUACION CON QUE SE CALCULA LA VARIANCIA DE LA ESTIMACION

$$S^2_{y|x} = S^2(y)(1 - \rho^2_{xy}) \tag{7}$$

SE CONCLUYE QUE $\rho^2_{xy} \leq 1$, YA QUE $S^2_{y|x} \geq 0$; EN CONSECUENCIA

$$-1 \leq \rho_{xy} \leq 1 \tag{8}$$

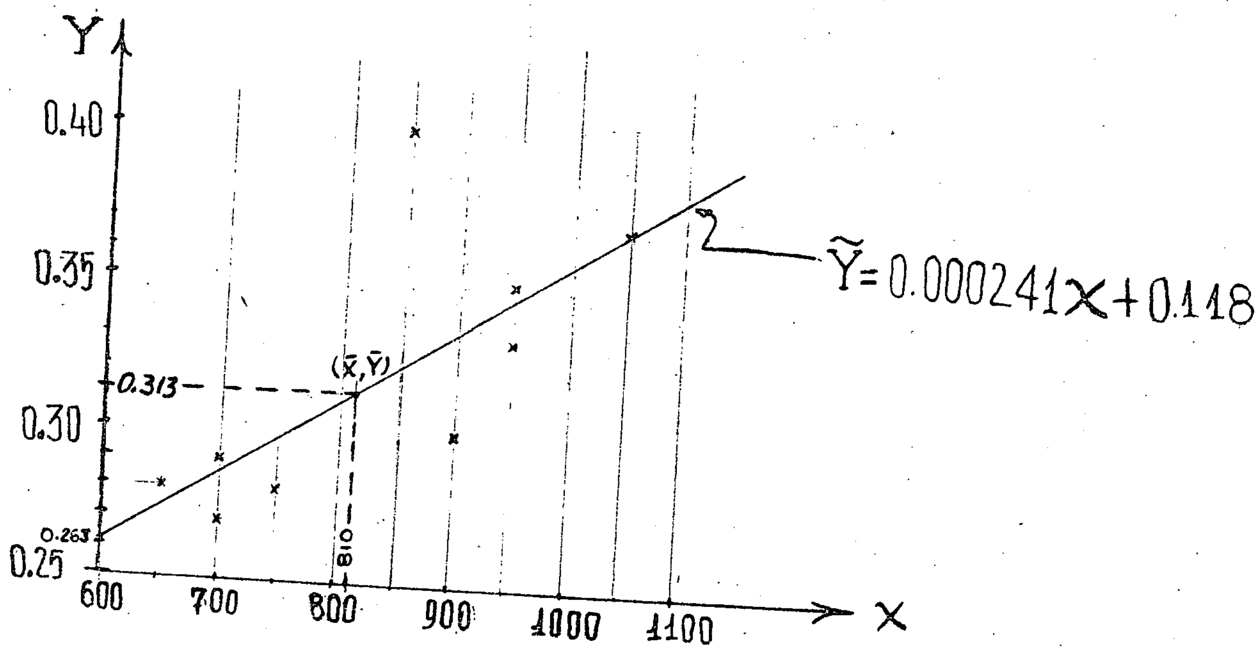
EJEMPLO

CALCULAR EL COEFICIENTE DE CORRELACION MEDIANTE EL METODO CORTO DE LOS DATOS LA SIGUIENTE TABLA. OBTENER TAMBIEN LA ECUACION DE LA RECTA DE REGRESION CORRESPONDIENTE.

CALIFICACION, X	TIEMPO, Y, EN MIN.	CALIFICACION X	TIEMPO, Y, EN MIN.
97	77	87	83
97	86	87	58
95	60	87	79
95	52	86	60
94	62	85	62
94	86	83	72
94	80	82	68
93	79	82	66
93	92	83	71
93	88	81	70
92	74	80	65
92	43	80	84
92	61	79	82
92	75	79	93
91	79	79	76
90	62	78	71
90	81	77	89
90	80	77	71
90	76	77	98
90	70	76	92
89	67	76	82
88	69	74	98
88	81	72	78
88	80	79	93
87	91	70	78

$b = 0.313 - 0.000241 \times 810 = 0.118$ CM

$\tilde{y} = 0.000241\,x + 0.118$; SI $X = 600$, $\tilde{y} = 0.145 + 0.118 = 0.263$



$s^2_{y|x} = s^2(y)(1 - \rho^2_{xy}) = 0.002001\ (1 - 0.77)^2 = 0.000106$

$s_{y|x} = 0.0103$

$s^2(y) = s^2(\tilde{y}) + s^2_{y|x} \rightarrow s^2(\tilde{y}) = 0.002001 - 0.000106 = 0.001895$

$s(\tilde{y}) = 0.0435$

y \ x	72.5			78.5			84.5			90.5			96.5			f_y	y'	$f_y y'$	y'^2	$f_y y'^2$	$\Sigma f_{ixy} x'y'$
45.5										-2	1	-2				1	-2	-2	4	4	-2
55.5							0	2	0				-2	2	-4	4	-1	-4	1	4	-4
65.5				0	2	0	0	3	0	0	5	0	0	1	0	11	0	0	0	0	0
75.5	-2	2	-4	-1	3	-3	0	3	0	1	7	7	2	2	4	17	1	17	1	17	4
85.5				-2	4	-8	0	1	0	2	3	6	4	2	8	10	2	20	4	40	6
95.5	-6	2	-12	-3	3	-9	0	1	0	3	1	3				7	3	21	9	63	-18
f_x	4			12			10			17			7			50		52		128	-14
x'	-2			-1			0			1			2								
$f_x x'$	-8			-12			0			17			14			11					
x'^2	4			1			0			1			4								
$f_x x'^2$	16			12			0			17			28			73					
$\Sigma f_{ixy} x'y'$	-16			-20			0			14			8			-14					

