Técnica Estadística general

(para luchas y profesiones sanitarias)

Cuadros sinópticos-nemotécnicos

Reducción de los cuadros murales utilizados en los Cursos de Diplomados Sanitarios de la Escuela Departamental de Barcelona

POR C. GARCÍA LUQUERO
MÉDICO DEL CUERPO DE SANIDAD NACIONAL

Clemente

BARCELONA, 1954

Reducción lineal 1/6
Reducción superficial 1/36

TÉCNICA ESTADÍSTICA GENERAL

ÍNDICE DE CONCEPTOS
(ORDEN DIDÁCTICO)

1. Diagrama múltiple: Leyes de Higiene.
2. Definiciones (I.).
3. Promedios (serie Ma.).
4. Variación gausiana. Desviación standard (Var.).
5. Representación gráfica:
   (Serie R. G.) - Diagrama polar (ciclo estacional). — Barras segmentadas (evol. mort. inf.). — Nomograma de probabilidades.
6. Concentración (serie Con.). Variación no gausiana.
7. Tendencia: Números índices. — Ajustamiento de curvas (serie Aj.).
8. Correlación (serie Corr.). — Tabla de concentración. — Líneas de regresión (suministro de agua y tifoidea hasta 1948).
9. Probabilidades (serie P.).
10. Curva de Gauss. Errores probables (serie G.).
11. Asociación: Tabla tetracórica. Test χ^2. (serie As.).
12. Tasas o coeficientes sanitarios (serie T.).
13. Estudio estadístico de un fenómeno demográfico (serie M. I.).
14. Distribuciones estadísticas y constantes (serie E, ep.) (para histogramas y curvas).

E. ep. 1/

Distribución por edades de seis enfermedades evitables.-(Depart. de Higiene de Maryland.)

