ESTADISTICA DESCRIPTIVA

Octavio A. Rascón Chávez

1ª Edición

Sección: Matemáticas
Noviembre de 1979

1

ESTADISTICA DESCRIPTIVA

Octavio A. Rascón Chávez

Sección: Matemáticas
Noviembre de 1979

1

ESTADISTICA DESCRIPTIVA

POR DR. OCTAVIO A. RASCON CH.

EXPERIMENTO

PARA FINES DE ESTE CURSO, SE ENTENDERA POR EXPERIMENTO A TODO PROCESO DE OBSERVACION DE UN FENOMENO O VARIABLE DE INTERES. 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 O COLECCION 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

MUESTREO: ES EL 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; ES EL FENOMENO EN ESTUDIO

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

EJEMPLOS

1. EXPERIMENTO: LANZAMIENTO DE UNA MONEDA DIEZ VECES

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

   MUESTRA: GRUPO DE 10 OBSERVACIONES

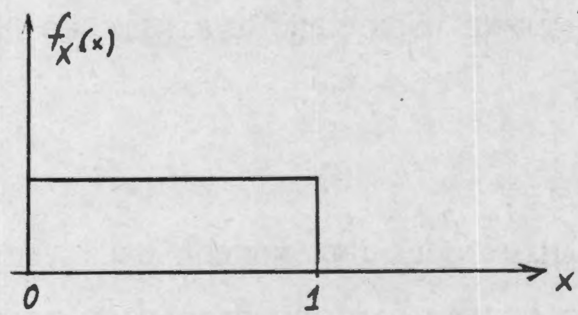
2. EXPERIMENTO: MEDICION DE LA PRECIPITACION PLUVIAL DIARIA EN LA CIUDAD DE MEXICO DURANTE DIEZ AÑOS

   POBLACION: SUCESION INFINITA NO NUMERABLE DE VALORES (CONTINUA)

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

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

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



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

METODO DE MUESTREO ALEATORIO

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

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

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, 0394, 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
	$\Sigma=30$	$\Sigma=30/30=1$	

¿CUAL ES LA FRECUENCIA RELATIVA DE VALORES MENORES O IGUALES QUE 93?: 27/30

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

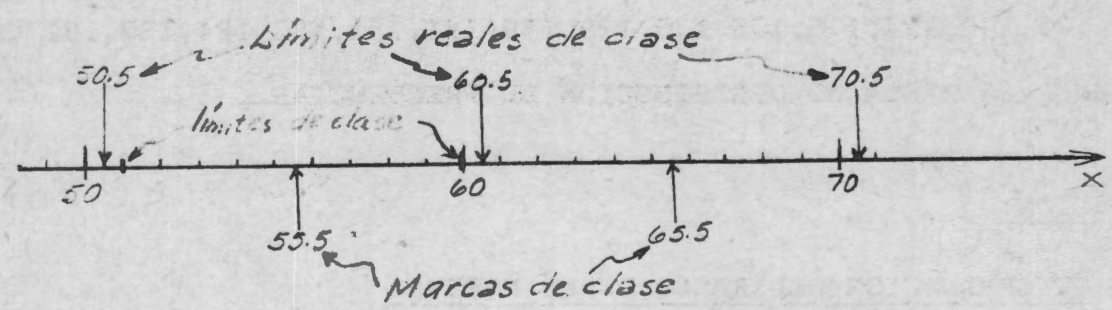
LIMITES REALES DE CLASE: SON LOS VALORES MINIMO Y MAXIMO QUE SON FRONTERA ENTRE LOS INTERVALOS. ESTOS DEBEN TENER UNA CIFRA DECIMAL MAS QUE LOS DATOS.

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

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



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

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

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

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

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

LIMITES REALES INFERIORES DE CLASE

LIMITES REALES SUPERIORES DE CLASE

PROCEDIMIENTO DE AGRUPAMIENTO

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

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

EJEMPLO

EN UN ESTUDIO ANTROPOLOGICO SE OBTUVO UNA MUESTRA DE 30 ESTATURAS

DE LOS VARONES ADULTOS RESIDENTES EN UNA REGION. LOS DATOS, ORDENADOS EN FORMA CRECIENTE DE VALORES, FUERON LOS SIGUIENTES:

159, 161,163,163,163,167,167,167,167,168,168,168,169,169,170, 171,171,173,174,175,175,175,178,179,181,181,183,184,187,191 CM.

OBTENER LA TABLA DE DISTRIBUCION DE FRECUENCIAS.

SOLUCION:

1. DETERMINACION DEL RANGO DE LA MUESTRA

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

2. DETERMINACION DEL NUMERO DE INTERVALOS

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

3. DETERMINACION DE LOS LIMITES DE CLASE

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

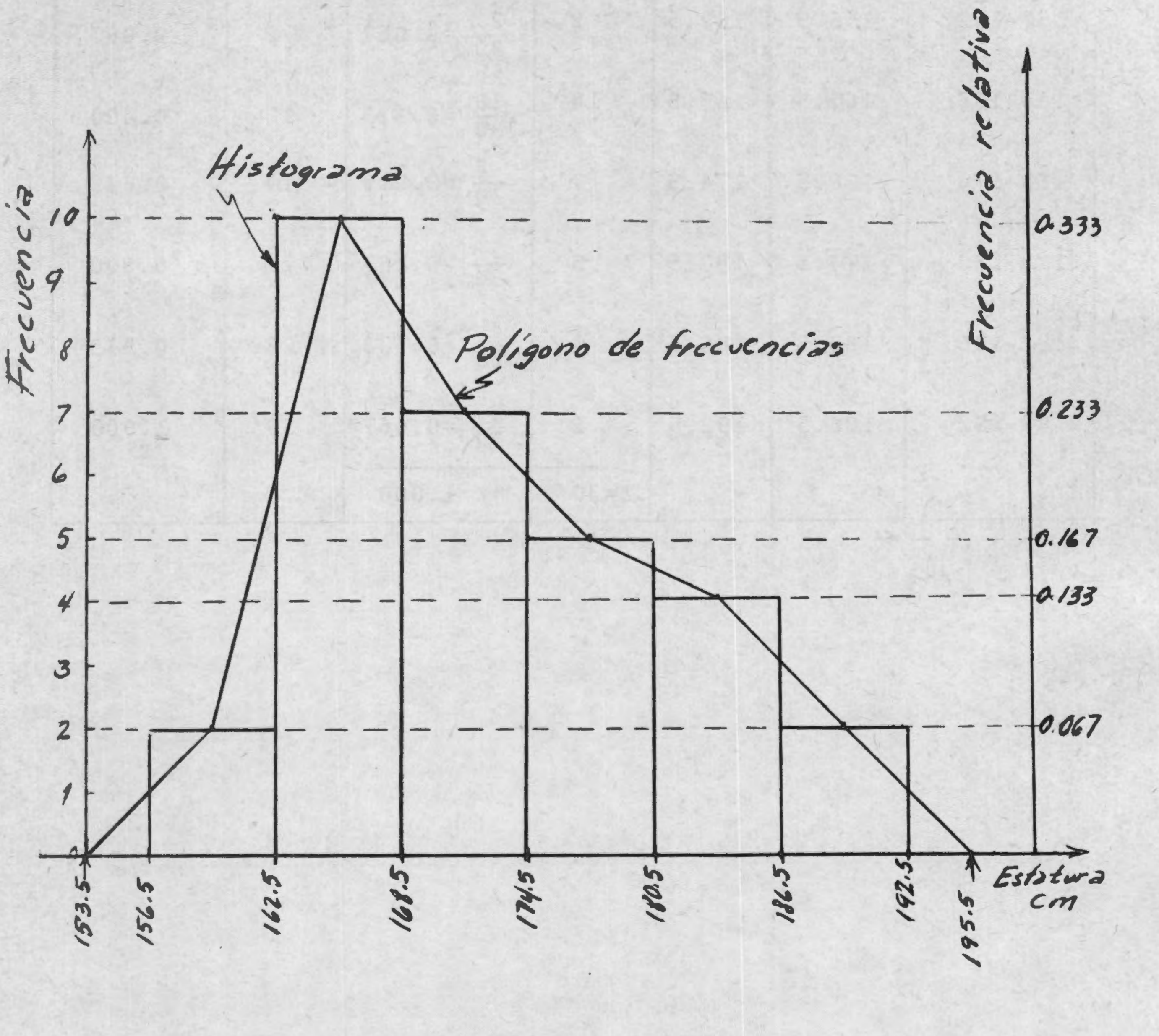
   TOMAREMOS UN ANCHO DE 6 CM, CON LO CUAL EL RANGO DEL AGRUPAMIENTO ES 6 x 6 = 36 CM. LA DIFERENCIA DE RANGOS ES 36-32=4, QUE SE REPARTE EN LOS DOS INTERVALOS EXTREMOS EQUITATIVAMENTE. POR LO TANTO, LOS INTERVALOS RESULTAN SER:

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

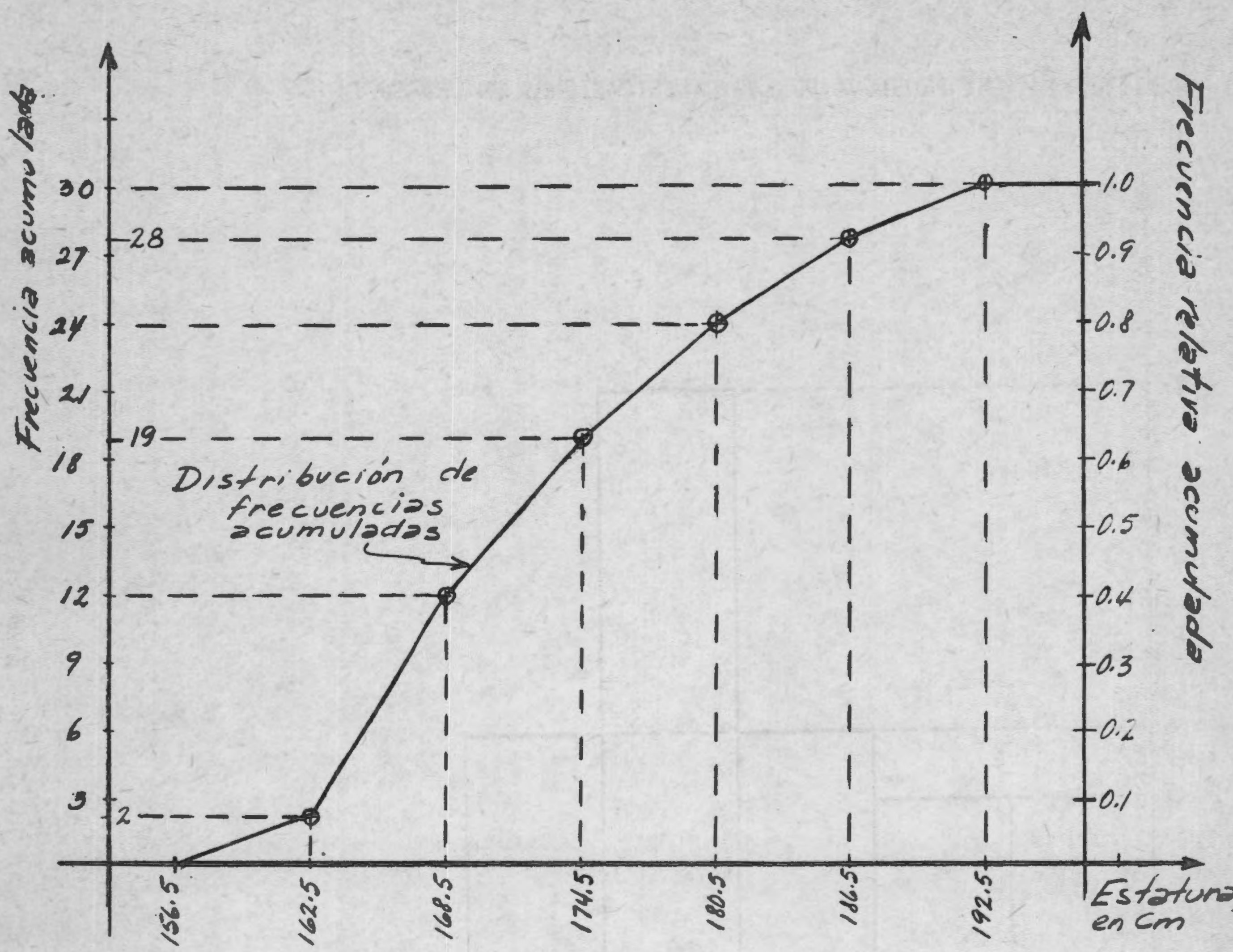
4. INTEGRACION DE LA TABLA:

INTERVALO	LIMITES REALES INF.	LIMITES REALES SUP.	FREC.	FREC. REL.	FREC. ACUM.	FREC.REL. ACUM.
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
			Σ=30	Σ=1.000		