$C_1 = 84.5;\ C_2 = 6;\ C_3 = 65.5;\ C_4 = 10$

$\bar{x}' = 11/50 = 0.22;\ \bar{y}' = 52/50 = 1.04$

$\bar{x} = 6 \times 0.22 + 84.5 = 85.82;$

$\bar{y} = 10 \times 1.04 + 65.5 = 75.9$

$S^2(x') = 73/50 - (0.22)^2 = 1.42;\quad S(x') = \sqrt{1.42} = 1.19$

$S^2(y') = 128/50 - (1.04)^2 = 1.49;\quad S(y') = \sqrt{1.49} = 1.22$

$S(x) = 1.19 \times 6 = 7.14;\ S(y) = 1.22 \times 10 = 12.2$

$$\rho_{xy} = \frac{-14/50 - 0.22 \times 1.04}{1.19 \times 1.22} = -0.35$$

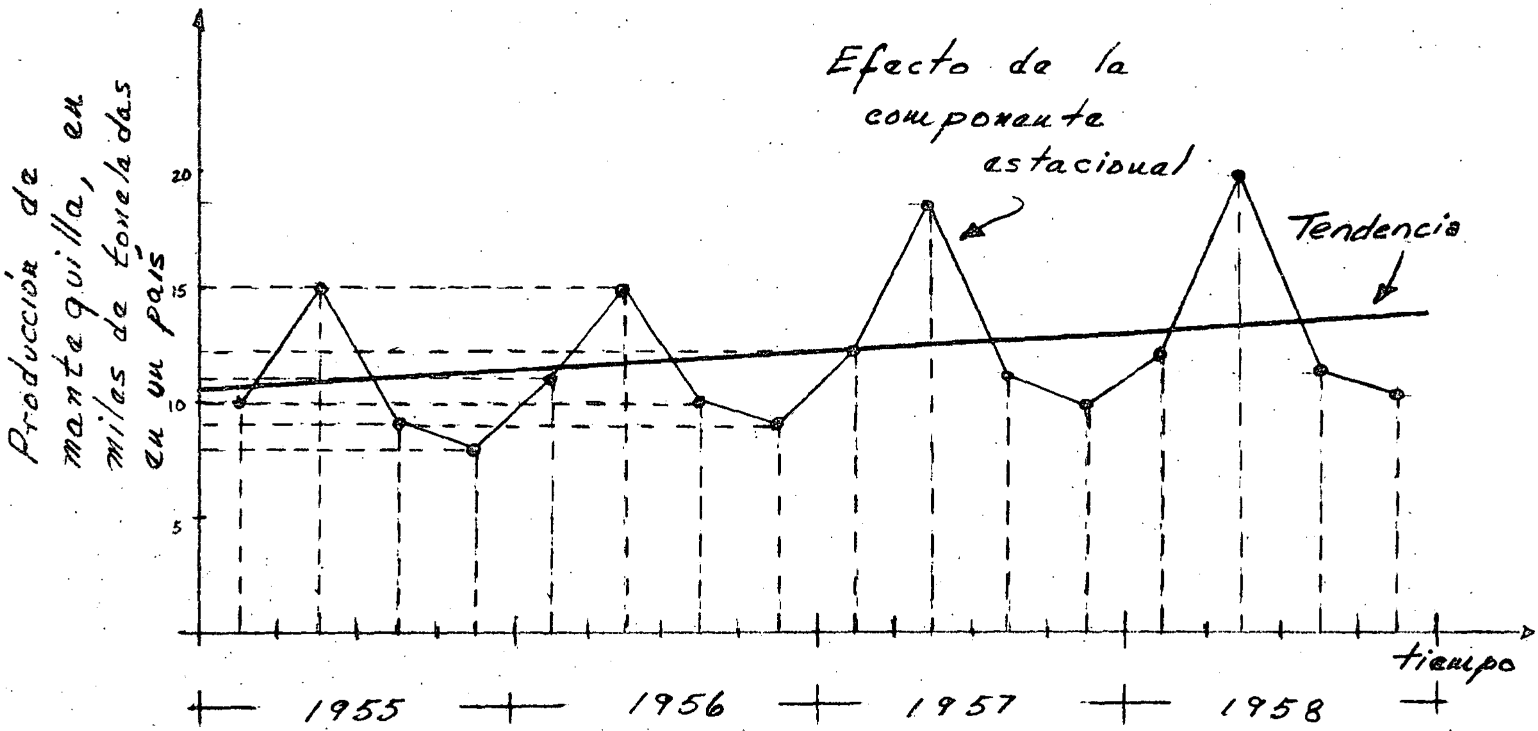
$m = -0.35 \times 12.2 / 7.14 = -0.60$

$b = 75.9 - (-0.60)\ 85.82 = 127.39$

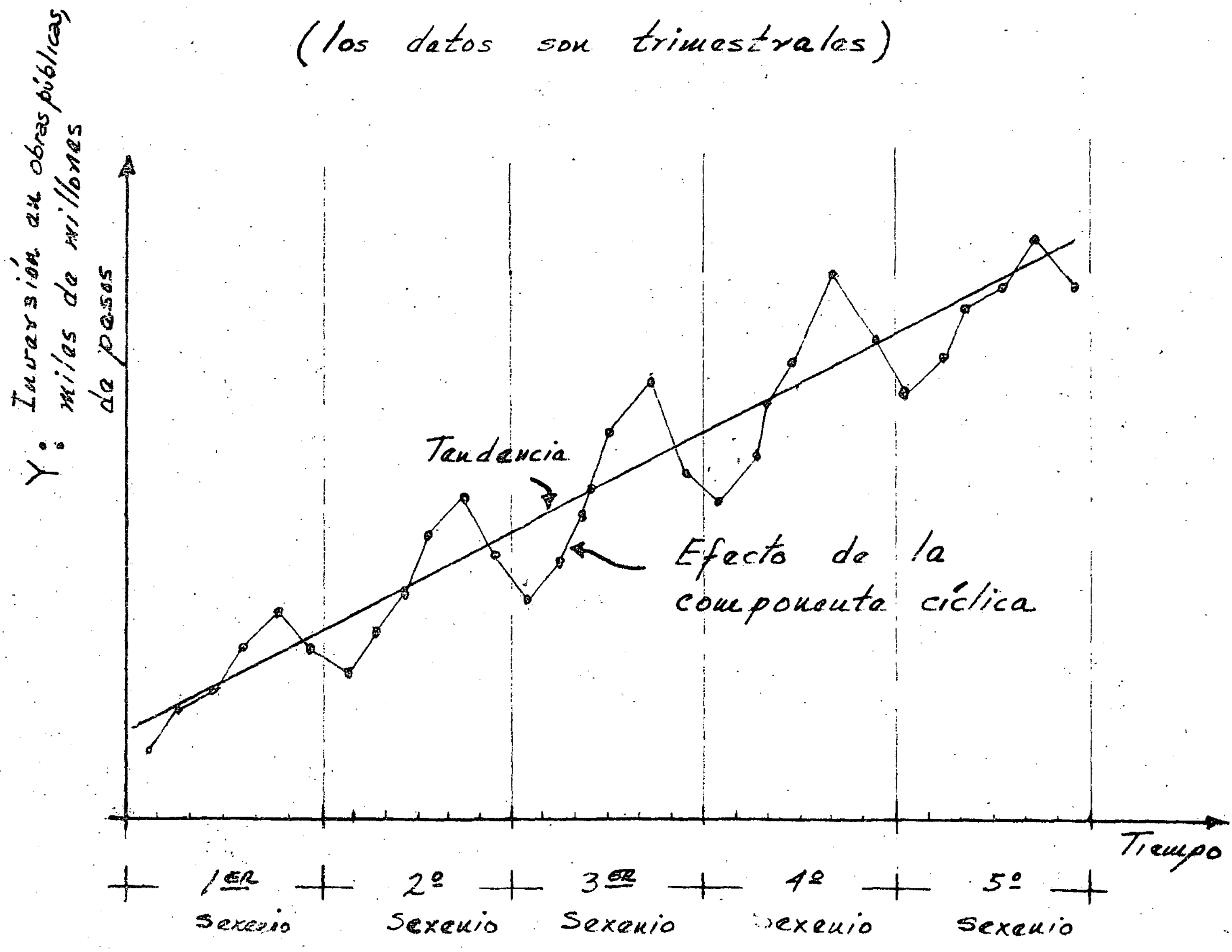
TAREA: TRAZAR EL DIAGRAMA DE CORRELACION Y LA RECTA DE REGRESION CORRESPONDIENTE, Y CALCULAR EL ERROR ESTANDAR DE LA ESTIMACION.

DISTRIBUCION CONJUNTA DE FRECUENCIAS:

Y \ X	70-75	76-81	82-87	88-93	94-99
41-50				1	
51-60			2		2
61-70		2	3	5	1
71-80	2	3	3	7	2
81-90		4	1	3	2
91-100	2	3	1	1	



(los datos son trimestrales)