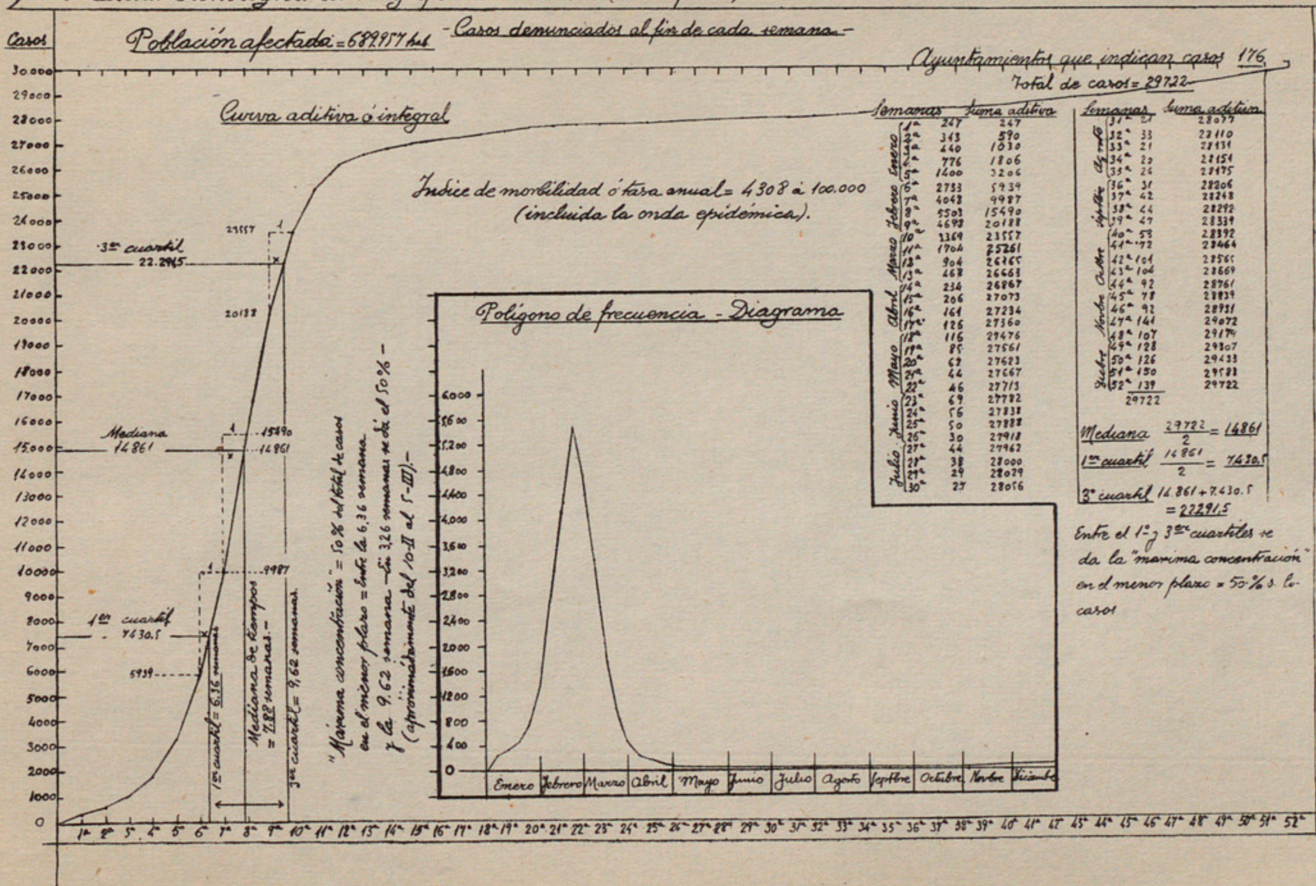
Grupos de Edades	Sarampión Casos	Sarampión Defund.	Tifoidea Casos	Tifoidea Defund.	Escarlatina Casos	Escarlatina Defund.	Difteria Casos	Difteria Defund.	Tos ferina Casos	Tos ferina Defund.	Gripe Casos	Gripe Defund.
1 año	1807	292	18	70	149	2011	301	913	1724	3442	73	164
1 año	3848	319	72	119	544	2562	1245	1857	1631	1516	181	163
2 años	4613	138	184	144	784	1117	1286	1781	1675	596	241	99
3 años	5213	75	316	149	1225	584	1715	1498	1603	301	315	49
4 años	5757	37	487	145	1556	302	1666	1293	1614	152	408	53
-5 años	21258	861	1077	627	4260	6576	6213	7342	8247	6007	1218	528
5 años	5513								1453			
6 años	6347								1348			
7 años	6224								1143			
8 años	5304								848			
9 años	3614	98	3396	759	7569	660	6125	3171	573	246	2221	139
10 años	2795								353			
11 años	1802								283			
12 años	1699								188			
13 años	1142								130			
14 años	1213	33	3659	825	3392	185	2578	704	87	40	2583	126
15 años	918								71			
16 años	947								51			
17 años	755.								33			
18 años	746								27			
19 años	736	26	3981	1544	991	161	1215	214	20		3210	167
20 a 25	2754	14	3892	1817	403	138	823	123	20-30 { —	20-30 { —	2507	300
25 à 30	1430	10	2874	1366	209	70	685	80	111	7	2110	419
30 à 40	727	16	3169	1741	190	109	656	123	80	6	3102	626
40 à 50	291	5	1612	1230	62	90	272	73	53	4	1425	207
50 à 60	90	4	724	754	8	47	95	42	29	3	497	98
60 y más	31	16	336	627	7	73	35	41	23	11	251	130
	66340	1083	24720	11290	17091	8109	18697	11913	15151	6324	19124	2840

(La morbilidad y mortalidad proceden de series distintas, no relacionables. 1920)