PRESENTACION GRAFICA DE LAS DISTRIBUCIONES DE FRECUENCIAS



DISTRIBUCION DE FRECUENCIAS DE LOS DATOS DE LAS ESTATURAS DE LOS VARONES RESIDENTES EN UNA REGION

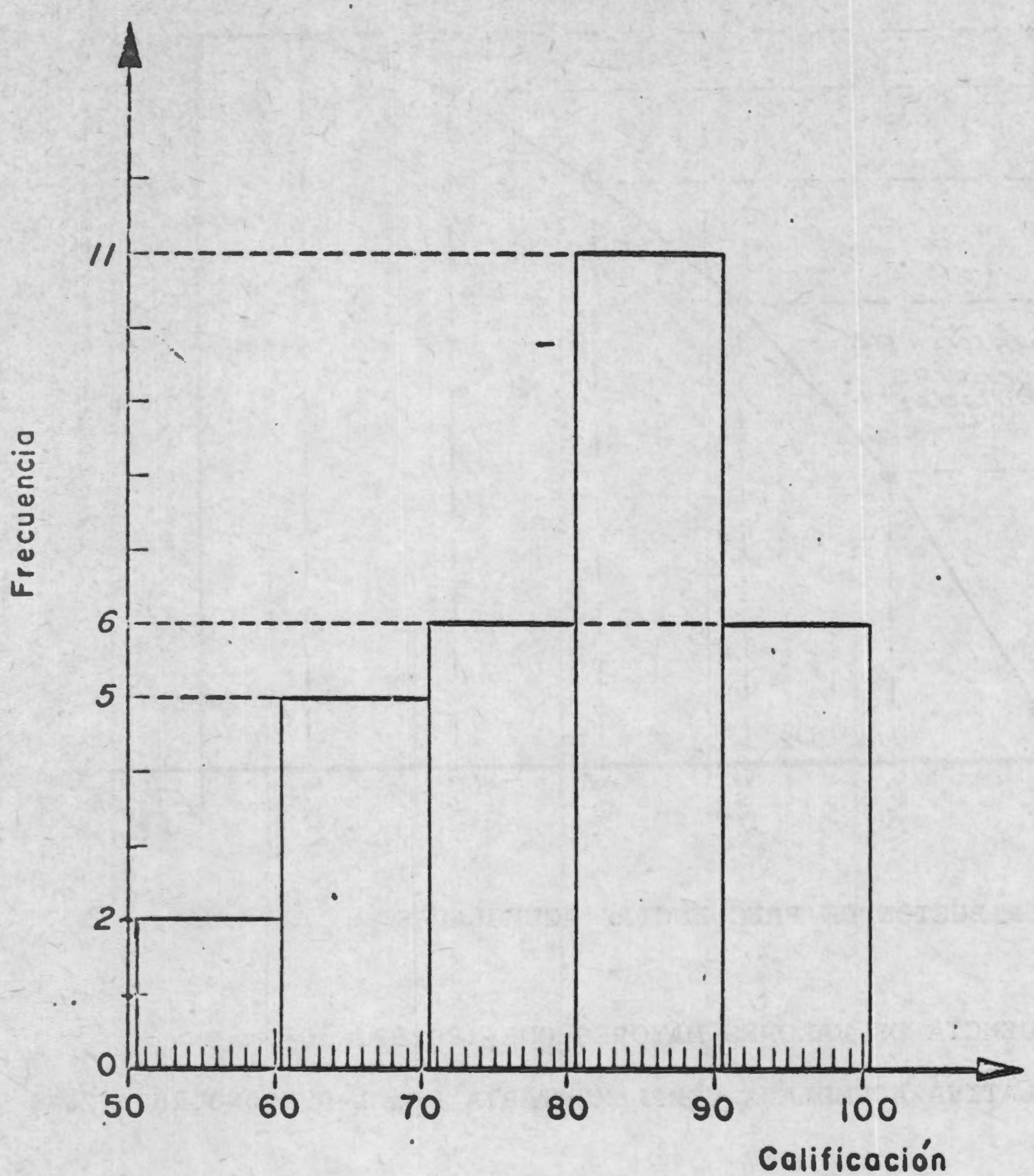


DISTRIBUCION DE FRECUENCIAS ACUMULADAS

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

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

HISTOGRAMA DEL PROBLEMA DE LAS CALIFICACIONES EN PEDAGOGIA

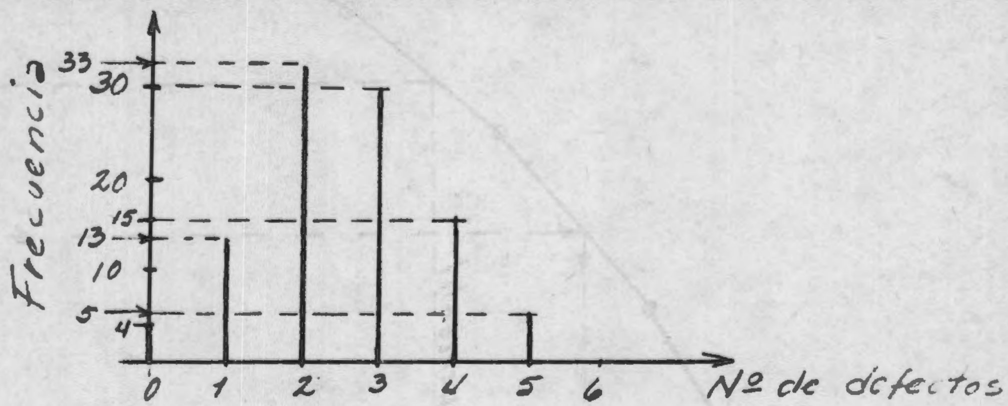


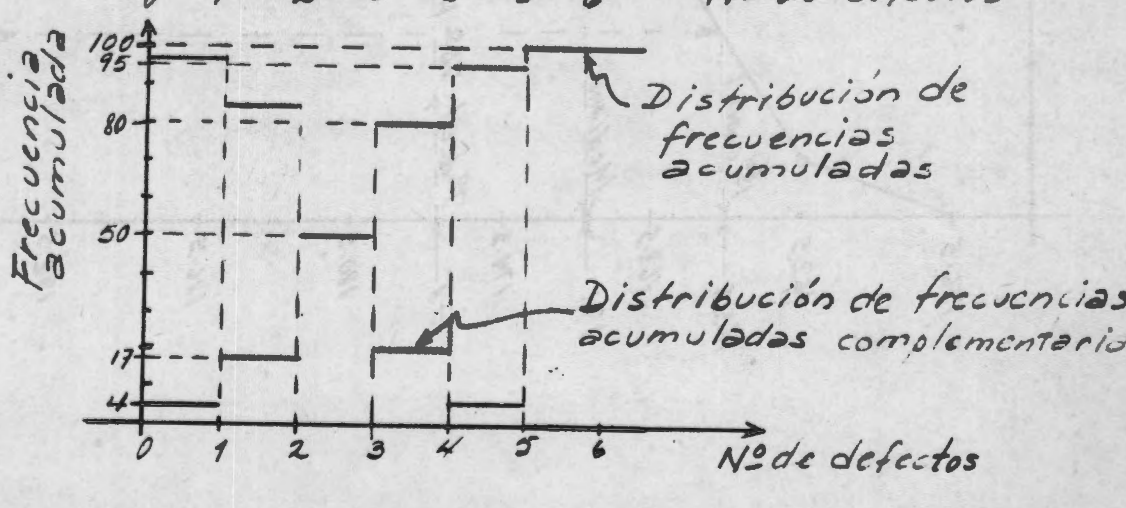
TAREA: DIBUJAR EL POLIGONO, DE FRECUENCIAS Y LAS CURVAS DE FRECUENCIAS RELATIVAS ACUMULADAS Y COMPLEMENTARIAS.

EJEMPLO

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

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



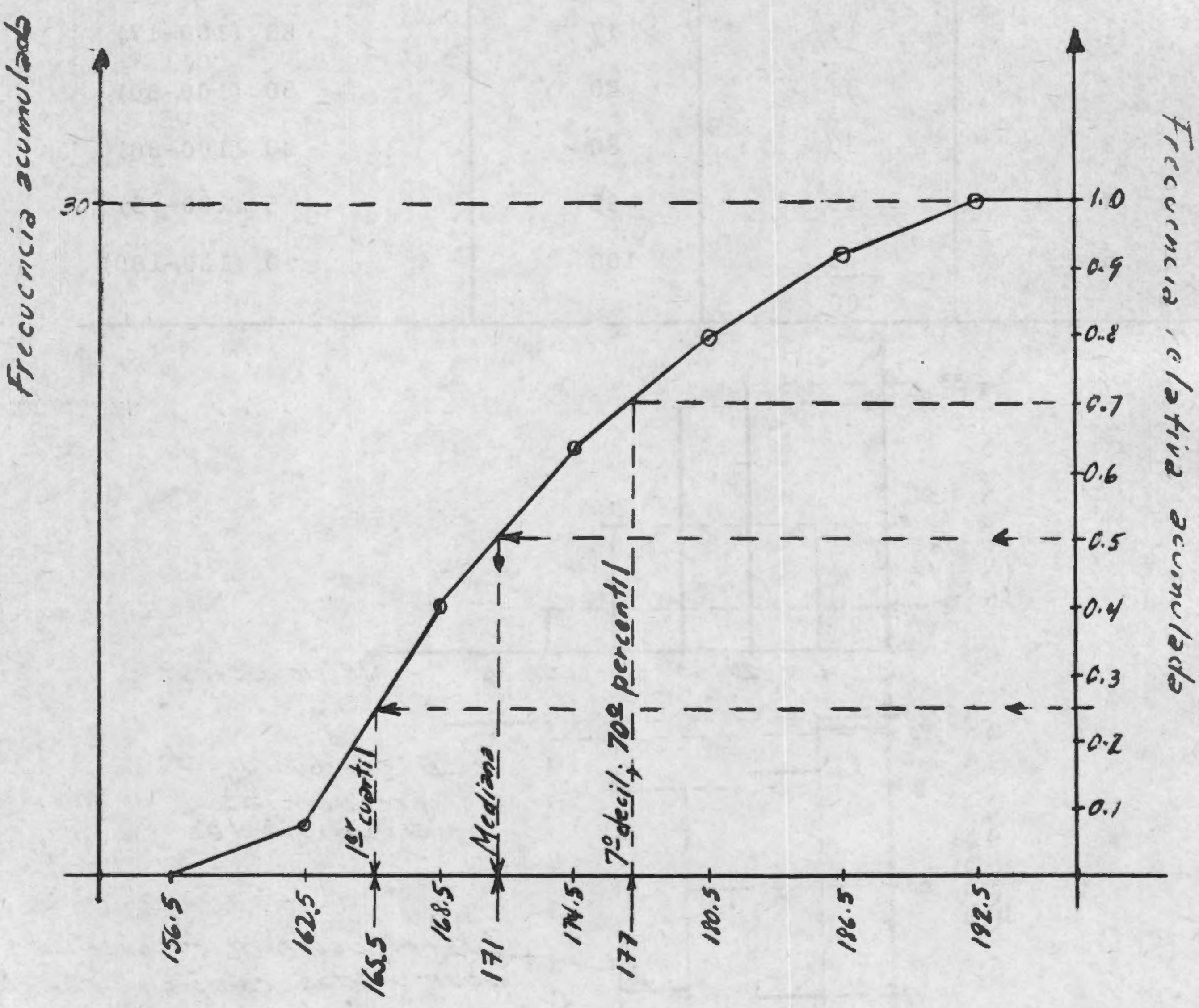


<u>PERCENTILES</u>:- SON LOS VALORES DE LA VARIABLE CORRESPONDIENTES A FRECUENCIAS RELATIVAS ACUMULADAS QUE SON MULTIPLOS DE 1 POR CIENTO.

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

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

<u>MEDIANA</u>:- VALOR DE LA VARIABLE CORRESPONDIENTE A LA FRECUENCIA RELATIVA ACUMULADA DE 50%.



MEDIDAS REPRESENTATIVAS DE LOS DATOS

MEDIDAS DE TENDENCIA CENTRAL

VALOR MEDIO O PROMEDIO ARITMETICO

PARA DATOS NO AGRUPADOS

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

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

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

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

EJEMPLO

SEA EL EJEMPLO ENUNCIADO ANTERIORMENTE DE LOS DEFECTOS EN MONOBLOCKS.