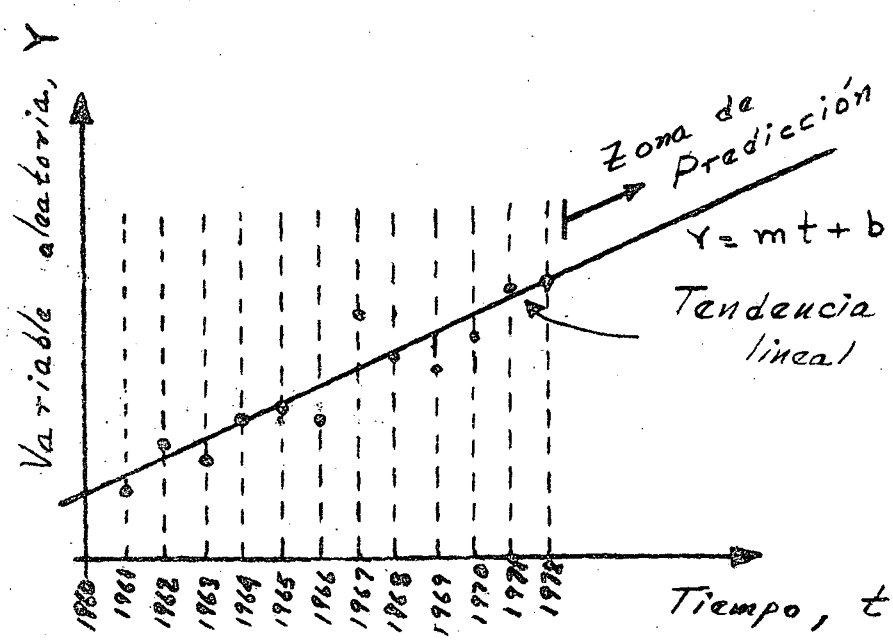


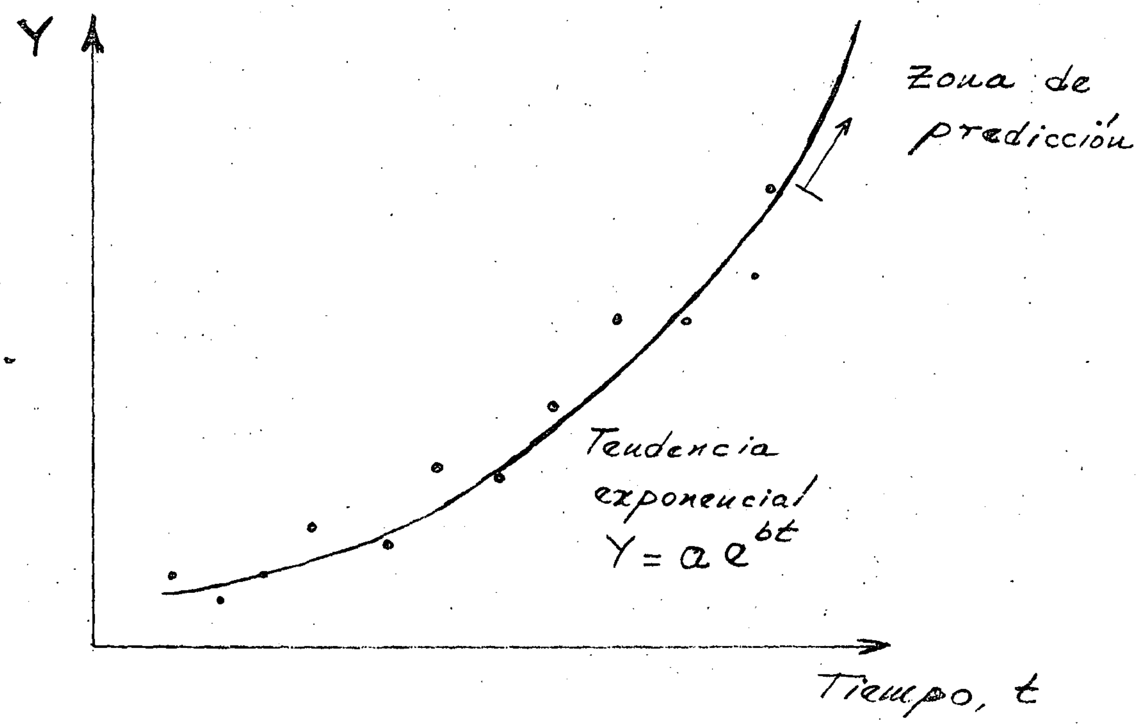
SERIES CRONOLOGICAS O SERIES DE TIEMPO

Se le denomina serie cronológica o de tiempo a toda serie de observaciones (datos) tomados en tiempos específicos, que en general están igualmente espaciados (cada hora, cada semana, cada mes, cada año, etc.)

Componentes de una serie cronológica:

- Tendencia general.- Indica hacia dónde tiende la serie cronológica
- Componente estacional.- Indica las variaciones periódicas que ocurren a corto plazo (en periodos menores de un año)
- Componente cíclica.- Indica las variaciones periódicas que ocurren a largo plazo (en periodos mayores de un año)
- Componente irregular.- Indica las variaciones que ocurren al azar.





TENDENCIA GENERAL

Métodos de cálculo $\begin{cases} \text{Mínimos cuadrados} \\ \text{Dos promedios} \\ \text{Promedios móviles} \end{cases}$

MINIMOS CUADRADOS

El método de mínimos cuadrados se estudió en el capítulo de regresión lineal para el caso de tendencia modelada mediante una línea recta.

Si la tendencia no se puede modelar razonablemente mediante una recta, es necesario emplear una relación no lineal, que puede ser un polinomio de orden M, dado por

$$\tilde{y}(t) = b_0+b_1t+b_2t^2+\ldots+b_Mt^M$$

En este caso las constantes b_i que hacen mínimo el error cuadrático respecto a la línea de regresión, q, se obtienen de resolver un sistema de ecuaciones simultáneas que resultan de:

$$\frac{\partial q}{\partial b_0} = 0,\ \frac{\partial q}{\partial b_1} = 0,\ \frac{\partial q}{\partial b_2} = 0,\ldots,\ \frac{\partial q}{\partial b_M} = 0$$

en donde

$$q = \sum_{i=1}^{n}(y_i-\tilde{y}_i)^2$$

En el caso de un ajuste parabólico (m = 2), por ejemplo,

$$\tilde{y}(t) = b_0+b_1t+b_2t^2$$

$$\tilde{y}_i = \tilde{y}(t_i) = b_0+b_1t_i+b_2t_i^2$$

$$q = \sum_{i=1}^{n}(y_i-b_0-b_1t_i-b_2t_i^2)^2$$

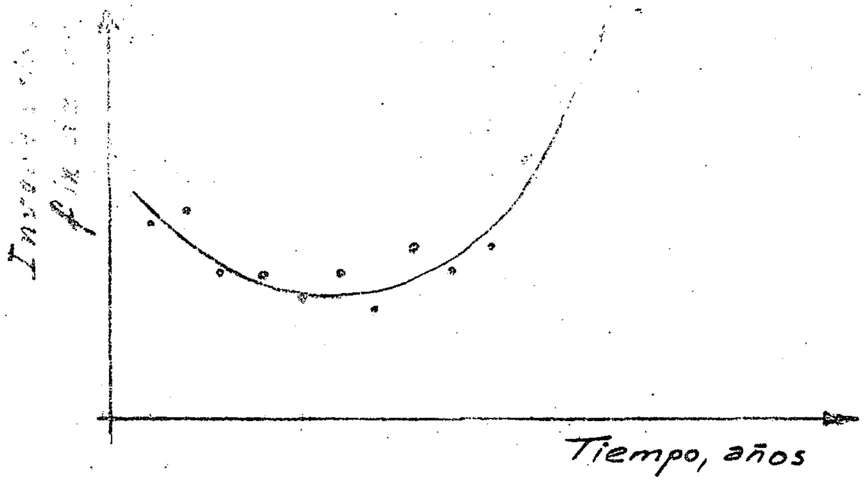
$$\left.\begin{aligned}
\frac{\partial q}{\partial b_0} &= -2\sum_{i=1}^{n}(y_i-b_0-b_1t_i-b_2t_i^2) = 0\\
\frac{\partial q}{\partial b_1} &= -2\sum_{i=1}^{n}(y_i-b_0-b_1t_i-b_2t_i^2)t_i=0\\
\frac{\partial q}{\partial b_2} &= -2\sum_{i=1}^{n}t_i^2(y_i-b_0-b_1t_i-b_2t_i^2) = 0
\end{aligned}\right\}$$

Estas tres últimas ecuaciones constituyen un sistema con tres incógnitas, b_0, b_1 y b_2. Este sistema se puede reescribir en la forma:

$$b_0 n + b_1 \sum t_i + b_2 \sum t_i^2 = \sum y_i$$

[illegible]

[illegible]



Cuando al observar la gráfica de Y contra t se concluye que es razonable ajustar una función exponencial de la forma

$$\tilde{Y}(t) = ae^{mt}$$

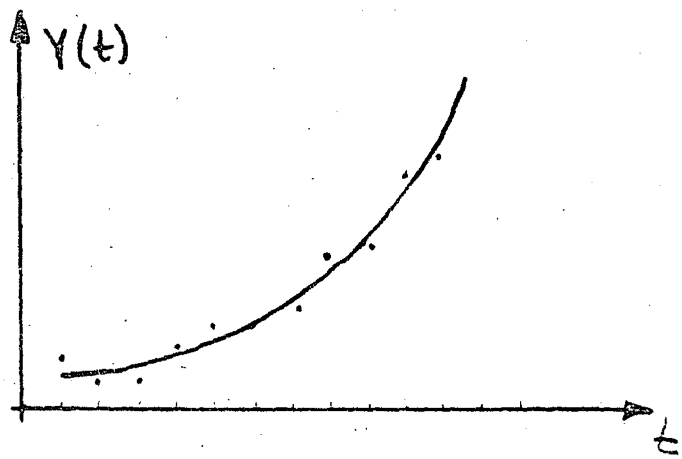
se puede resolver el problema trabajando con logaritmos, ya que, en tal caso,