Constantes estadísticas		
1ª Sarampión	Casos	Defund.
Media aritmética	8,73 años	6,1 años
Mediana	7,94 años	1,78 años
Modo	6,52 años	1,26 años
Máxima concentración	3,62 à 8,37	0 à 3,41
2ª Fiebre tifoidea	Casos	Defund.
Media aritmética	22 años	28,5 años
Mediana	20,3 años	25,05 años
Modo	17,64 años	22,25 años
Máxima concentración	9,15 à 25,4	10 à 30,5
3ª Escarlatina	Casos	Defund.
Media aritmética	9,1 años	5,7 años
Mediana	7,6 años	1,76 años
Modo	6,42 años	1,32 años
Máxima concentración	3,27 à 9,7	0 à 1,8
4ª Difteria	Casos	Defund.
Media aritmética	10,61 años	5,9 años
Mediana	7,15 años	3,9 años
Modo	3,61 años	1,32 años
Máxima concentrac.	0 à 9,6	0 à 3,94
5ª Tos ferina	Casos	Defund.
Media aritmética	5,6 años	1,7 años
Mediana	4,71 años	0,81 años
Modo	2,49 años	0,45 años
Máxima concentrac.	0,5 à 5,2	0 à 0,9
6ª Gripe	Casos	Defund.
Media aritmética	22,35 años	25,4 años
Mediana	18,95 años	25,7 años
Modo	17,25 años	24,5 años
Máxima concentrac.	8,7 à 26,7	12,3 à 29,0

Rg.3

Marcha cronológica de la gripe en Barcelona (sin capital) durante el año 1953.- Constantes estadísticas.-



Semanas		Suma aditiva
Enero 1ª	247	247
2ª	343	590
3ª	440	1030
4ª	776	1806
5ª	1400	3206
Febrero 6ª	2733	5939
7ª	4048	9987
8ª	5503	15490
9ª	4698	20188
Marzo 10ª	3369	23557
11ª	1704	25261
12ª	904	26165
13ª	468	26663
Abril 14ª	234	26867
15ª	206	27073
16ª	161	27234
17ª	126	27360
Mayo 18ª	116	27476
19ª	85	27561
20ª	62	27623
21ª	44	27667
22ª	46	27713
Junio 23ª	69	27782
24ª	56	27838
25ª	50	27888
26ª	30	27918
Julio 27ª	44	27962
28ª	38	28000
29ª	29	28029
30ª	27	28056

Semanas		Suma aditiva
Agosto 31ª	21	28077
32ª	33	28110
33ª	21	28131
34ª	20	28151
35ª	24	28175
Septbre 36ª	31	28206
37ª	42	28248
38ª	44	28292
39ª	47	28339
Octbre 40ª	53	28392
41ª	72	28464
42ª	101	28565
43ª	104	28669
44ª	92	28761
Novbre 45ª	78	28839
46ª	92	28931
47ª	141	29072
48ª	107	29179
Dicbre 49ª	128	29307
50ª	126	29433
51ª	150	29583
52ª	139	29722
	29722	

Mediana $\frac{29722}{2} = 14861$

1er cuartil $\frac{14861}{2} = 7430.5$

3º cuartil $14.861 + 7.430.5 = 22291,5$

Entre el 1º y 3er cuartiles se da la "maxima concentración" en el menor plazo = 50 % de los casos

Con. 3)

Concentracion – Variación no gausiana – Diferencia media – Indices –

1. Diferencia media en serie aditiva: Mortalidad por tuberculosis (T.T.) – Barcelona 1945

Óbitos por edades en graduatoria aditiva – Series inversas		Diferencias sin signos		Factor de posicion	Productos (p) (a × b)	
167	1621	(a) 1454	1 a 12	11 (b)	15994	
251	1383	1132	2 a 11	9	10188	
342	1217	875	3 a 10	7	6125	
442	1060	618	4 a 9	5	3090	
554	917	363	5 a 8	3	1089	
672	792	120	6 a 7	1	120	
792	672	120	7 a 6	1	120	36606
917	554	363	8 a 5	3	1089	
1060	442	618	9 a 4	5	3090	
1217	342	875	10 a 3	7	6125	
1383	251	1132	11 a 2	9	10188	
1621	167	1454	12 a 1	11	15994	36606
9418					Σp ---	73212