SE TENIA:

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

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

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

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

EJEMPLO

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

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):

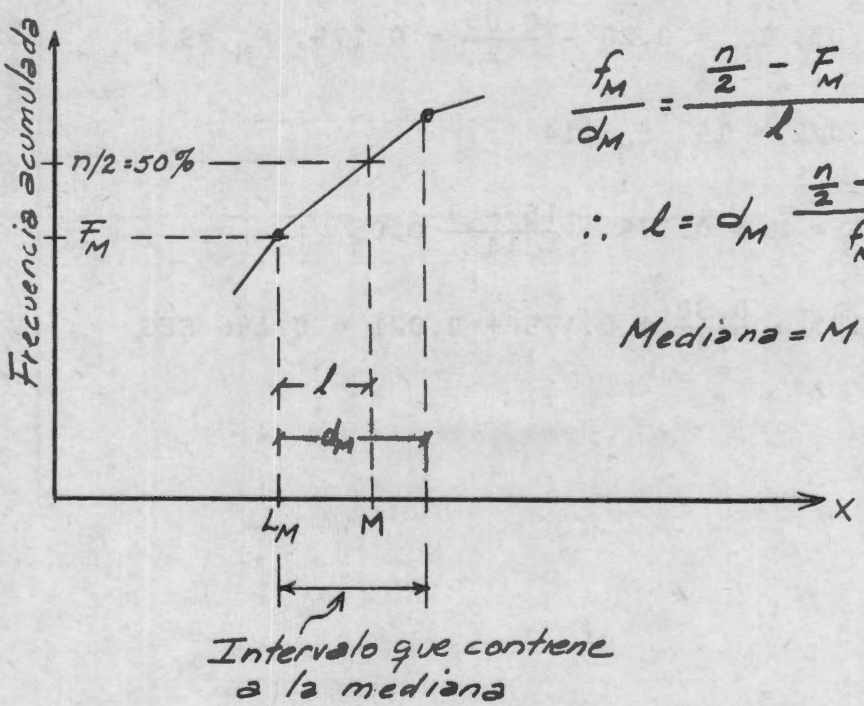
$$\text{MEDIANA} = M = L_M + \frac{\frac{n}{2} - F_M}{f_M} d_M$$

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



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

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

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

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
			$\Sigma=30$		$\sum_{j=1}^{5} f_i x_i = 5.95$

PROMEDIO ARITMETICO

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

MODO = 0.20 SEG

MEDIANA

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

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

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

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

MEDIDAS DE DISPERSION

RANGO = MAXIMO VALOR OBSERVADO - MINIMO VALOR OBSERVADO

VARIANCIA : SI LOS DATOS *NO* ESTAN AGRUPADOS;

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

SI LOS DATOS ESTAN AGRUPADOS:

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

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

DESVIACION ESTANDAR

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

COEFICIENTE DE VARIACION

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

EJEMPLO

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

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

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

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

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

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

MODO = 86

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

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

EJEMPLO

MEDIDA DE DISPERSION (DATOS AGRUPADOS POR VALORES)

Rango = 1.48 - 0.18 = 1.30

Datos x	Frecuencia	xf	x^2	x^2f
0.18	4	0.72	0.032	0.128
0.28	1	0.28	0.078	0.078
0.36	2	0.72	0.130	0.260
0.38	1	0.38	0.144	0.144
0.48	7	3.36	0.230	1.610
0.49	1	0.49	0.240	0.240
0.51	1	0.51	0.260	0.260
0.55	1	0.55	0.302	0.302
0.57	3	1.71	0.325	0.975
0.65	12	7.80	0.422	5.064
0.72	9	6.48	0.518	4.662
0.78	14	10.92	0.608	8.512
0.83	7	5.81	0.689	4.823
0.88	2	1.76	0.774	1.548
0.92	5	4.60	0.846	4.230
0.96	8	7.68	0.922	7.376
1.00	1	1.00	1.000	1.000
1.03	4	4.12	1.061	4.244
1.06	2	2.12	1.124	2.248
1.09	3	3.27	1.189	3.567
1.12	2	2.24	1.254	2.508
1.18	1	1.18	1.392	1.392
1.21	2	2.42	1.464	2.928
1.23	1	1.23	1.513	1.513
1.26	1	1.26	1.588	1.588
1.34	1	1.34	1.796	1.796
1.36	1	1.36	1.850	1.850
1.40	1	1.40	1.960	1.960
1.43	1	1.43	2.045	2.045
1.48	1	1.48	2.190	2.190
		$\Sigma=79.62$		$\Sigma=71.041$

$$\bar{X} = \frac{79.62}{100} = 0.796$$

$$S_X^2 = \bar{x}^2 - \bar{x}^2 = 0.710 - 0.634$$
$$= 0.076$$

$$\bar{x}^2 = 0.634$$

$$S_x^2 = 0.076$$

$$S_x = \sqrt{S_x^2} = \sqrt{0.076} = 0.276$$

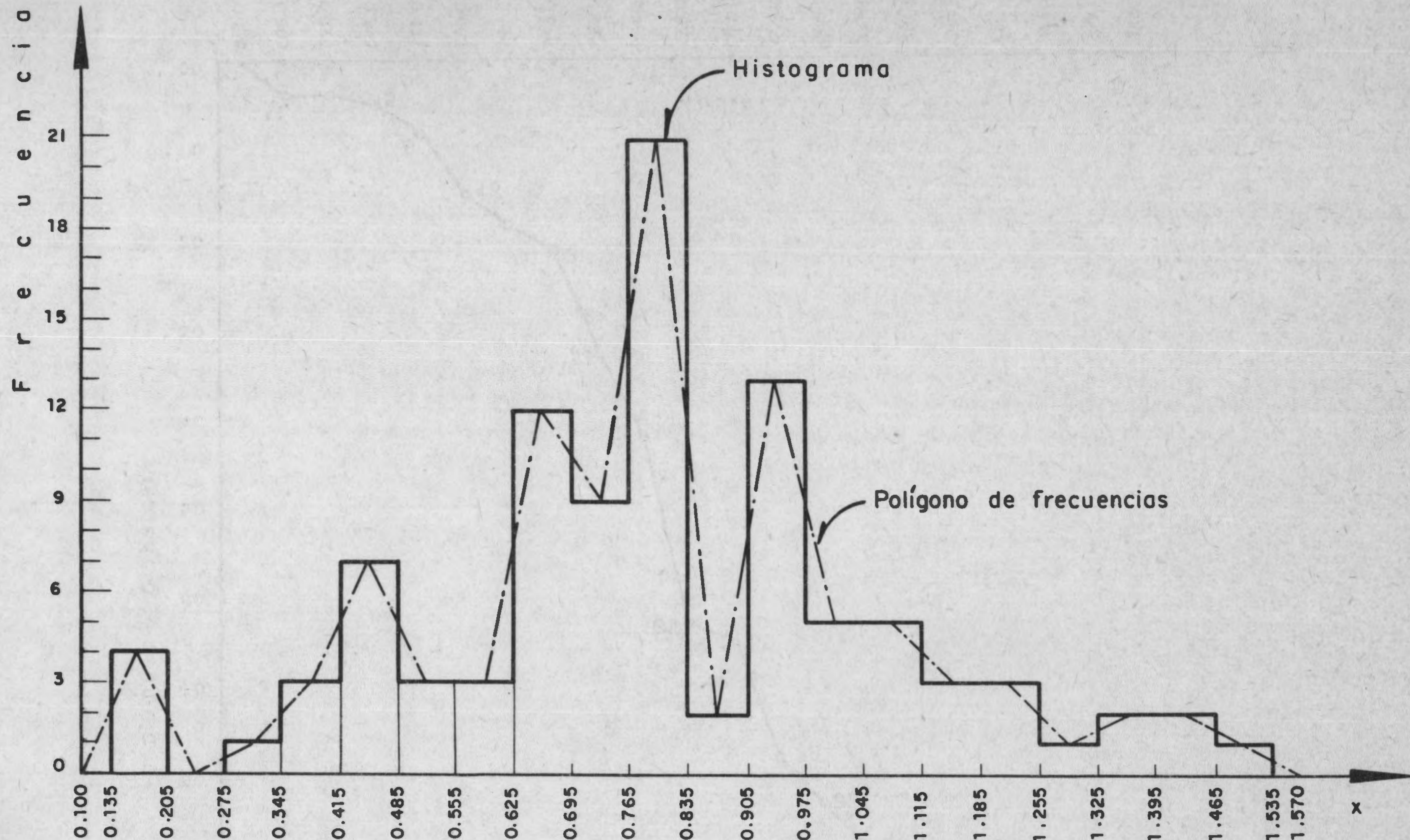
$$\text{Coeficiente de variación} = V_x = \frac{S_x}{\bar{x}} = \frac{0.276}{0.796} = 0.347 = 34.7\%$$

	Intervalo	Límites reales		FREC.	FRECUENCIA RELATIVA	FREC. ACUM.	FREC. REL. ACUMULADA
		INF.	SUP.				
1	0.14-0.20	0.135	0.205	4	4/100=0.04	4	0.04
2	0.21-0.27	0.205	0.275	0	0/100=0.00	4	0.04
3	0.28-0.34	0.275	0.345	1	1/100=0.01	5	0.05
4	0.35-0.41	0.345	0.415	3	3/100=0.03	8	0.08
5	0.42-0.48	0.415	0.485	7	7/100=0.07	15	0.15
6	0.49-0.55	0.485	0.555	3	3/100=0.03	18	0.18
7	0.56-0.62	0.555	0.625	3	3/100=0.03	21	0.21
8	0.63-0.69	0.625	0.695	12	12/100=0.12	33	0.33
9	0.70-0.76	0.695	0.765	9	9/100=0.09	42	0.42
10	0.77-0.83	0.765	0.835	21	21/100=0.21	63	0.63
11	0.84-0.90	0.835	0.905	2	2/100=0.02	65	0.65
12	0.91-0.97	0.905	0.975	13	13/100=0.13	78	0.78
13	0.98-1.04	0.975	1.045	5	5/100=0.05	83	0.83
14	1.05-1.11	1.045	1.115	5	5/100=0.05	88	0.88
15	1.12-1.18	1.115	1.185	3	3/100=0.03	91	0.91
16	1.19-1.25	1.185	1.255	3	3/100=0.03	94	0.94
17	1.26-1.32	1.255	1.325	1	1/100=0.01	95	0.95
18	1.33-1.39	1.325	1.395	2	2/100=0.02	97	0.97
19	1.40-1.46	1.395	1.465	2	2/100=0.02	99	0.99
20	1.47-1.53	1.465	1.535	1	1/100=0.01	100	1.00
				$\Sigma = 100$	$\Sigma = 1.00$		

21/111

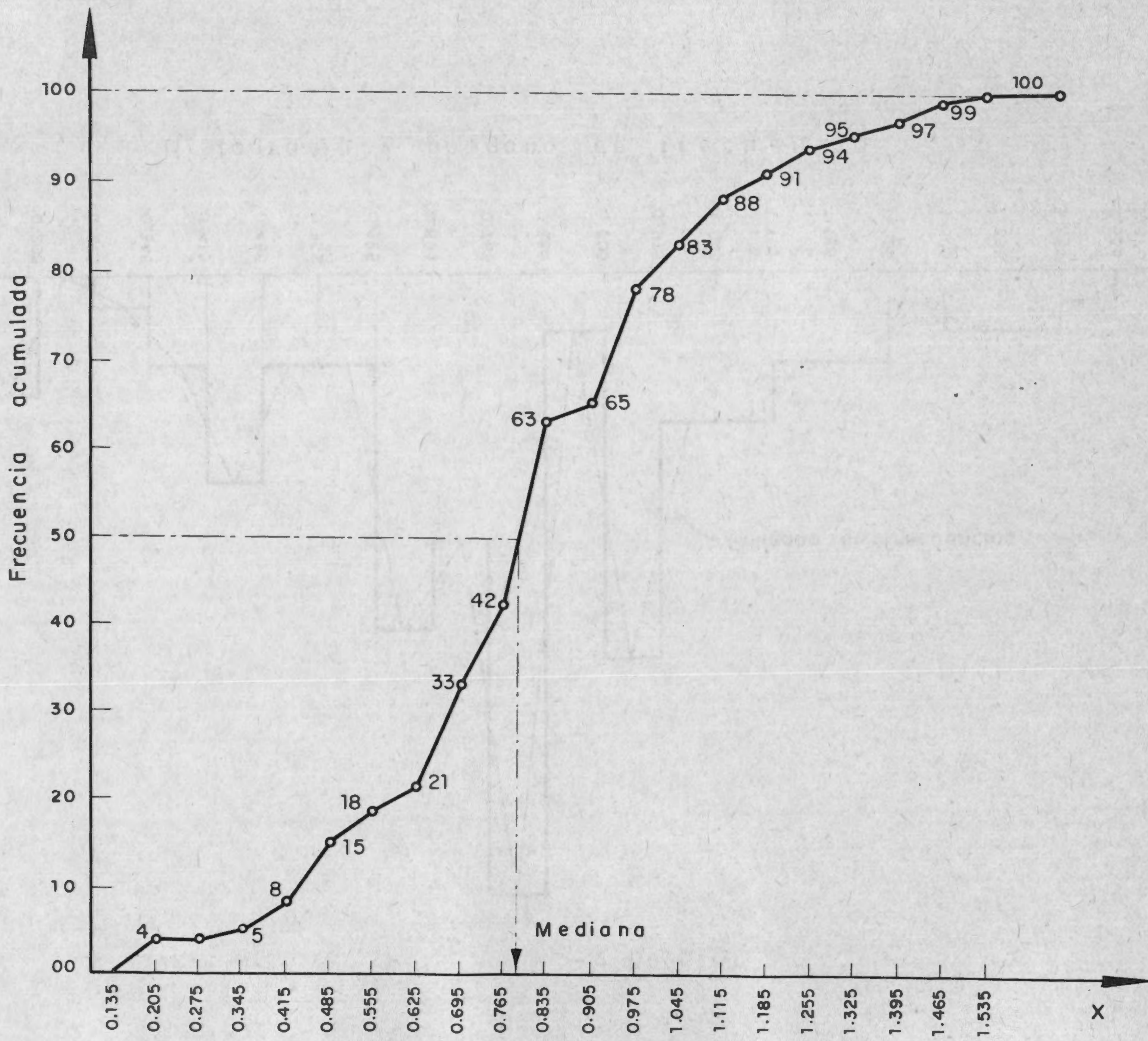
						0.78							
						0.78							
						0.78							
						0.78							
						0.78							
						0.78							
						0.78							
						0.78							
						0.78							
						0.78							
						0.78		0.96					
					0.65	0.78		0.96					
					0.65	0.78		0.96					
					0.65	0.78		0.96	1.09				
					0.65	0.72	0.88	0.96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