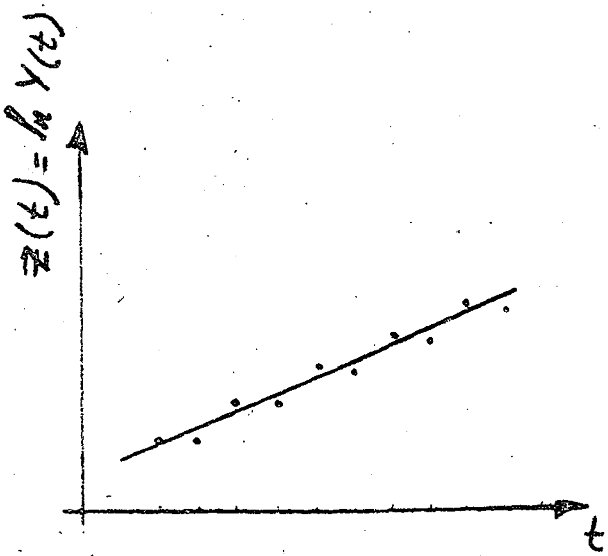
$$\ln\tilde{Y}(t) = \ln a + mt$$

o sea

$$\tilde{Z}(t) = mt+b$$

que es la ecuación de una línea recta y, por lo tanto, las constantes m y $b = \ln a$ se calculan mediante las fórmulas que se obtuvieron para el caso de regresión lineal, con $\tilde{z}_i = \ln\tilde{Y}(t_i)$



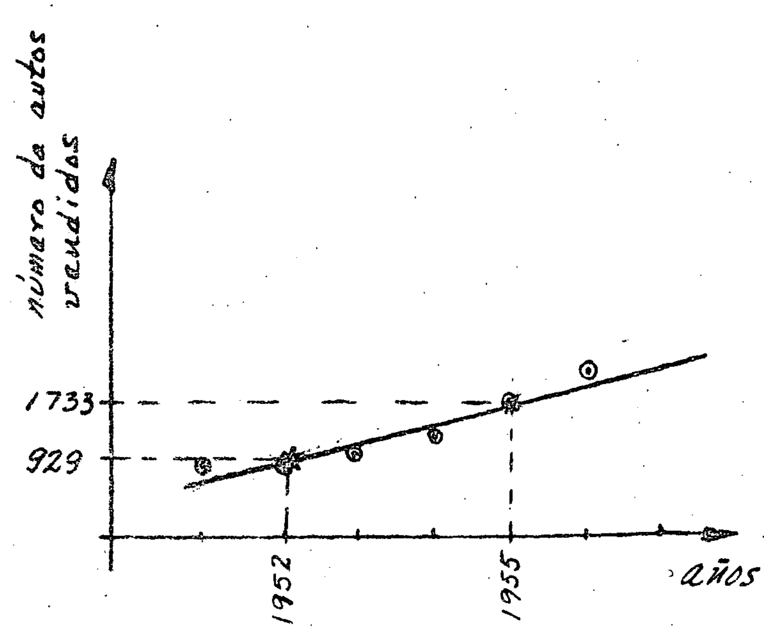


METODO DE DOS PROMEDIOS

El método de dos promedios consiste en dividir los datos en dos partes y calcular el promedio de las Y_i y los tiempos centrales correspondientes a cada parte, con lo cual se obtienen los puntos $(\bar{t}_1, \bar{Y}_1)$ y $(\bar{t}_2, \bar{Y}_2)$ por los cuales pasa la recta buscada

EJEMPLO

Año	Número de autos vendidos	Año central	Promedio
1951	860		
1952	910	1952	929
1953	1018		
1954	1326		
1955	1749	1955	1,733
1956	2125		



PROMEDIOS MOVILES

Los promedios móviles de orden N de la serie de números y_1, $y_2, y_3, \ldots, y_N$, $y_{N+1}, \ldots, y_n$ es la secuencia de promedios aritméticos

$$\frac{y_1+y_2+\ldots+y_N}{N}, \quad \frac{y_2+y_3+\ldots+y_{N+1}}{N}, \quad \ldots$$

Por ejemplo, los promedios móviles de orden 2 de los números 3, 9, 5 y 1 es la secuencia

$$\frac{3+9}{2} = 6, \quad \frac{9+5}{2} = 7, \quad \frac{5+1}{2} = 3$$

mientras que los de orden 3 son

$$\frac{3+9+5}{3} = 5.67, \quad \frac{9+5+1}{3} = 5.00$$

Cuando se obtienen los promedios móviles de los datos de una serie cronológica, cada promedio se asocia con el tiempo central de los tiempos que le corresponden.

EJEMPLO

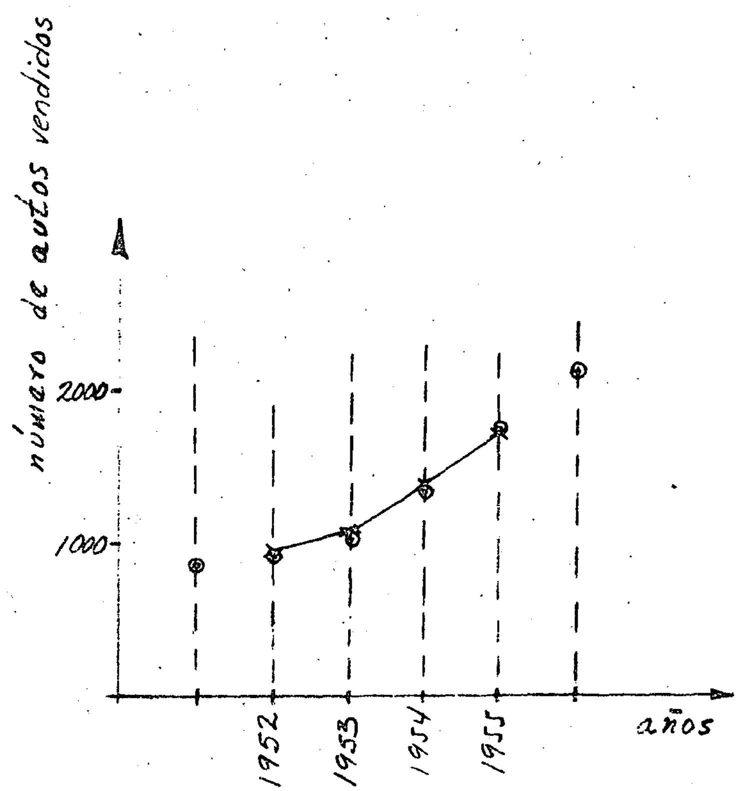
AÑO	Número de Autos vendidos	Promedio móvil
1951	860	
1952	910	929
1953	1,018	1,085
1954	1,326	1,364
1955	1,749	1,733
1956	2,121	

$$\frac{860+910+1018}{3} = 929$$

$$\frac{910+1018+1326}{3} = 1,085$$

$$\frac{1018+1326+1749}{3} = 1,364$$

$$\frac{1326+1749+2121}{3} = 1,733$$



COMPONENTE ESTACIONAL

Como ya se dijo, la componente estacional sirve para indicar las variaciones periódicas que ocurren a corto plazo (periodos menores de un año), tales como los aumentos en las ventas en navidad de cada año, o el aumento en la demanda de servicio en un banco cada fin de semana.

Al proceso de separar las cuatro componentes de una serie de tiempos se le denomina proceso de descomposición. En este proceso usaremos los símbolos, T, E, C e I para denotar, respectivamente, las componentes de tendencia general, estacional, cíclica e irregular. Además, consideraremos que la serie se compone del producto de las cuatro componentes, es decir, que

$$Y = T\,E\,C\,I$$

Si eliminamos a la componente estacional de una serie nos queda solamente T C I; si luego dividimos a la serie entre T C I obtenemos E, es decir

$$\frac{Y}{TCI} = \frac{TECI}{TCI} = E$$

Si la componente estacional se requiere para los meses del año y los datos son mensuales, para eliminarla se requiere sacar los promedios móviles de orden 12. Si se requiere para las semanas y los datos son semanales, E se elimina calculando los promedios móviles de orden 52.