9418 : 12 = Ma = 784,83

2Ma = 1569.66

Diferencia media; $\Delta = \Sigma p : n(n-1)$; $\Delta = 73212 : 132$; $\Delta = 554.6$

Relación de concentracion $R_c = \Delta : 2\,Ma$; 554,6 : 1569,6; $R_c = 0.353$

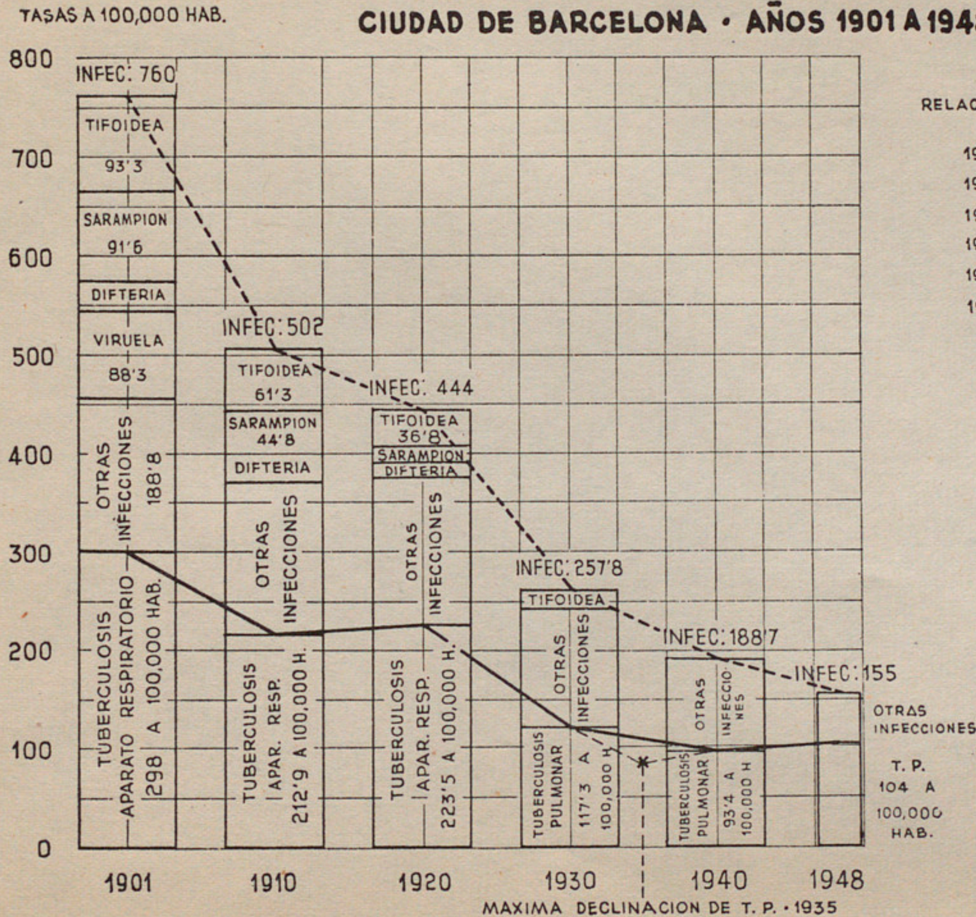
Variacion procentual $n\Delta : \Sigma f = 6655.2 : 9418$ ——— $I = 0.70$

G.4/ Concepto del error probable.- Error standard.- Significación en probabilidades del azar.- Aplicaciones.-