HISTOGRAMA



Histograma y polígono de frecuencias

21v



Curva de frecuencias acumuladas

21 VI

21 VII

MEDIDAS DE TENDENCIA CENTRAL

J	Marca de clase x	Límites reales	Frecuencia f	Frecuencia A-acumulada, F	fx
1	0.17	0.135-0.205	4	4	0.68
2	0.24	0.205-0.275	0	4	0.00
3	0.31	0.275-0.345	1	5	0.31
4	0.38	0.345-0.415	3	8	1.14
5	0.45	0.415-0.485	7	15	3.15
6	0.52	0.485-0.555	3	18	1.56
7	0.59	0.555-0.625	3	21	1.77
8	0.66	0.625-0.695	12	33	7.92
9	0.73	0.695-0.765	9	42	6.57
10	0.80	0.765-0.835	21	63	16.80
11	0.87	0.835-0.905	2	65	1.74
12	0.94	0.905-0.975	13	78	12.22
13	1.01	0.975-1.045	5	83	5.05
14	1.08	0.045-1.115	5	88	5.40
15	1.15	1.115-1.185	3	91	3.45
16	1.22	1.185-1.255	3	94	3.66
17	1.29	1.255-1.325	1	95	1.29
18	1.36	1.325-1.395	2	97	2.72
19	1.43	1.395-1.465	2	99	2.86
20	1.50	1.465-1.535	1	100	1.50
					$\sum_{j=1}^{20} f_j x_j =$ =79.79

Promedio aritmético $\bar{X} = \frac{79.79}{100} = 0.7979$

MODO =0.80

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

$d_M = 0.07$ $L_M = 0.765$ $F_M = 42$

$f_M = 21$ $n = 100$

$$\text{Mediana} = 0.765 + \frac{50-42}{21}\ 0.07$$

$$= 0.765 + 0.026 = 0.794$$

MEDIDAS DE DISPERSION (DATOS AGRUPADOS)

Rango = máximo valor observado - mínimo valor observado
= 1.48 - 0.18 = 1.30

J	Intervalo	Marca de clase x	Frecuencia f	xf	$x-\bar{x}$	$(x-\bar{x})^2$	$(x-\bar{x})^2$ f
1	0.14-0.20	0.17	4	0.68	-0.628	0.394	1.576
2	0.21-0.27	0.24	0	0.00	-0.558	0.311	0.00
3	0.28-0.34	0.31	1	0.31	-0.488	0.238	0.238
4	0.35-0.41	0.38	3	1.14	-0.418	0.175	0.525
5	0.42-0.48	0.45	7	3.15	-0.348	0.121	0.487
6	0.49-0.55	0.52	3	1.56	-0.278	0.077	0.231
7	0.56-0.62	0.59	3	1.77	-0.208	0.043	0.129
8	0.63-0.69	0.66	12	7.92	-0.138	0.019	0.228
9	0.70-0.76	0.73	9	6.57	-0.068	0.004	0.036
10	0.77-0.83	0.80	21	16.80	0.002	0.00	0.000
11	0.84-0.90	0.87	2	1.74	0.072	0.005	0.010
12	0.91-0.97	0.94	13	12.22	0.142	0.020	0.260
13	0.98-1.04	1.01	5	5.05	0.212	0.045	0.225
14	1.05-1.11	1.08	5	5.40	0.282	0.079	0.395
15	1.12-1.18	1.15	3	3.45	0.352I	0.123	0.369
16	1.19-1.25	1.22	3	3.66	0.422	0.178	0.534
17	1.26-1.32	1.29	1	1.29	0.492	0.242	0.242
18	1.33-1.39	1.36	2	2.72	0.562	0.316	0.632
19	1.40-1.46	1.43	2	2.86	0.632	0.399	0.798
20	1.47-1.53	1.50	1	1.50	0.702	0.493	0.493
			100	79.79			7.768

Mediana = $\bar{X} = \dfrac{79.79}{100} = 0.7979$

S_x^2 = Variancia = $\dfrac{7.768}{100} = 0.077$

Desviac. estándar = $S_x = 0.277$ $\left(S_x = \sqrt{S_x^2} \right)$

Coeficiente de variación = $V_x = \dfrac{S_x}{\bar{x}} = \dfrac{0.277}{0.7979} = 0.347$

TRANSFORMACION LINEAL DE VARIABLES

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

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

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

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.

EN OCASIONES LA TRANSFORMACION LINEAL SE PLANTEA COMO

$$y = \frac{x - c_1}{c_2}$$

DONDE c_1 y c_2 SON CONSTANTES. EN ESTE CASO SE OBTIENEN

$$\bar{y} = \frac{\bar{x} - c_1}{c_2} \quad y \quad S^2\,(y) = \frac{1}{c_2^2}\; S^2(x)$$

DE ESTAS ECUACIONES SE LLEGA A

$$\bar{x} = c_2\,\bar{y} + c_1$$

$$S^2(x) = c_2^2\; S^2(y)$$

EJEMPLO

EN EL PROBLEMA DE LOS RESULTADOS, x_i, DE UN EXAMEN SOBRE PEDAGOGIA SE OBTUVO LA DISTRIBUCION DE FRECUENCIAS INDICADA 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
	$\Sigma=30$		$\Sigma=14$		

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

$\bar{x} = 0.467 \times 10 + 75.5 = 80.17$, $S^2(x) = 1.382 \times 10^2 = 138.2$

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

$$y = \frac{x - C_1}{C_2}$$

CON C_1 = MARCA DE CLASE CENTRAL Y

C_2 = ANCHO DE CLASE

TOMANDO $C_1 = 75.5$ y $C_2 = 10$, SE TIENE $y_i = (x_i - 75.5)/10$

POR LO QUE

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

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

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

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

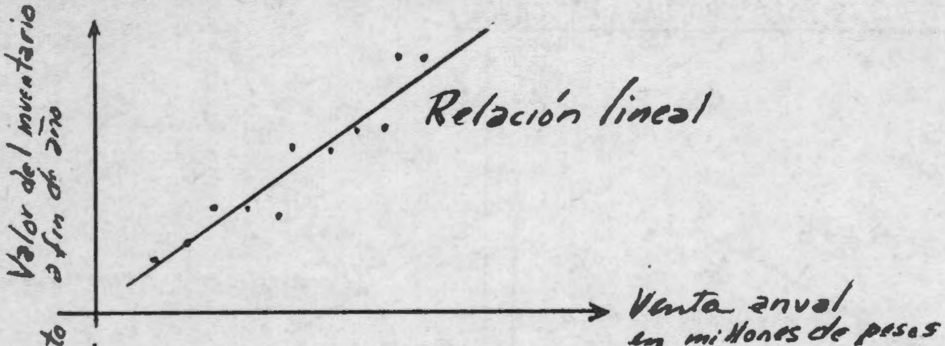
$$y_5 = (-75.5 + 95.5)/10 = 2$$

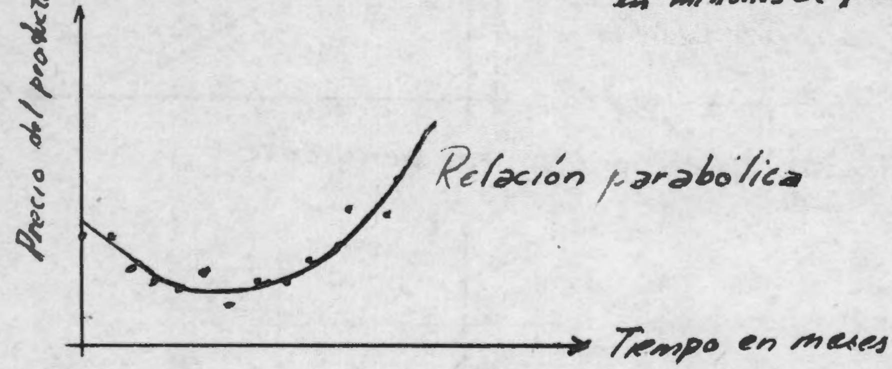
OBSERVESE QUE SE OBTIENE $y = 0$ PARA EL INTERVALO CORRESPONDIENTE A $X_i = C_1$, Y PARA LOS INTERVALOS CON VALORES MAYORES DE X BASTA CON IRLE SUMANDO UNA UNIDAD, MIENTRAS QUE A LOS DE VALORES MENORES, IRLE RESTANDO UNA UNIDAD.

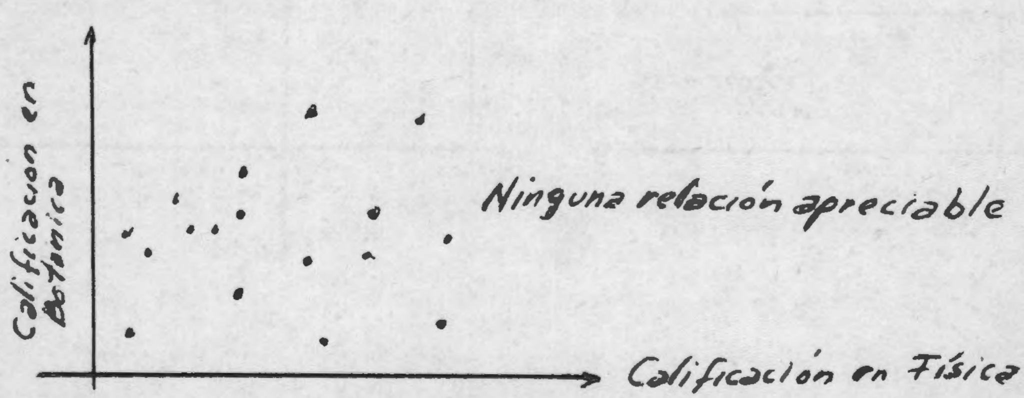
REGRESION LINEAL

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)



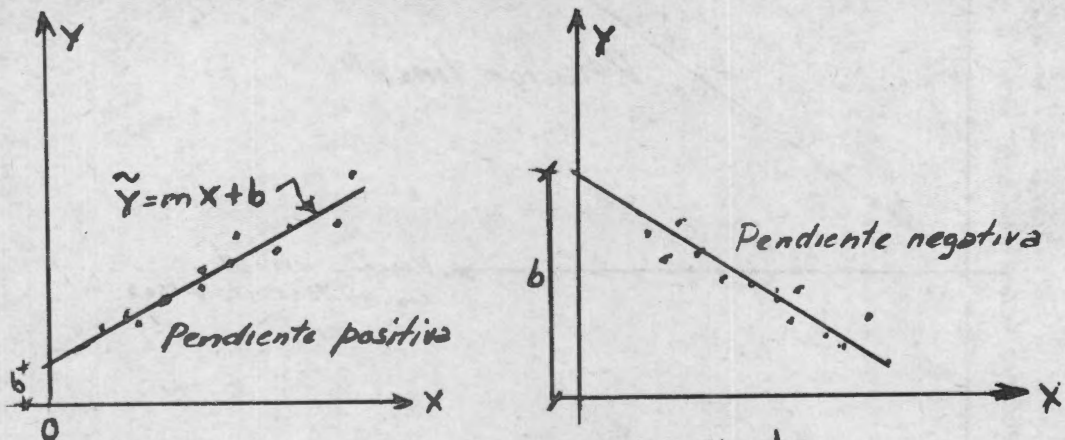




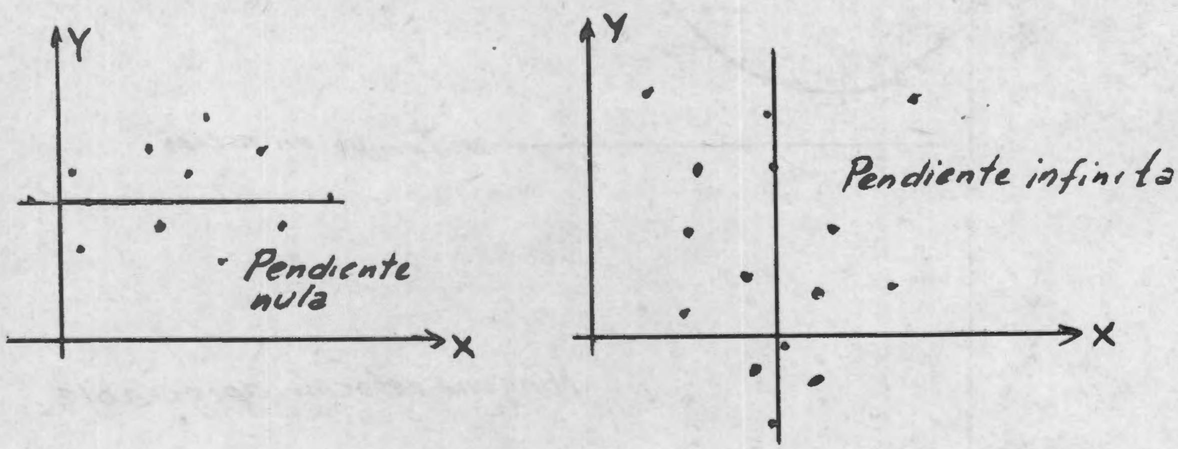
PARA AJUSTAR ALGUNA CURVA A UN GRUPO DE DATOS SE PUEDE PROCEDER DE DIFERENTES MANERAS, DE LAS CUALES LA MAS SENCILLA ES "A OJO", PERO TIENE LA DESVENTAJA DE QUE, POR NO SER SISTEMATICO, DIFERENTES PERSONAS PROPONEN DISTINTAS CURVAS. DE LOS METODOS ANALITICOS O MATEMATICOS, EL MAS COMUN ES EL DE <u>*MINIMOS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