Resumiendo, para calcular la componente estacional se practica el siguiente procedimiento:

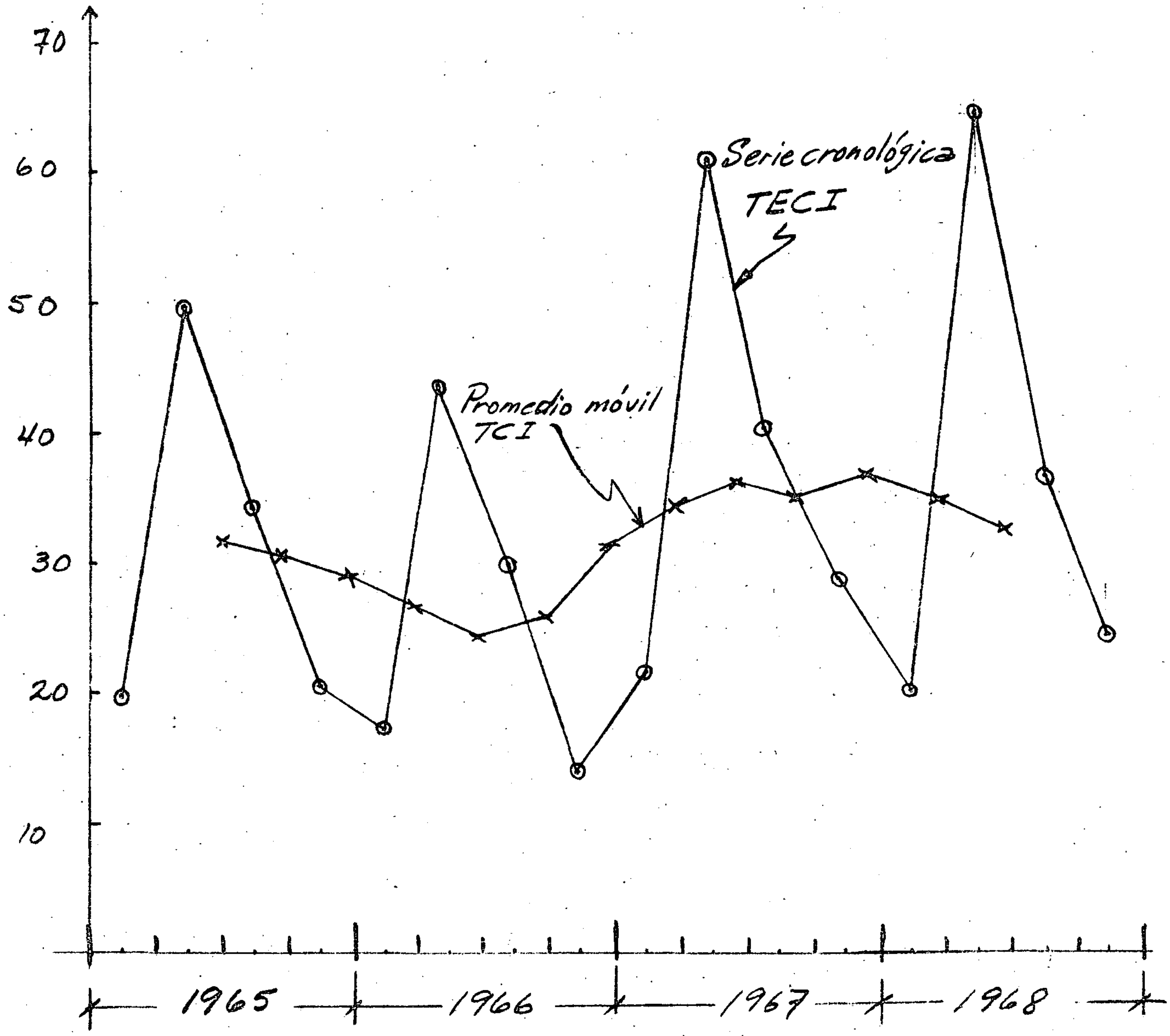
1. Cálculo de los promedios móviles (se obtiene TCI)
2. Cálculo de los porcentajes de los promedios móviles (se obtiene TECI/TCI = E)
3. Se calculan las medianas de los valores correspondientes a cada periodo (también se pueden usar los promedios aritméticos). Con esto se obtiene un valor representativo (de tendencia central) de los valores de cada periodo.
4. Se calculan los índices estacionales haciendo que el promedio de estos por periodo sea de 100%.

EJEMPLO

En la siguiente tabla se presenta el consumo promedio por día de fertilizante que se consumió en una región agrícola. Obtener la componente estacional.

Año	Trimestre	Consumo, Y ton/día	Suma	Promedio móvil	Promedio móvil centrado, TCI	Porcentaje del promedio móvil Y/TCI = E, %
1	1	20				
9	2	50				
				31.5		
6	3	35			31.3	111.5
				31.0		
5	4	21	126		30.2	69.5
				29.3		
1	1	18	124		28.7	62.7
				28.0		
9	2	43	117		27.3	157.5
				26.5		
6	3	30	112		27.0	111.1
				27.5		
6	4	15	106		29.8	50.3
				32.0		
1	1	22	110		33.2	66.2
				34.3		
9	2	61	128		35.9	169.9
				37.5		
6	3	39	137		37.4	104.2
				37.3		
7	4	28	150		37.6	74.5
				37.8		
1	1	21	149		37.6	55.8
				37.3		
9	2	63	151		36.8	171.2
				36.3		
6	3	37	149			
8	4	24	145			

Puesto que los datos están dados por trimestre el índice estacional que se obtendrá será para los trimestres, por lo cual los promedios móviles para eliminar, como primer paso, a la componente estacional deben ser de orden 4.



Eliminación de la componente estacional, E, mediante promedios móviles.

$$\frac{20+50+35+21}{4} = \frac{126}{4} = 31.5; \quad \frac{126-20+18}{4} = 124; \quad \text{etc.}$$

$$\frac{31.5+31.0}{2} = 31.3; \quad \frac{31.0+29.3}{2} = 30.2; \text{ etc.}$$

$$(35/31.3) \times 100 = 111.5\%; \quad (21/30.2) \times 100 = 69.5\%; \text{ etc.}$$

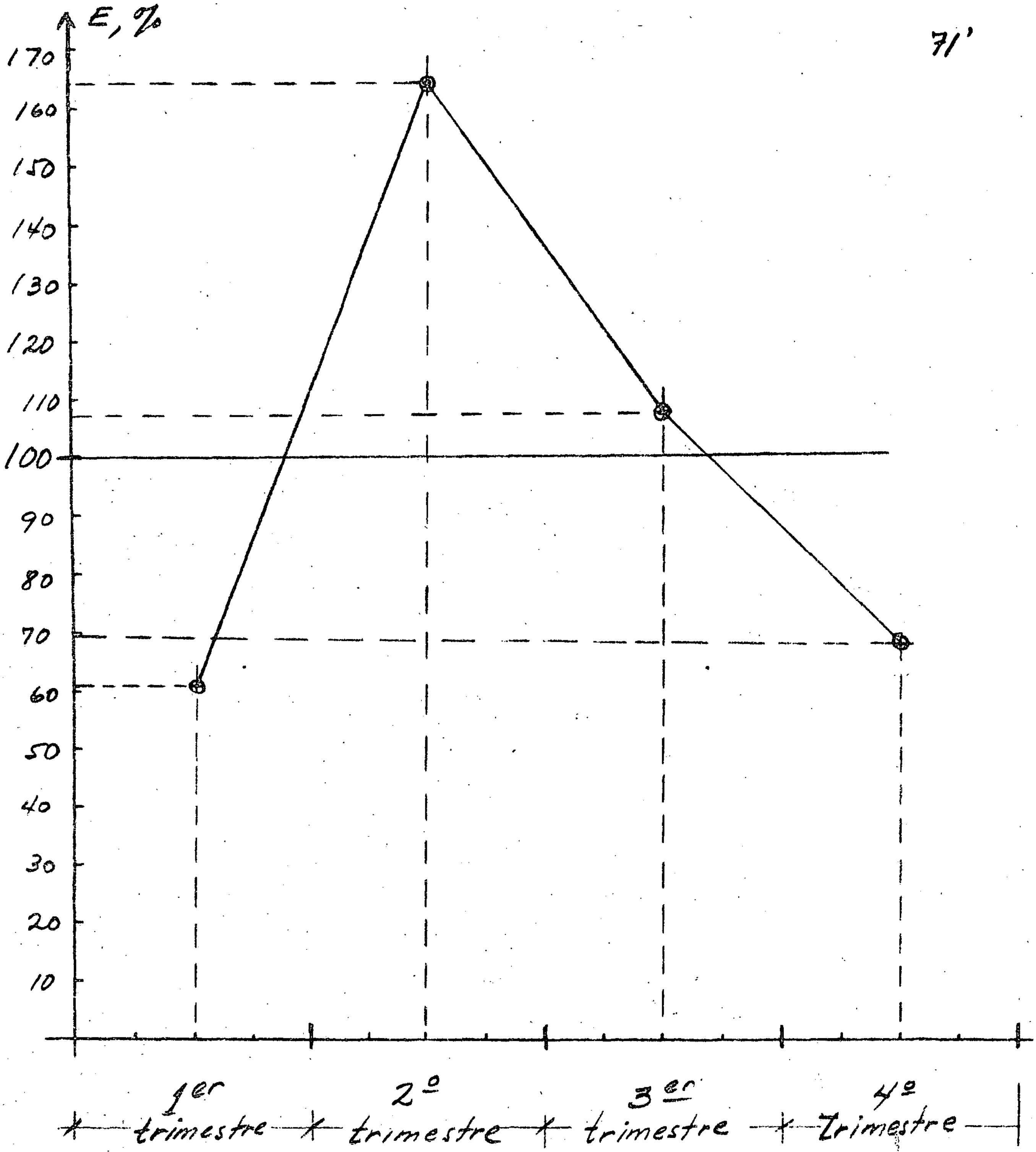
Trimestre	Porcentaje de los promedios móviles				Mediana	Indice estacional, %
	1965	1966	1967	1968		
1		62.7	66.2	55.8	62.7	60.8
2		157.5	169.9	171.2	169.9	164.5
3	111.5	111.1	104.2		111.1	107.5
4	69.5	50.3	74.5		69.5	67.3
					Σ = 413.2	Σ = 400.1 ≐ 400