Aplicada la σ al estudio de las probabilidades gaussianas ó de distribuciones de frecuencias normales e infinitas, abre el campo de la teoría de errores.- En la curva de Gauss, hemos indicado la importancia de esta unidad de medida, que tomada en el eje de abcisas desde el pié de la Ma, marca las ordenadas que dividen en fracciones el area unitaria total, de probabilidades de la curva. Una σ á cada lado de la media en el eje, señala las ordenadas que incluyen el 0,3413 de cada mitad del área, las dos (±) suman 0,6826, la parte mayor de toda el área (68,26%).- Dos σ á cada lado de la Ma, limitan el área de 0,4772 de cada mitad, sumando en total (±), el 95,45%.- Tres σ á cada lado del pié de la Ma marcan el límite de 0'4987 en cada mitad, sumando en toda el área el 99,75%, límite practico del total; como la curva es asintótica, mas σ á cada lado, aumentarán en fracción despreciable este margen, hasta ∞, donde está la unidad, 100%. En probabilidades, la desviación standard se llama "error standard" y se expresa tambien σ.— Un dato cualquiera de un fenómeno, siempre cae dentro de la curva (= 100% = certeza); tiene la probabilidad del 34,13% de caer entre la Ma y la σ de un lado y la del 68,26% de caer entre el margen de la Ma y ± σ.— Caerá un 13,59% de los casos, entre los valores de 1 a 2σ; podrá caer solamente el 2,15% de los casos entre valores de 2 a 3σ, á uno de los lados de la Ma. La influencia del azar, acerca los valores fortuitos á la Ma, es allí tan fuerte su actuación, que no se estiman como de especifica influencia causal, los datos obtenidos en la investigación; los dá el azar por sus propias leyes; no tienen "significacion alguna".- Solo se estiman como significativos, aquellos valores que exceden en 2 ó mas σ, la Ma, pues en esa zona es remota la influencia del azar. Cualquier cifra que se obtenga en una investigación debe tener una variación de ± 2σ; solo así se le puede ó no, dar valor de causalidad. Es incompleto todo estudio numerico que no lleva anejo este cálculo.-

Si la certeza es el área total de la curva, debe conocerse en probabilidades, la situacion fija de la indiferencia. Es el punto en que los valores al crecer ó disminuir desde la Ma, pierden esta representación: es el limite en que la probabilidad es tanta en favor del promedio, como en contra; cualquier dato fortuito, tanto puede incidir dentro como fuera de dicho margen ó límite (Probabilidad total 50%; á cada lado de la Ma = 25%). Esta medida de desviación se llama "Error probable" (Pe).- Como la medida de las probabilidades se calcula por las ordenadas trazadas sobre medidas de desviación unitaria standard en el eje de abcisas (1σ, marca a cada lado de la Ma, el 34,13% de probabilidad), para fijar un 25%, no pudiéndolo precisar por proporciones, dada la forma irregular de la curva. Acudimos á las tablas de probabilidades de las fracciones de dicha area, resueltas, como áreas de curva, por integrales en cada fraccion de σ, interpolando, por diferencias, entre los valores contiguos de probabilidad. Las tablas dan: Probabilidad de 24,857% con desviación de 0,67σ; probabilidad de 25,175% con desviación de 0,68σ; como buscamos la desviacion correspondiente á 25% de probabilidad tendremos: $\frac{\text{Dif. de P. en la tabla}}{\text{Dif. de } \sigma \text{ en la tabla}} = \frac{\text{Dif. de P. menor, hasta } 25\%}{x}$; $\frac{0.318\,\%}{0.01\,(\sigma)} = \frac{0.143}{x}$; $x = \frac{0.00143}{0.318} = 0,4498$, factor adicional al menor (0,67), que nos da la desviación en abcisas de $Pe = 0,6745\sigma$.— Equivalencias con errores standard: $\pm 1 Pe = 0,6745\sigma$; probabilidad $= \pm 25\%$. $\pm 2 Pe = \pm 1,35\sigma$; probabilidad $= \pm 41,15\%$. $\pm 3 Pe = \pm 2,023\sigma$; probabilidad $= \pm 47,84\%$.—