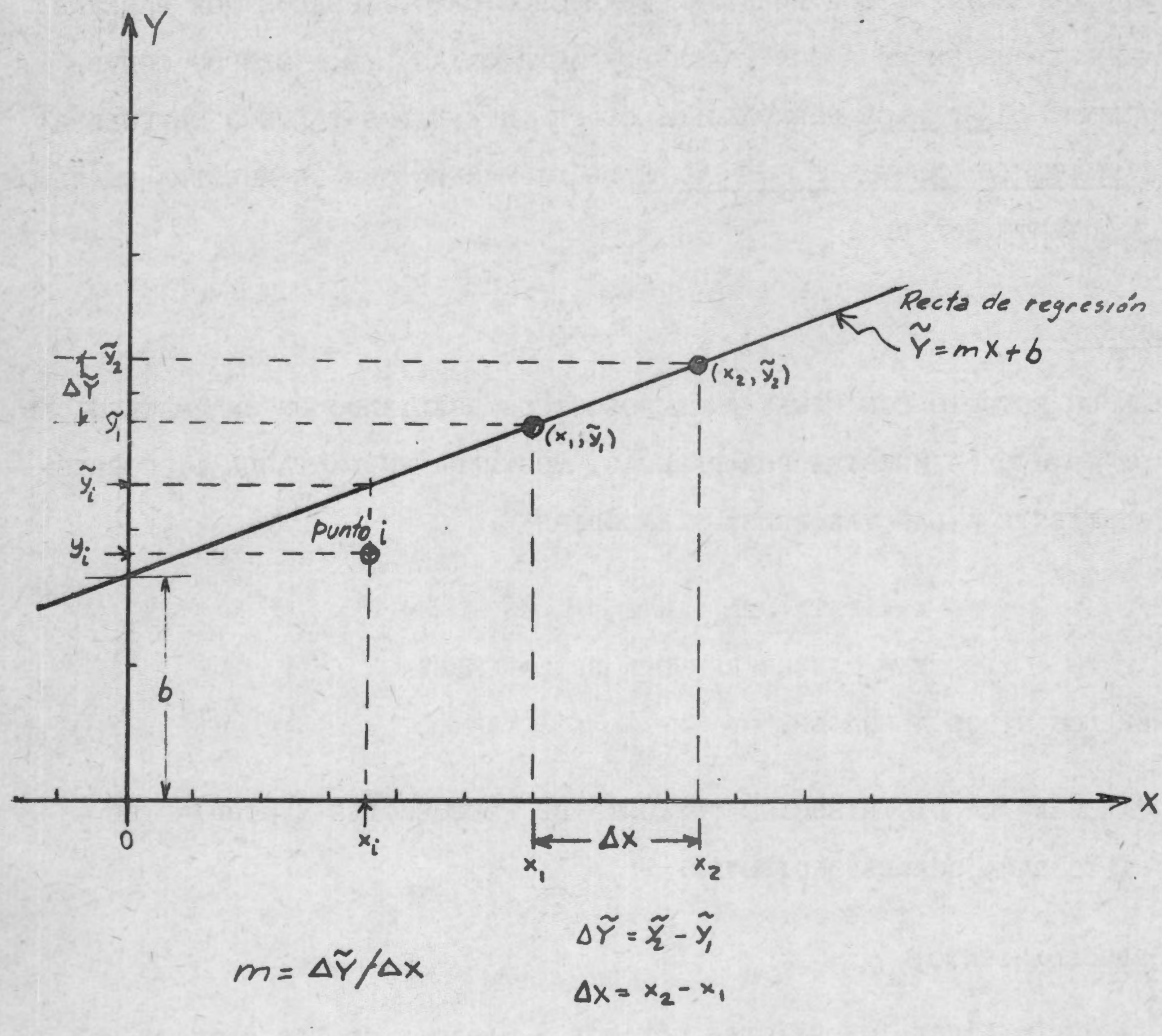
SI X ES LA VARIABLE INDEPENDIENTE Y Y LA DEPENDIENTE, SE DICE QUE LA REGRESION ES DE <u>Y CON BASE EN X</u>, Y VICEVERSA.

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



b = ordenada en el origen ; m = pendiente




Y
Recta de regresión
$\tilde{Y} = mX + b$
$\tilde{y}_2$
$\Delta\tilde{Y}$
$\tilde{y}_1$
$(x_2, \tilde{y}_2)$
$(x_1, \tilde{y}_1)$
$\tilde{y}_i$
punto i
y_i
b
X
0
x_i
ΔX
x_1
x_2
$\Delta\tilde{Y} = \tilde{y}_2 - \tilde{y}_1$
$m = \Delta\tilde{Y}/\Delta X$
$\Delta X = x_2 - x_1$

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.

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

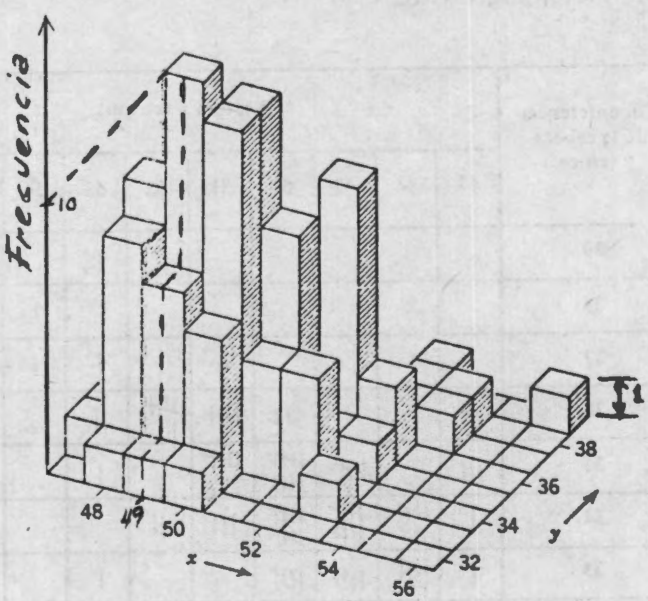
TABLA 2, GRAFICA DE CONTEO CORRESPONDIENTE A LA MUESTRA DE LA TABLA 1.

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	𝍸 II	II			
35	III		𝍸	𝍸 IIII	𝍸 I	I	I			
34	I	𝍸 II	𝍸 𝍸	𝍸 IIII	III	III				
33	I	𝍸 I	𝍸	IIII			I			
32		I	I	I						

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



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,

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

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:

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

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

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

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

EL SEGUNDO TERMINO DE LA MISMA FORMULA QUEDA:

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

$$C_1C_4\bar{y}' + C_1C_3$$

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.

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

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

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

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

$b = 6.5 - 0.13 \times 7 = 5.59$

$\tilde{y} = 0.13\,x + 5.59$

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

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)					

CALCULO DE $\bar{x}$ y S(x)

INTERVALOS	MARCAS DE CLASE	f	xf	x^2	x^2f
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

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

$$\overline{x^2} = \frac{2,201,304.50}{22} = 100,059.30$$

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

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

CALCULO DE $\bar{y}$ y S(y)

INTERVALOS	MARCAS DE CLASE, y	f	yf	y^2	$y^2 f$
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.

MARCAS DE CLASE		FRECUENCIAS		
x	y	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
		Σ = 22		Σ = 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

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

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

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

EJEMPLO

RESOLVER EL PROBLEMA ANTERIOR MEDIANTE EL METODO CORTO.

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

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

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

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

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					

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 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.

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

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

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

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

SE OBTIENE

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

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

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

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

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

DESVIACIONES INEXPLICADAS (NO EXPLICADAS).

COMO CONSECUENCIA DE LA ECUACION ANTERIOR, SE PUEDE DEMOSTRAR QUE

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

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

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

MEDIDAS DE CORRELACION

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

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

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

EN DONDE

(x_i, y_i) = PAREJAS DE DATOS

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

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

N = TOTAL DE PAREJAS DE DATOS

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

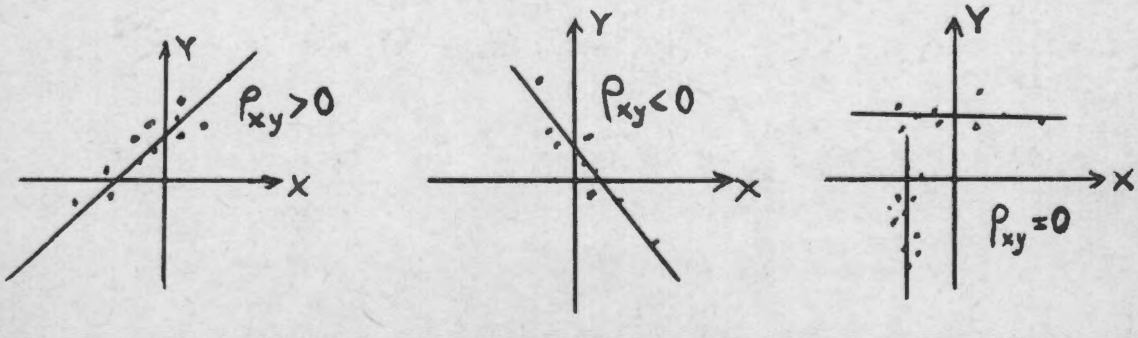
$$\rho_{xy} = \frac{S^2_{xy}}{S(x)\ S(y)}; \quad -1 \leq \rho_{xy} \leq 1$$

EN DONDE

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

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

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



CASO DE CORRELACION PERFECTA

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

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

EN DONDE

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

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

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

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

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

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

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

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

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

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,

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

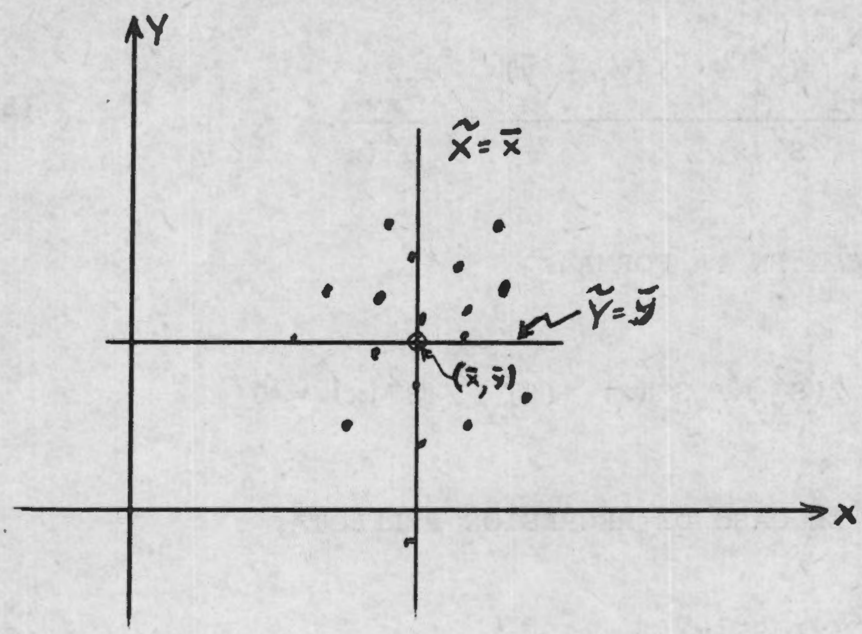
Y, SI $S(x)>0$ Y $S(y)>0$,

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

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

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=∞).