Para que el promedio por estación sea de 100% se requiere multiplicar cada mediana por el factor

$$\frac{400}{413.2} = 0.968$$

$$62.7 \times 0.968 = 60.8\%; \quad 169.9 \times 0.968 = 164.5$$

$$111.5 \times 0.968 = 107.5; \quad 69.5 \times 0.968 = 67.3$$



INDICE ESTACIONAL

COMPONENTE CICLICA

Como ya se indicó, la componente cíclica de una serie cronológica indica las variaciones periódicas que ésta tiene a largo plazo (tiempos mayores de un año).

Si dividimos la serie de tiempo entre TE obtenemos

$$\frac{Y}{TE} = \frac{TECI}{TE} = CI$$

Cuando los datos de la serie están dados por años, ésta no contiene a la componente estacional (el índice estacional vale 100% para cada periodo), por lo que para obtener CI será suficiente con dividir a la serie entre T.

El procedimiento para obtener la componente cíclica consiste en:

1. Dividir la serie cronológica entre TE, con lo cual se obtiene CI.
2. Calcular los promedios móviles de orden N de la serie CI, con lo cual se elimina I, en donde N debe ser <u>menor</u> que el periodo de los ciclos (nos queda C).

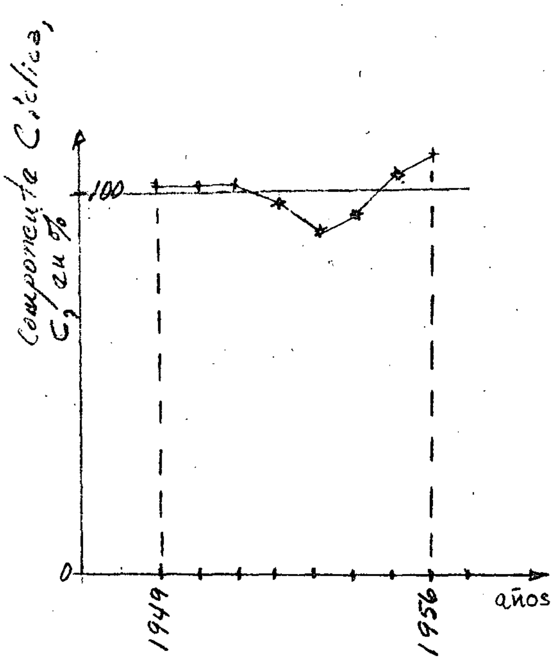
EJEMPLO

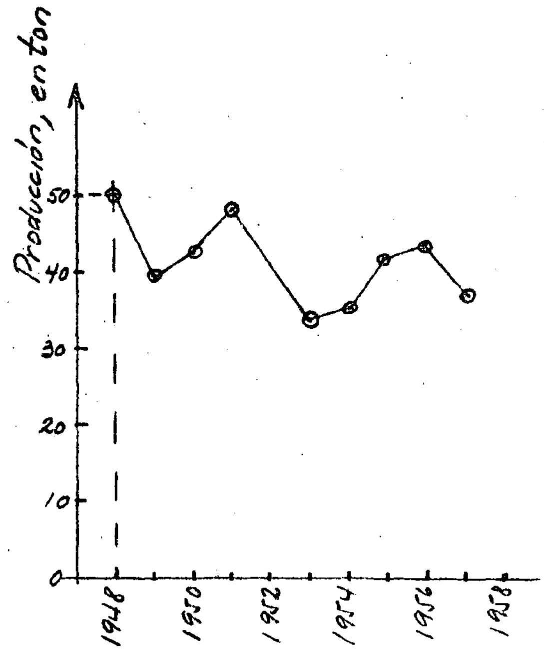
En la siguiente tabla se presentan los datos de producción de uva en una granja, así como sus promedios móviles de orden 5, calcular la componente cíclica.

AÑO	Producción, Y en ton	Promedio móvil, T, en ton	CI = Y/T, en %	Promedio móvil, C en %	Componente irregular CI/C, en %
1948	50.0	41.3	121.1		
1949	39.0	43.1	90.5	103.0	88
1950	41.9	43.0	97.4	102.9	95
1951	48.0	39.7	120.9	103.0	117
1952	36.1	39.8	90.7	98.8	92
195[illegible]	33.8	39.8	84.9	89.3	95
1954	35.3	38.2	92.4	95.5	97
1955	42.0	38.5	109.1	106.4	103
1956	43.6	37.0	117.8	109.8	107
1957	37.9	37.0	102.4		

Puesto que el más pequeño de los ciclos va de 1948 a 1951 (el periodo es de 3 años) y el mayor va de 1951 a 1956 (el periodo es de 5 años), tomaremos promedios móviles de orden 3 para eliminar a I.

$$\frac{121.1+90.5+97.4}{3} = 103.0; \qquad \frac{103.0+121.1+120.9}{3} = 102.9$$





Para evaluar los índices de la componente cíclica es recomendable contar por lo menos con tres periodos completos de datos. Los índices cíclicos se calculan de manera semejante a los estacionales.

EJEMPLO

Supóngase que la componente cíclica de las inversiones anuales en un país con periodo sexenal de gobierno federal es la indicada en la siguiente tabla. Calcular los índices cíclicos.

Año	C, en %
1947	85
48	102
49	117
50	126
51	129
52	137
53	79
54	98
55	121
56	127
57	132
58	143
59	59
60	86
61	121
62	122
63	137
64	149
65	89
66	100
67	112
68	129
69	136
70	138

Año del ciclo	Comp. cíclica CICLOS				Promedio	Indice cíclico, en %
	1	2	3	4		
1	85	79	59	89	78.0	67.5
2	102	98	86	100	96.5	83.5
3	117	121	121	112	117.8	101.9
4	126	127	122	129	126.0	109.0
5	129	132	137	136	133.5	115.4
6	137	143	149	138	141.8	122.7
					Σ= 693.6	600.0

$$\frac{600}{693.6} = 0.865$$

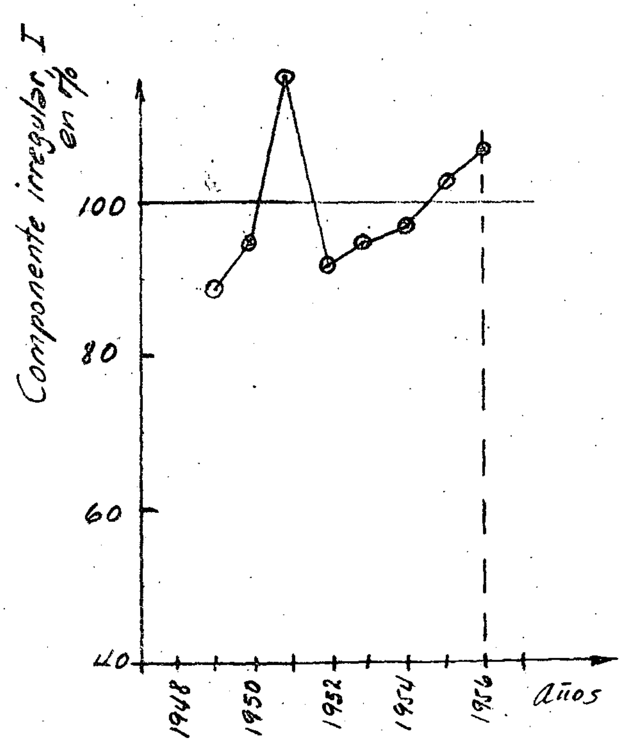
COMPONENTE IRREGULAR

Como se indicó, la componente irregular de una serie cronológica indica las variaciones que en ésta ocurren al azar.