EVOLUCION DE LA MORTALIDAD INFECCIOSA EN LA CIUDAD DE BARCELONA · AÑOS 1901 A 1948



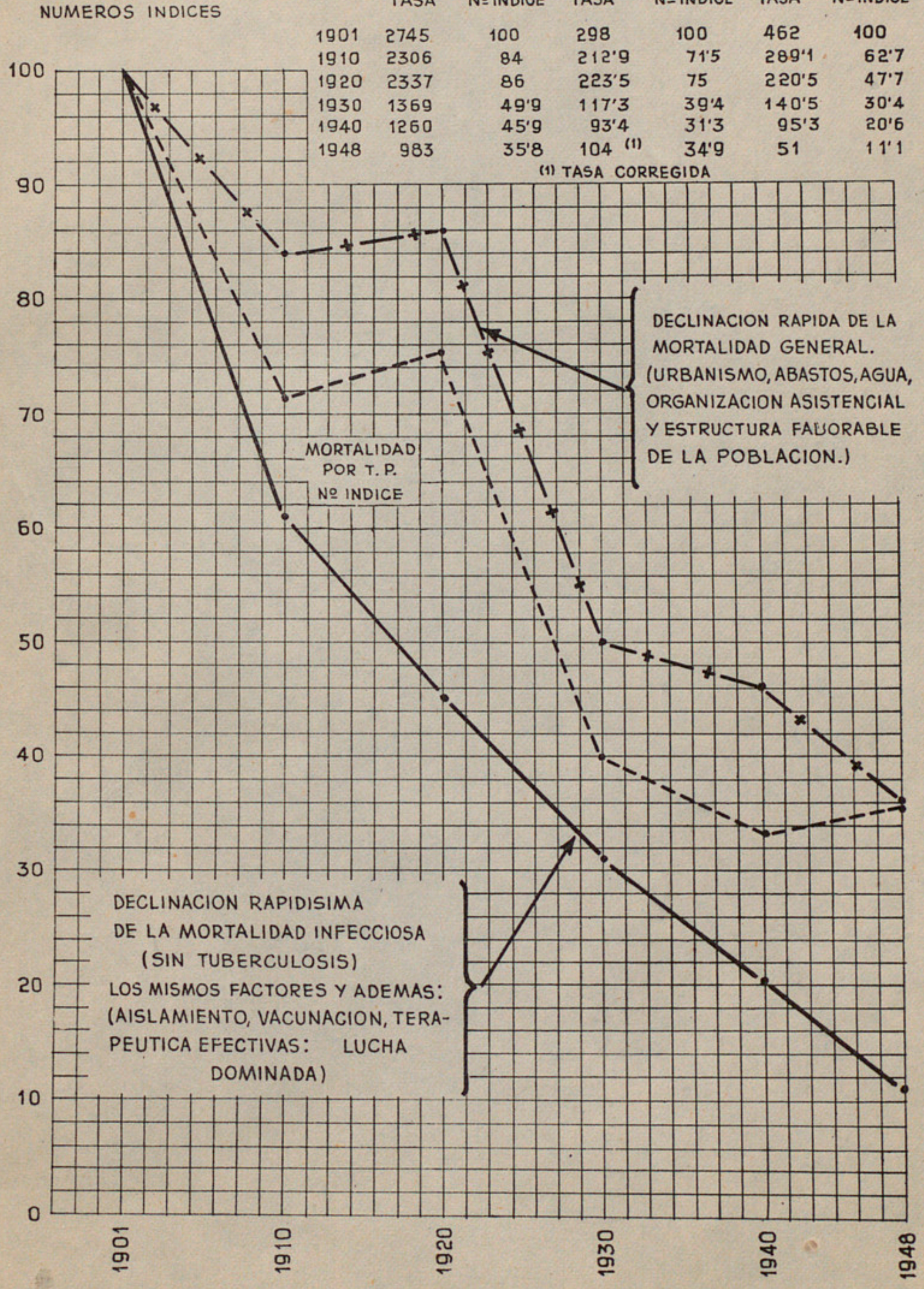
RELACION DE LA T. P. AL TOTAL DE INFECCIONES