RELACION ENTRE EL COEFICIENTE DE CORRELACION Y LA PENDIENTE DE LA RECTA DE REGRESION

TOMANDO EN CUENTA QUE

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

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

$$\rho_{xy} = \frac{\frac{1}{N}\sum_{i=1}^{N} f_{ixy}x'y' - \overline{x'}\ \overline{y'}}{S(x')\ S(y')} \qquad (10)$$

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

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

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

RANGO DEL COEFICIENTE DE CORRELACION

DE LA ECUACION CON QUE SE CALCULA LA VARIANCIA DE LA ESTIMACION

$$S^2_{y|x} = S^2(\underline{y})(1 - \rho^2_{xy}) \qquad (7)$$

SE CONCLUYE QUE $\rho^2_{xy} \leq 1$, YA QUE $S^2_{y|x} \geq 0$; EN CONSECUENCIA

$$-1 \leq \rho_{xy} \leq 1 \qquad (8)$$

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

$$\bar{x} = \frac{8100}{10} = 810 \text{ KG}; \quad \bar{y} = \frac{3.13}{10} = 0.313 \text{ CM}; \quad S^2_{xy} = \frac{49.20}{10} = 4.92$$

$$S^2(x) = \frac{204,000}{10} = 20,400; \quad S(x) = \sqrt{20,400} = 142,83$$

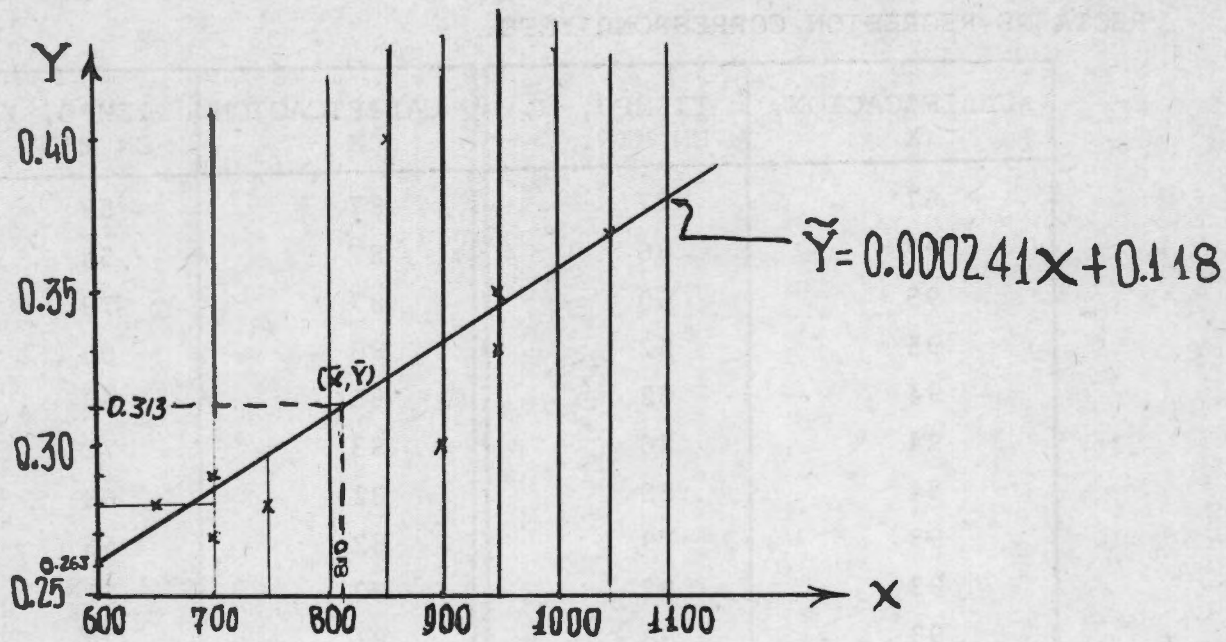
$$S^2(y) = \frac{0.020010}{10} = 0.002001; \quad S(y) = \sqrt{0.002001} = 0.0447$$

$$\rho_{xy} = \frac{4.92}{142.83 \times 0.0447} = 0.77$$

$$m = \rho_{xy}\, S(y)/S(x) = 0.77 \times 0.0447/142.83 = 0.000241 \text{ CM/KG}$$

$b = 0.313 - 0.000241 \times 810 = 0.118$ CM

$\tilde{y} = 0.000241\ x + 0.118$; SI $X = 600$, $\tilde{y} = 0.145 + 0.118 = 0.263$



$$S^2_{y|x} = S^2(y)(1 - \rho^2_{xy}) = 0.002001\ (1 - 0.77)^2 = 0.000106$$

$$S_{y|x} = 0.0103$$

$$S^2(y) = S^2(\tilde{y}) + S^2_{y|x} \rightarrow S^2(\tilde{y}) = 0.002001 - 0.000106 = 0.001895$$

$$S(\tilde{y}) = 0.0435$$

EL AGRIETAMIENTO QUE SE PREDICE,POR EJEMPLO, PARA UNA CARGA DE 1500Kg, SI EL COMPORTAMIENTO DE LA VIGA CONTINUA IGUAL, SERA

$\tilde{y} = 0.000241 \times 1500 + 0.118 = 0.480$ CM

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

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	

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					

$C_1 = 84.5;\ C_2 = 6;\ C_3 = 65.5;\ C_4 = 10$

$\bar{x}' = 11/50 = 0.22;\ \bar{y}' = 52/50 = 1.04$

$\bar{x} = 6 \times 0.22 + 84.5 = 85.82;$

$\bar{y} = 10 \times 1.04 + 65.5 = 75.9$

$S^2(x') = 73/50 - (0.22)^2 = 1.42;\quad S(x') = \sqrt{1.42} = 1.19$

$S^2(y') = 128/50 - (1.04)^2 = 1.49;\quad S(y') = \sqrt{1.49} = 1.22$

$S(x) = 1.19 \times 6 = 7.14;\ S(y) = 1.22 \times 10 = 12.2$

$$\rho_{xy} = \frac{-14/50 - 0.22 \times 1.04}{1.19 \times 1.22} = -\,0.35$$

$m = -0.35 \times 12.2 \ / \ 7.14 = -0.60$

$b = 75.9 - (-0.60)\ 85.82 = 127.39$

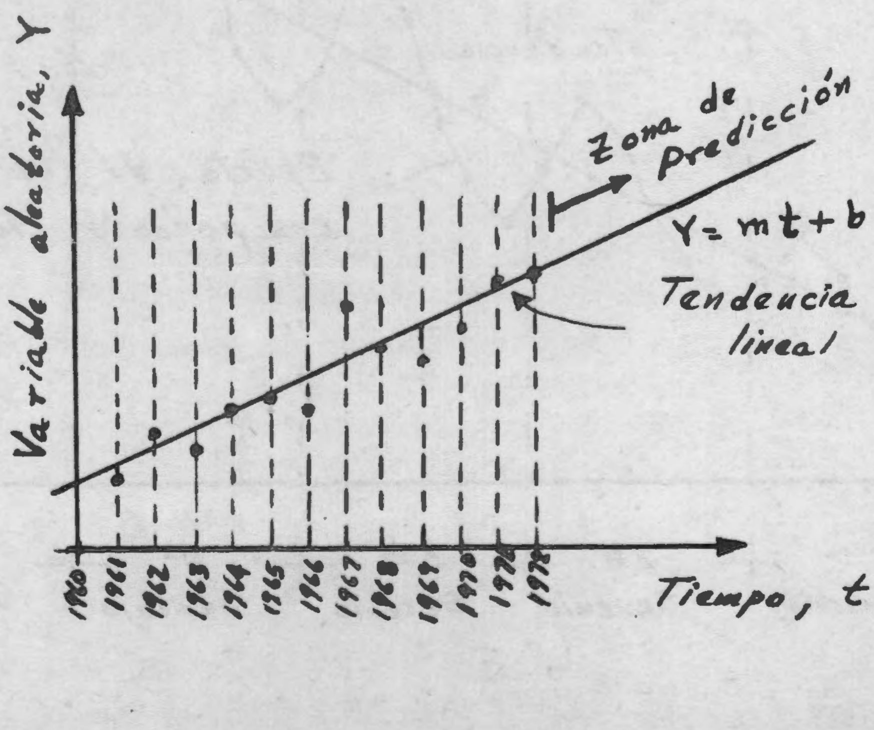
TAREA: TRAZAR EL DIAGRAMA DE CORRELACION Y LA RECTA DE REGRESION CORRESPONDIENTE, Y CALCULAR EL ERROR ESTANDAR DE LA ESTIMACION.

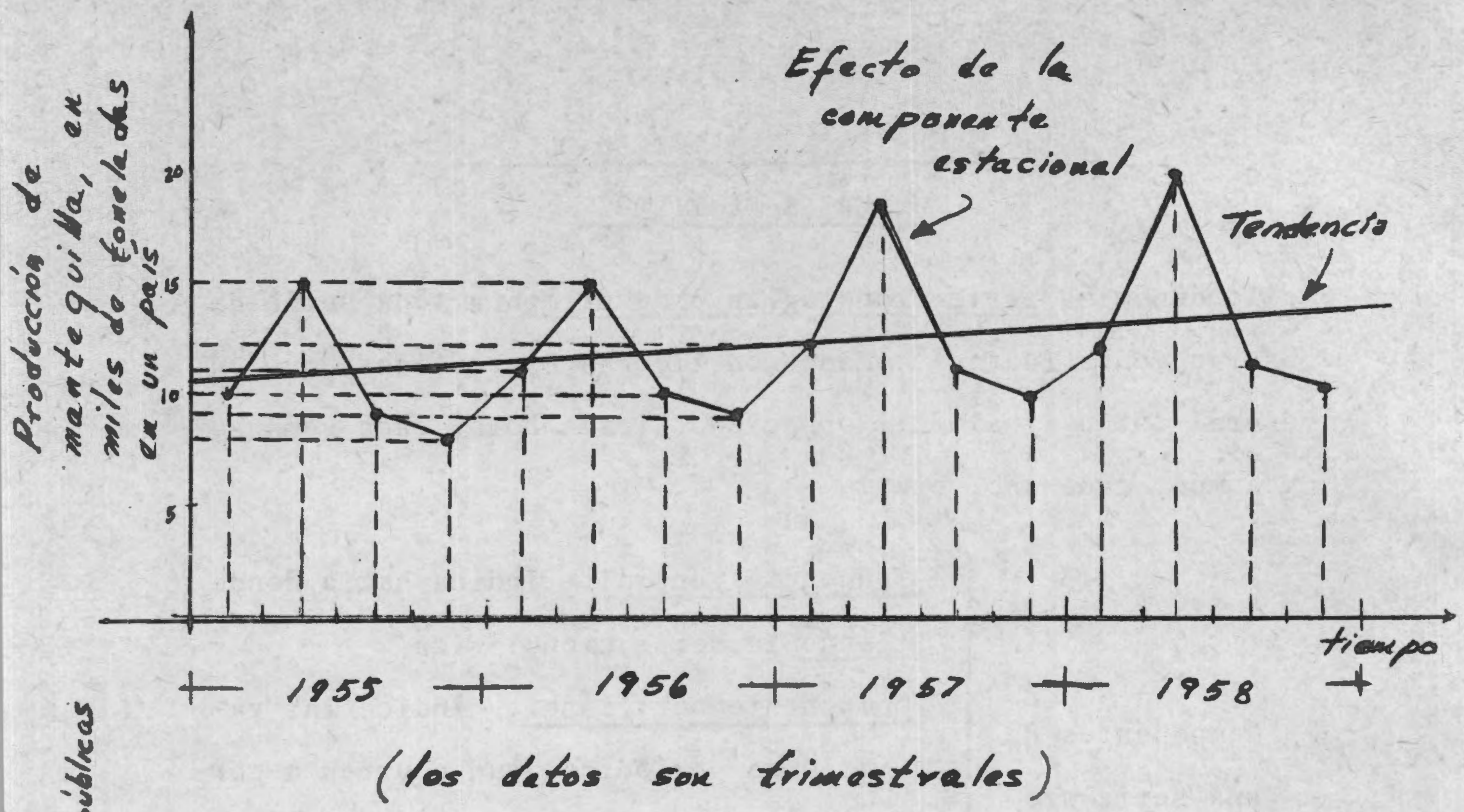
SERIES CRONOLOGICAS O SERIES DE TIEMPO

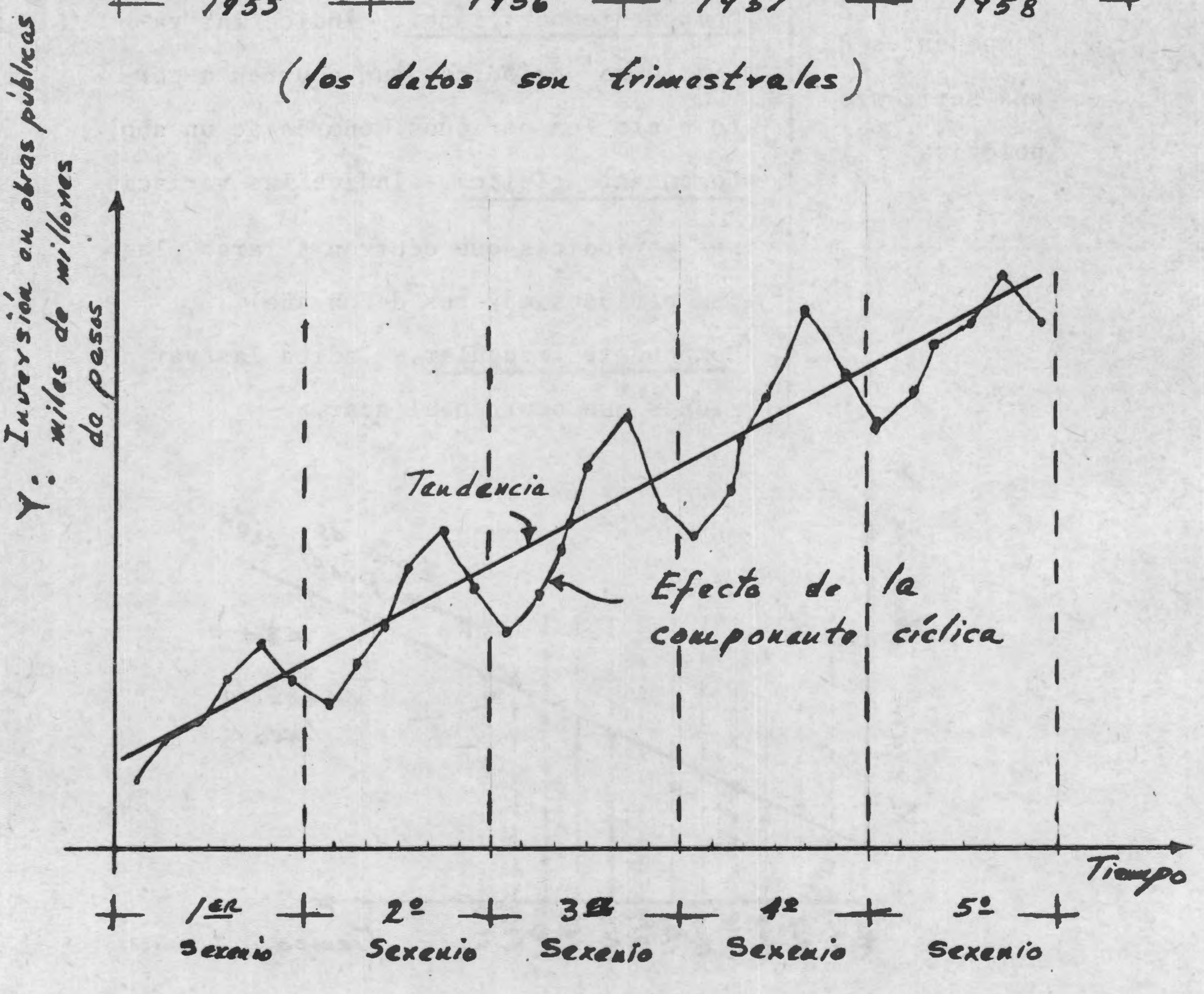
Se le denomina serie cronológica o de tiempo a toda serie de observaciones (datos) tomados en tiempos específicos, que en general están igualmente espaciados (cada hora, cada semana, cada mes, cada año, etc.)