Una vez que se ha calculado C, para obtener I basta dividir CI entre C, es decir

$$CI/C = I$$

En la tabla del penúltimo ejemplo se encuentra calculada la componente irregular de la serie cronológica correspondiente a la producción de uva en una granja.



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

PROBABILIDAD Y ESTADISTICA:
FUNDAMENTOS Y APLICACIONES

INFERENCIA ESTADISTICA

M. EN I. AUGUSTO VILLARREAL ARANDA

JUNIO, 1978

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 varianza, 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 $n (n \geq 30)$ 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

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

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

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

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

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

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

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

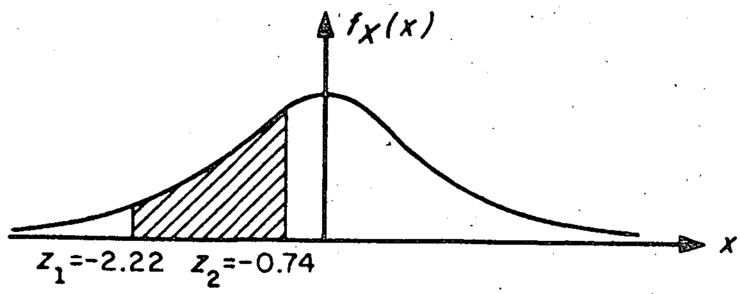


Fig 4.1 Distribución normal correspondiente al ejemplo

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

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

$$= 0.5 - 0.4985 = 0.0015$$

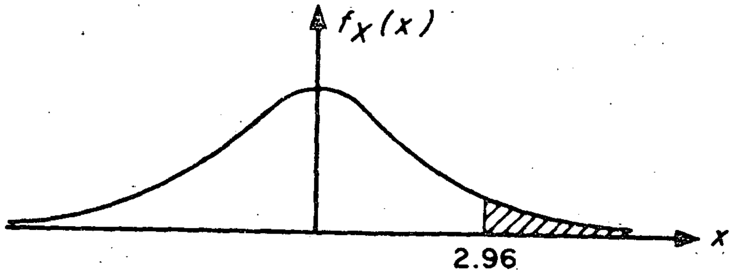


Fig 4.2 Distribución normal correspondiente al ejemplo

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:

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

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

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

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

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

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

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

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

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.

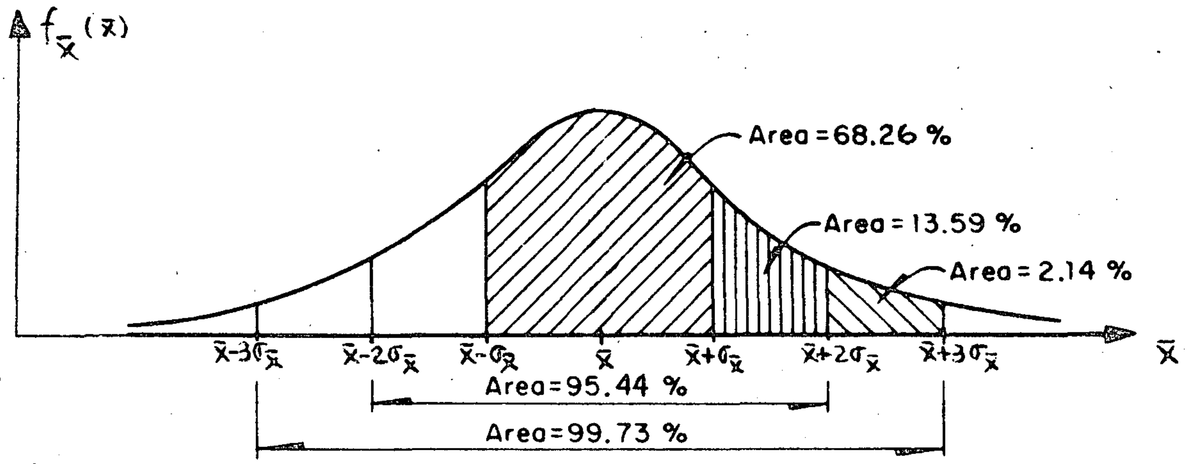


Fig 8.1

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

$$(69.296,\ \ 74.704)$$

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%,

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

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

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 (o 53.14%), tal como se muestra en la *fig* 9.1.

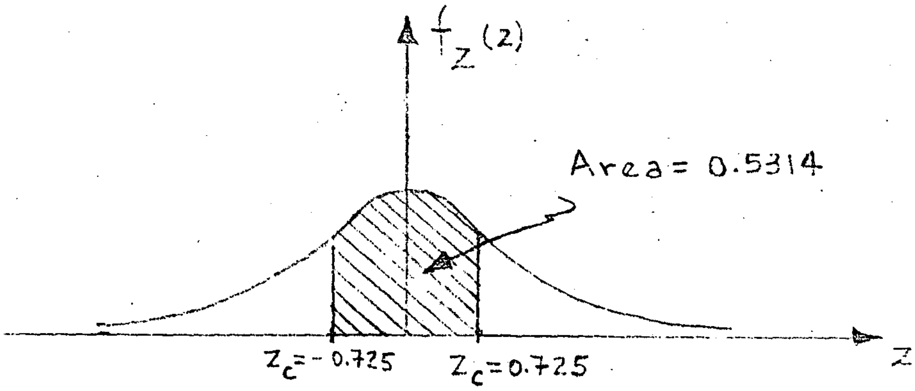


Fig 9.1

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

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

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

$$200 \pm 21.638$$

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

$$H_0 : \mu_S = \mu_1$$

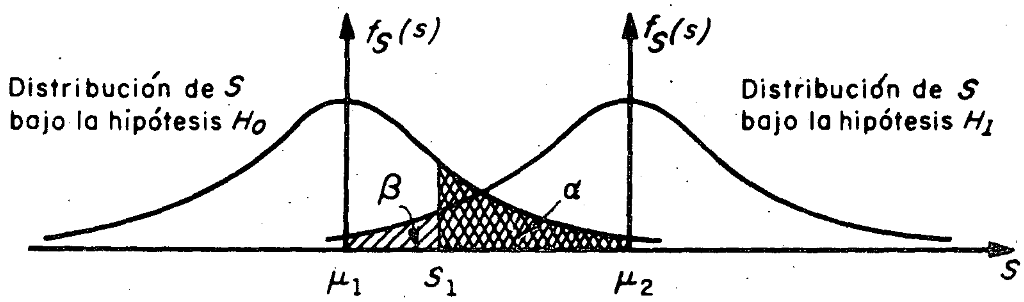
$$H_1 : \mu_S = \mu_2$$

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.



$$P[S>S_1] = \alpha \text{ (error tipo I)}$$
$$P[S<S_1] = \beta \text{ (error tipo II)}$$

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.