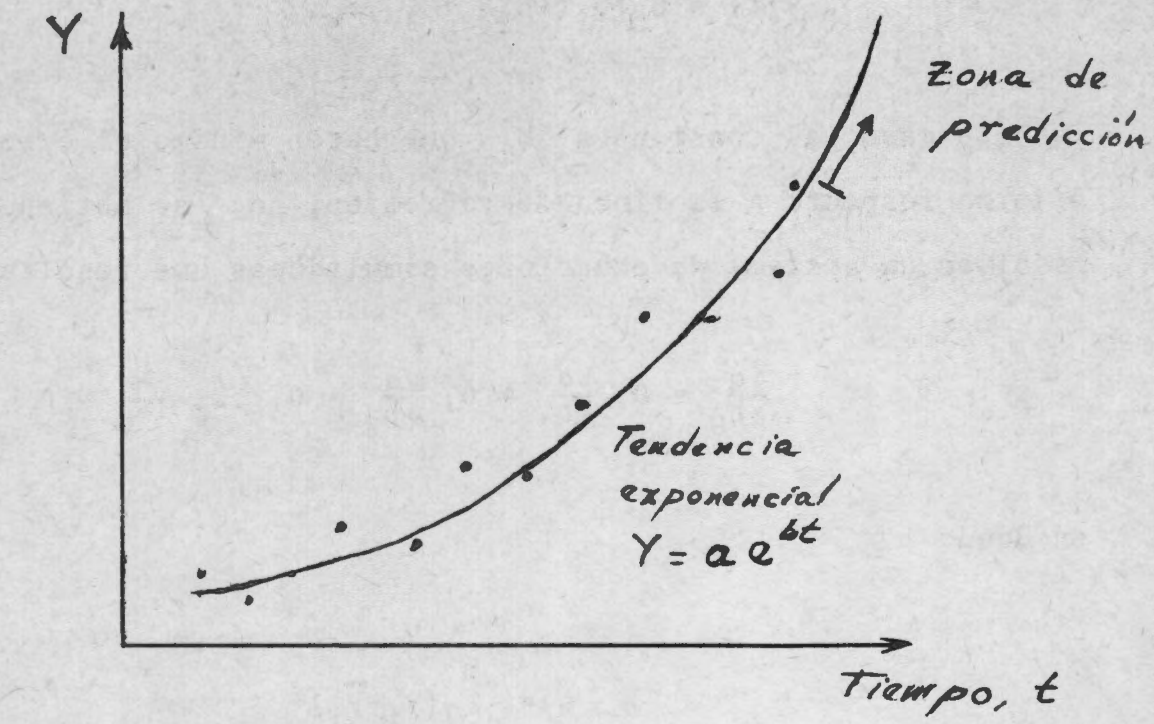
Componentes de una serie cronológica:

- 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.




Producción de mantequilla, en miles de toneladas en un país
Efecto de la componente estacional
Tendencia
tiempo
1955
1956
1957
1958
(los datos son trimestrales)


Y: Inversión en obras públicas miles de millones de pesos
Tendencia
Efecto de la componente cíclica
Tiempo
1er Sexenio
2º Sexenio
3er Sexenio
4º Sexenio
5º Sexenio



TENDENCIA GENERAL

Métodos de cálculo:
- Mínimos cuadrados
- Dos promedios
- Promedios móviles

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 <u>parabólico</u> $(M = 2)$, por ejemplo,

$$\tilde{y}(t) = b_0+b_1t+b_2t^2$$

$$\tilde{y}_i = \tilde{y}(t_i) = b_0+b_1t_i+b_2t_i^2$$

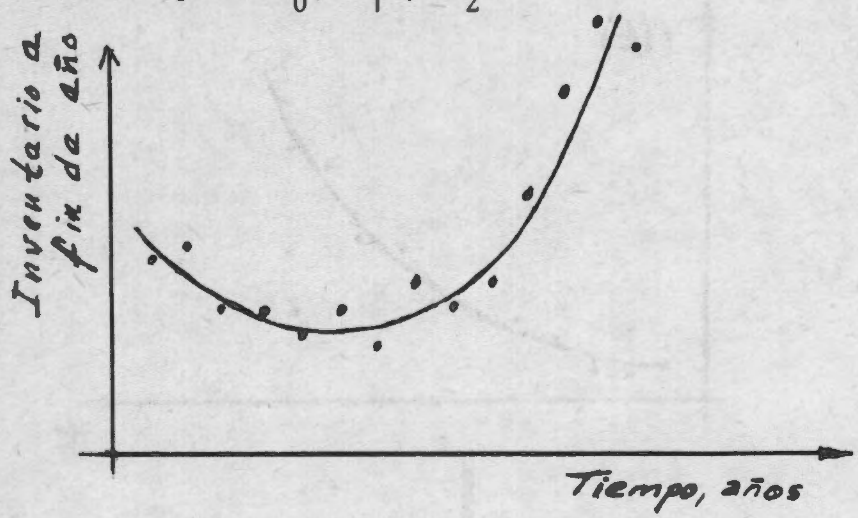
$$q = \sum_{i=1}^{n} (y_i-b_0-b_1t_i-b_2t_i^2)^2$$

$$\left.\begin{aligned}
\frac{\partial q}{\partial b_0} &= -2\sum_{i=1}^{n} (y_i-b_0-b_1t_i-b_2t_i^2) = 0 \\
\frac{\partial q}{\partial b_1} &= -2\sum_{i=1}^{n} (y_i-b_0-b_1t_i-b_2t_i^2)t_i=0 \\
\frac{\partial q}{\partial b_2} &= -2\sum_{i=1}^{n} t_i^2(y_i-b_0-b_1t_i-b_2t_i^2) = 0
\end{aligned}\right\}$$

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:

$$\left.\begin{array}{l} b_0 n + b_1 \Sigma t_i + b_2 \Sigma t_i^2 = \Sigma y_i \\ \\ b_0 \Sigma t_i + b_1 \Sigma t_i^2 + b_2 \Sigma t_i^3 = \Sigma t_i y_i \\ \\ b_0 \Sigma t_i^2 + b_1 \Sigma t_i^3 + b_2 \Sigma t_i^4 = \Sigma t_i^2 y_i^2 \end{array}\right\}$$

Y luego resolver para b_0, b_1 y b_2.



Cuando al observar la gráfica de Y contra t se concluye que es razonable ajustar una <u>función exponencial</u> de la forma

$$\tilde{Y}(t) = ae^{mt}$$

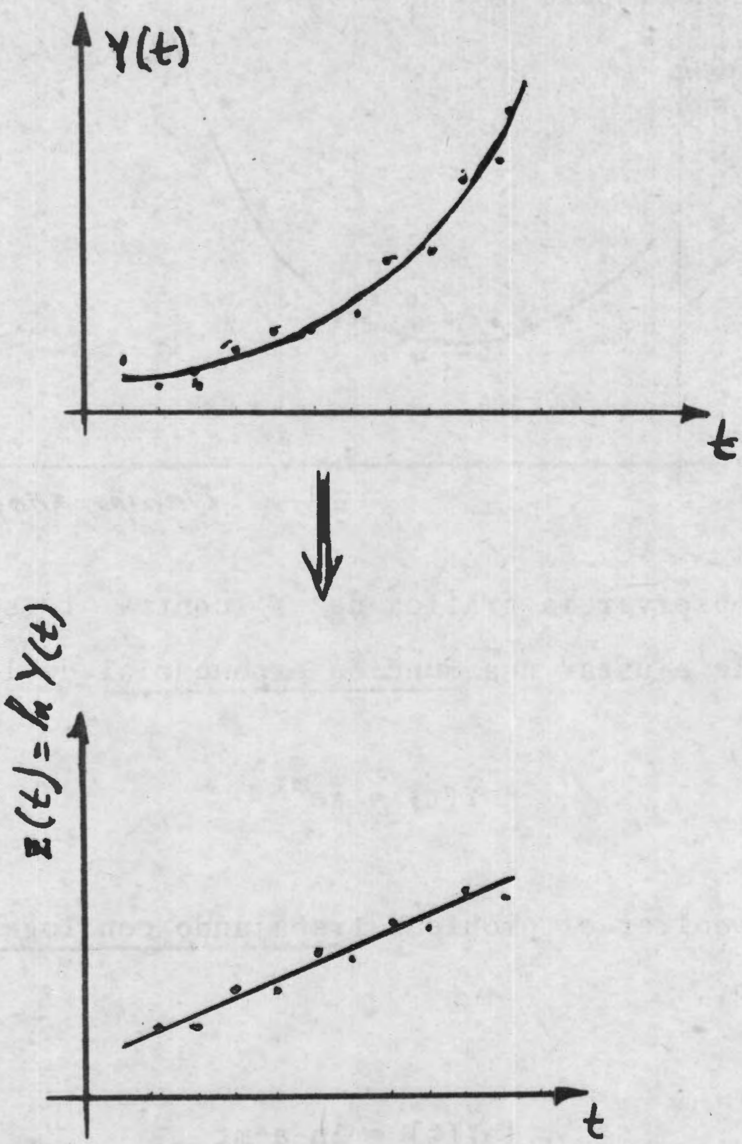
se puede resolver el problema <u>trabajando con logaritmos</u>, ya que, en tal caso,

$$\ln \tilde{Y}(t) = \ln a + mt$$

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 $\tilde{z}_i = \ln\tilde{Y}(t_i)$

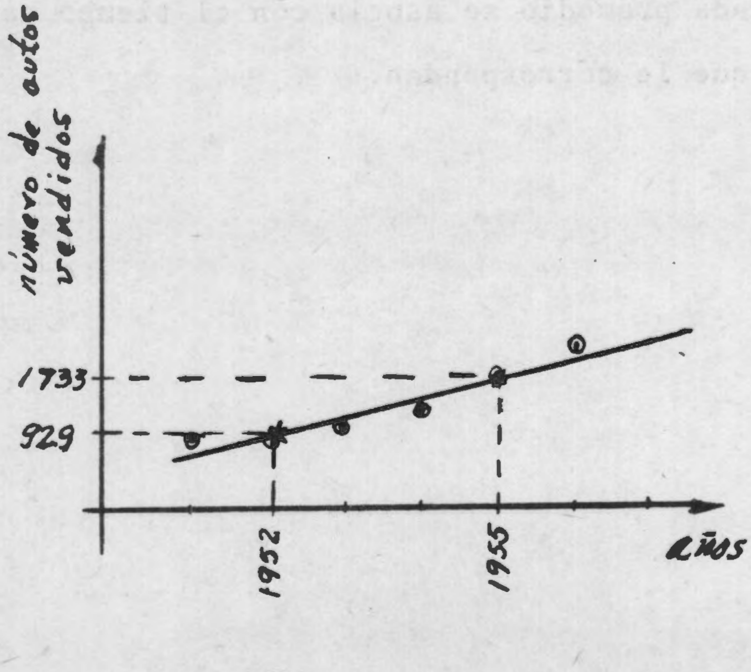


METODO DE DOS PROMEDIOS

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

$$\frac{y_1+y_2+\ldots+y_N}{N}, \quad \frac{y_2+y_3+\ldots+y_{N+1}}{N}, \ldots$$

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

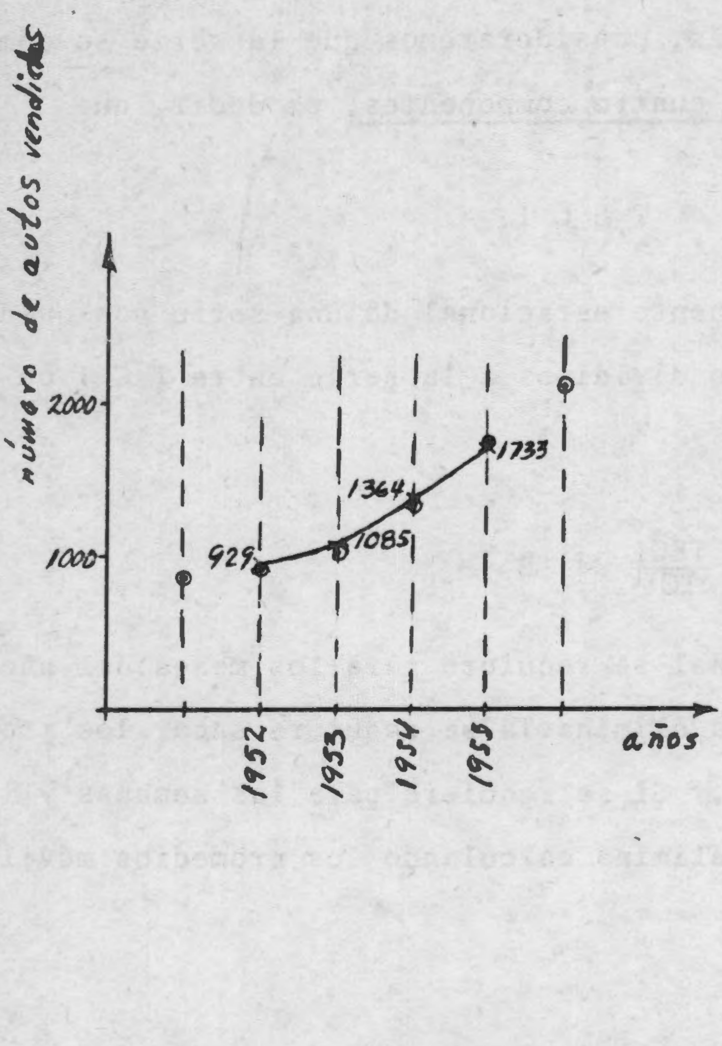
AÑO	Número de Autos vendidos	Promedio móvil de orden 3
1951	860	
1952	910	929
1953	1,018	1,085
1954	1,326	1,364
1955	1,749	1,733
1956	2,121	

$$\frac{860+910+1018}{3} = 929$$

$$\frac{910+1018+1326}{3} = 1,085$$

$$\frac{1018+1326+1749}{3} = 1,364$$

$$\frac{1326+1749+2121}{3} = 1,733$$



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:

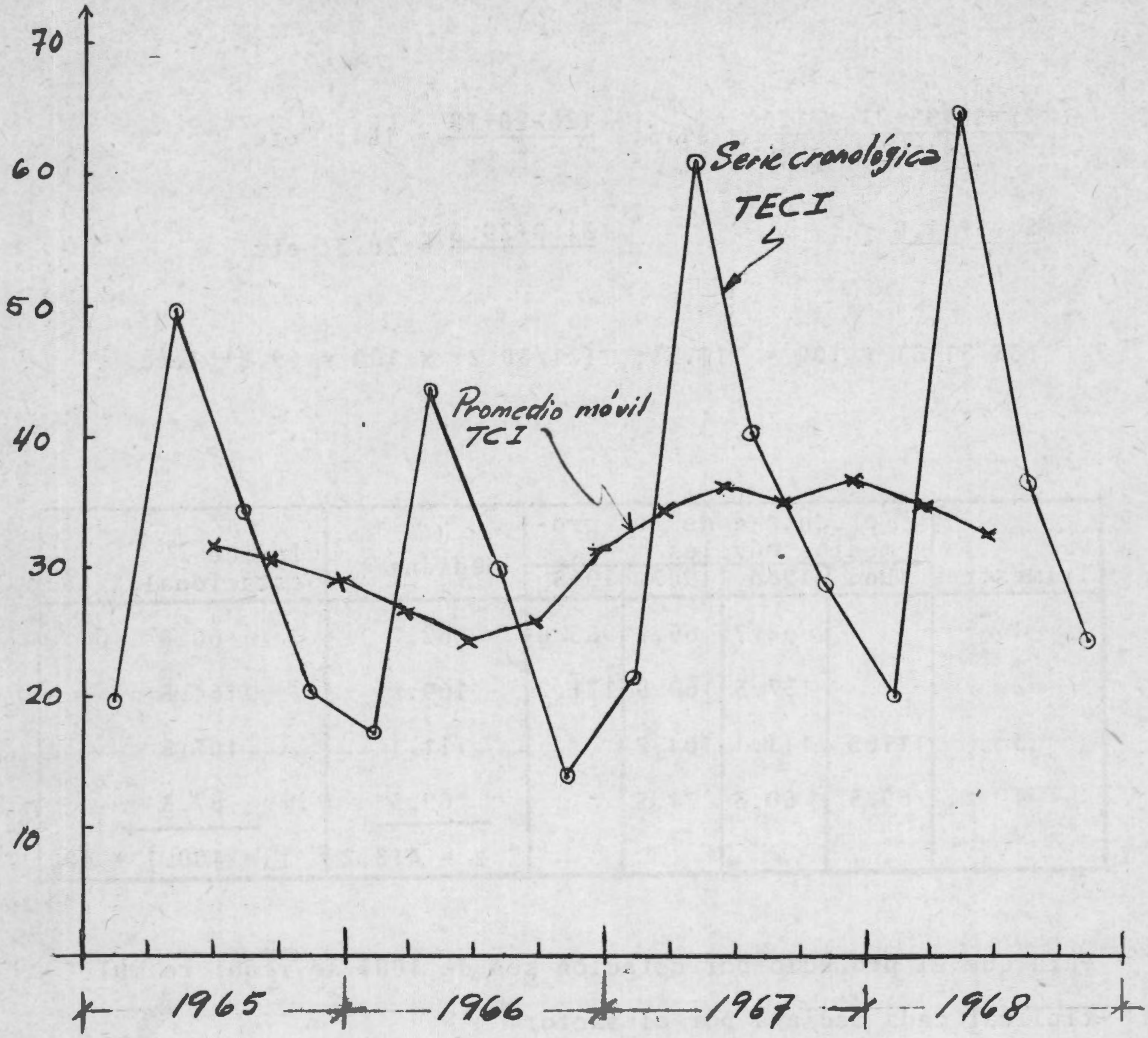
1. Cálculo de los promedios móviles (se obtiene TCI)
2. Cálculo de los porcentajes de los promedios móviles (se obtiene TECI/TCI = E)
3. Se calculan las medianas de los valores correspondientes a cada periodo (también se pueden usar los promedios aritméticos). Con esto se obtiene un valor representativo (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) x 100 = 111.5%; (21/30.2) x 100 = 69.5%; etc.

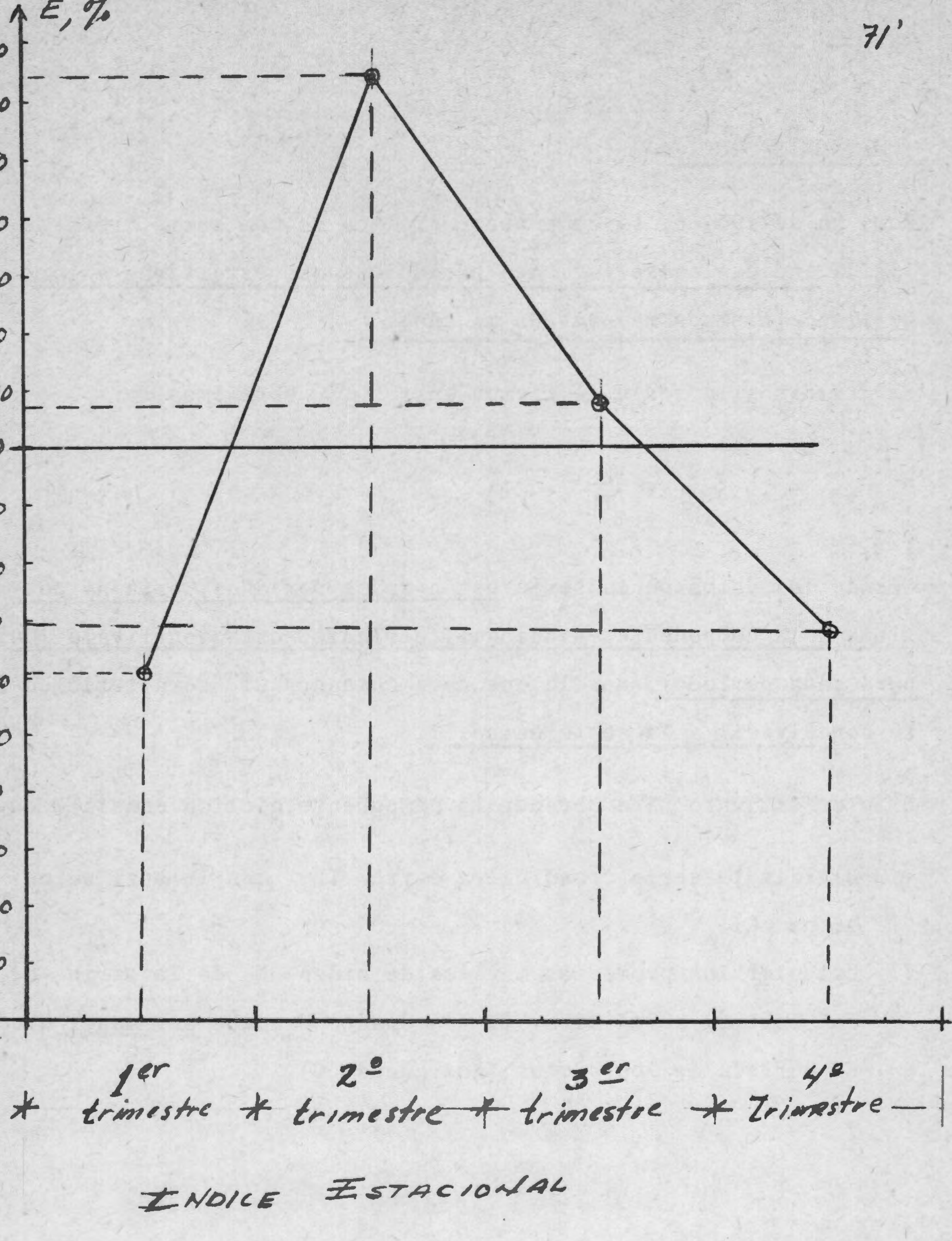
Trimestre	Porcentaje de los promedios móviles 1965	1966	1967	1968	Mediana	Indice estacional, %
1		62.7	66.2	55.8	62.7	60.8
2		157.5	169.9	171.2	169.9	164.5
3	111.5	111.1	104.2		111.1	107.5
4	69.5	50.3	74.5		69.5	67.3
					Σ = 413.2	Σ = 400.1 ± 400

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 x 0.968 = 60.8%; 169.9 x 0.968 = 164.5

111.5 x 0.968 = 107.5; 69.5 x 0.968 = 67.3


E, %
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
1er trimestre
2º trimestre
3er trimestre
4º Trimestre
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 menor que el periodo de los ciclos (nos queda C).

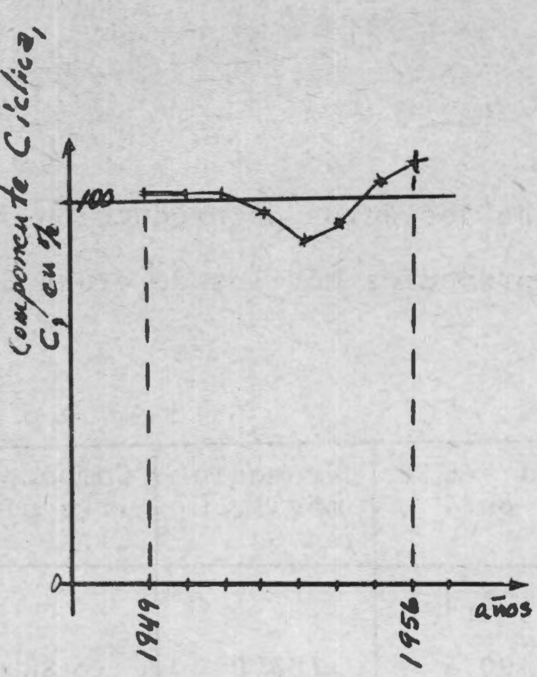
EJEMPLO

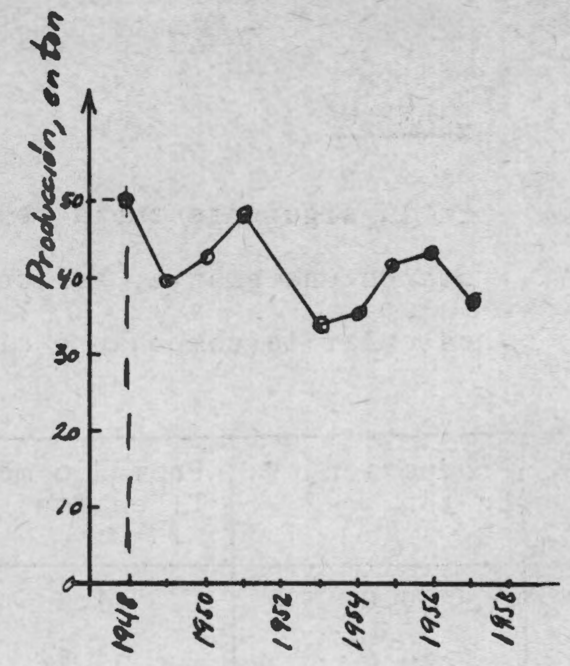
En la siguiente tabla se presentan los datos de producción de uva en una granja, así como sus promedios móviles de orden 5, calcular la componente cíclica.

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
1953	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 1	2	3	4	Promedio	Indice cíclico, en %
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.