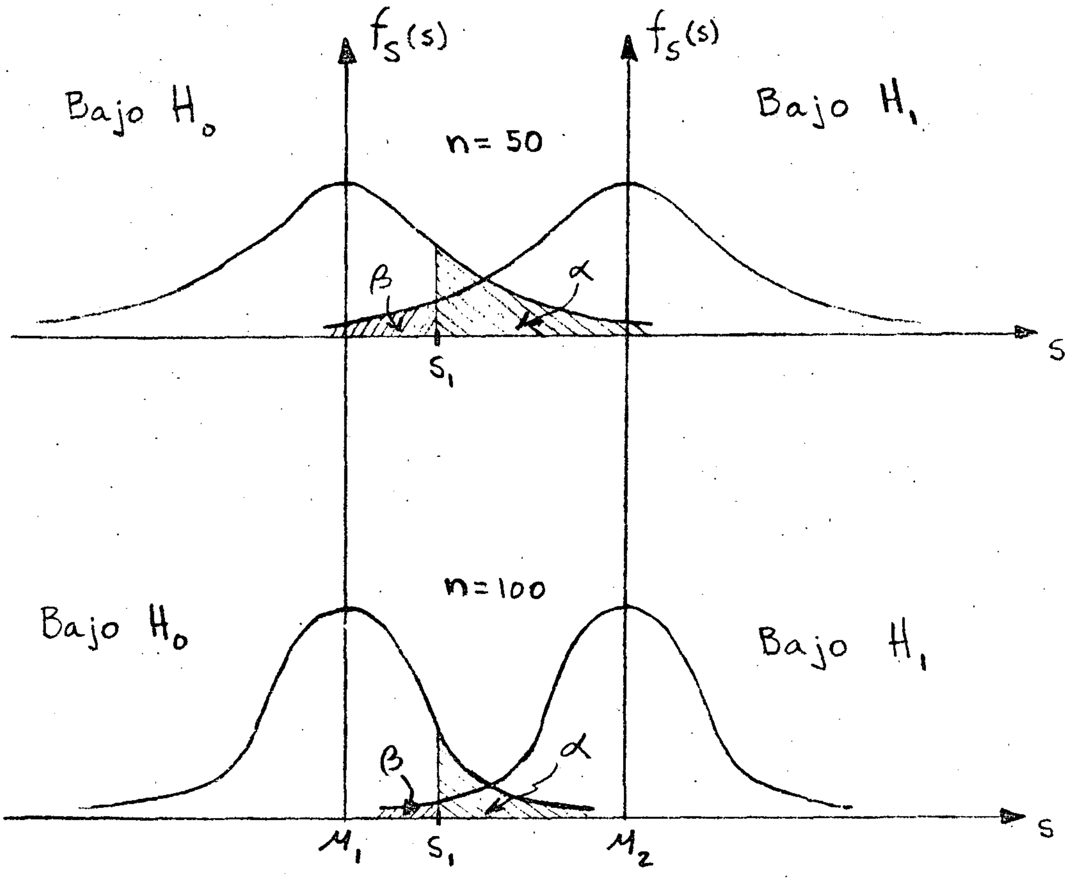


Fig 13.2

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 o 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 $Z < -2.58$.

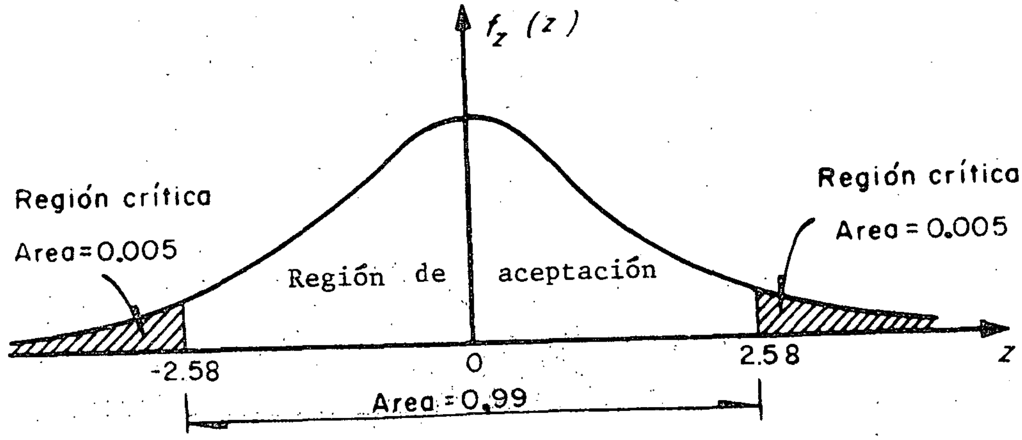


Fig 14.1 *Región de significancia*

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.

$$\chi^2 = \frac{n\ S_X^2}{\sigma^2} \tag{3.14}$$

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.

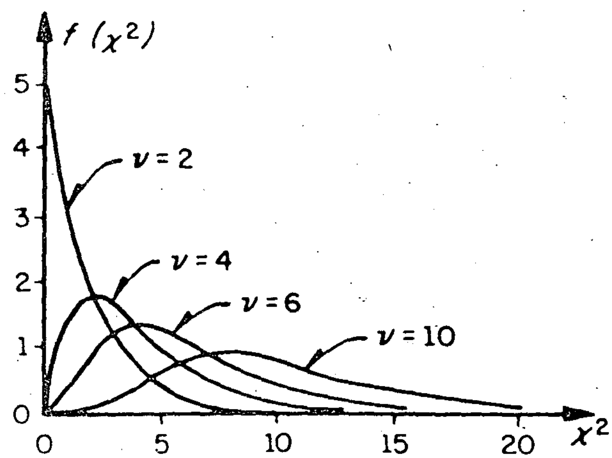
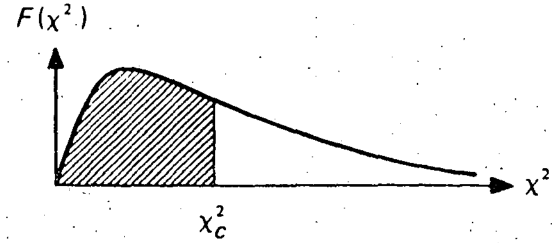


Fig 21. Distribución Ji cuadrada para distintos valores de ν

TABLA 8. VALORES CRITICOS χ_c^2



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

46

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 $\alpha = 0.01$

r_2 = Grados de libertad del denominador	r_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	4.83
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	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	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	5.79	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

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).

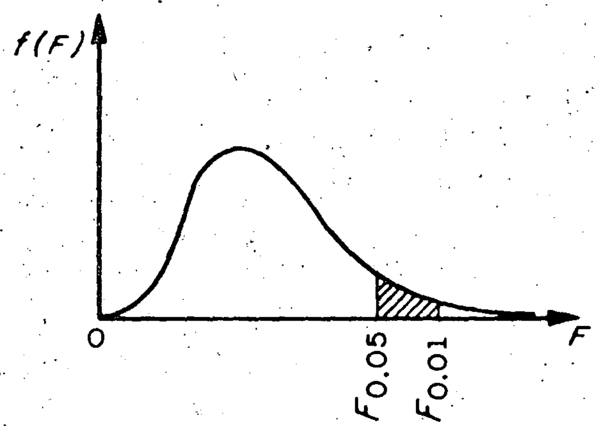


Fig 22. Distribución F.

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.

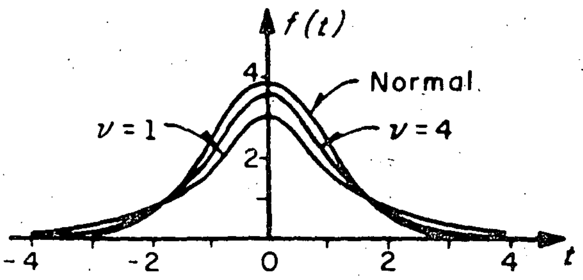


Fig 23. Distribución t de Student para distintos valores de ν

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,

$$- t_c \leqslant \frac{\bar{X} - \mu}{S_X} \sqrt{n - 1} \leqslant t_c$$

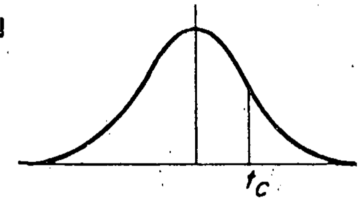
representa un intervalo de confianza para t, a partir del cual se puede estimar que μ se encuentra dentro del intervalo

$$\bar{X} - t_c \frac{\sigma_X}{\sqrt{n-1}} < \mu < \bar{X} + t_c \frac{\sigma_X}{\sqrt{n-1}}$$

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

$$T = \frac{\overline{X} - \overline{Y}}{\epsilon \sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}} \tag{3.17}$$

donde

$$\epsilon = \sqrt{\frac{n_X S_X^2 + n_Y S_Y^2}{n_X + n_Y - 2}} \tag{3.18}$$

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.