1901 = T. P. = 39'2 %	DEL	TOTAL
1910 = T. P. = 42'3 %	"	"
1920 = T. P. = 50'3 %	"	"
1930 = T. P. = 45'5 %	"	"
1940 = T. P. = 49'4 %	"	"
1948 = T. P. = 67'2 %	"	"

EVOLUCION COMPARADA ENTRE MORTALIDAD GENERAL, TUBERCULOSA Y OTRAS INFECCIONES · BARCELONA 1901-1948

	MORTALIDAD GENERAL		MORTALIDAD POR T. P.		MORTALIDAD POR OTRAS INFECCIONES	
	TASA	Nº INDICE	TASA	Nº INDICE	TASA	Nº INDICE
1901	2745	100	298	100	462	100
1910	2306	84	212'9	71'5	289'1	62'7
1920	2337	86	223'5	75	220'5	47'7
1930	1369	49'9	117'3	39'4	140'5	30'4
1940	1260	45'9	93'4	31'3	95'3	20'6
1948	983	35'8	104 (1)	34'9	51	11'1

(1) TASA CORREGIDA



T) Tasas ó coeficientes demográfico - sanitarios.-

- Relacion del número de nacidos, fallecidos, ó casos de enfermedad a cuantía fija de población = Tasa
  Tasa "cruda" (rate) = Referida á poblacion sin discriminar clase — "Corregida", si se modifican factores de error.
  "Tasa específica" ó de edades, en que los casos y la poblacion se toman solo en ciertas edades.- "Standardizada" = Al millón standard.
- Tasas crudas más usuales: Simple proporción: Si en toda la población, se dan un número de casos, en 100.000 hab. = Tasa á 100.000

1ª Natalidad á 1000 hab.	2ª Mortalidad á 1000 hab.	3ª Nupcialidad á 1000 hab.	4ª Mortalidad infantil.— a) <1 año á 1000 nacidos vivos	b) De 1 a 4 años á 1000 fallecidos totales	c) 5 á 14 años á 1000 falled. totales	5ª Mortinatalidad
= Nacidos totales x 1000 / Total población	= Obitos totales x 1000 / Total poblacion	= Matrimonios x 1000 hab / Total población	= Defunc. < 1 año x 1000 n v / total nacidos vivos	= Defunc. 1 y 4 años x 1000 / total defunciones	= Defunc. 5 á 14 años x 1000 / Total defunciones	Nacidos muertos a 100 — = Nacidos muertos y abortos x 100 / Nacidos vivos y muertos
Barcelona (c). 1951.	Barcelona (c). 1951.	Barcelona (c.) 1951.	Barcelona (c.) (1951).	Barcelona (c.) 1951.	Barcelona (c) 1951.	Barcelona (c) - 1951
Nacidos vivos = 18530 Población 1951 = 1.288.283 Tasa = 14,47 ‰	Fallecidos total = 15.117 - Pob. = idem (1951). Tasa = 11.80 ‰	Matrimonios = 10.063 Pob. id (1951) Tasa 7.86 ‰	Defunc. de < 1 año 832 Nacidos vivos = 18530 Tasa de < 1 año = 44.90 ‰ n.v.	Defunc. 1 y 4 a. = 440 Defunc. en total = 15117 Tasa = 29.1 ‰ d. (1951).	Defunc. 5-14 a. 1272 Defunc. total = 15117 Tasa = 84.1 ‰ d. (1951)	Nacidos muertos, muertos antes de 24 horas y abortos inscritos 777 Nacidos vivos y muertos 19307 Tasa 4,02 % n.v.y m (1951)
Madrid = 19,19 (1949)	Madrid = 10,52 (1949)	Madrid 1950 = 8,10 ‰				

- Tasas de enfermedades diversas, en general á 100.000 hab.

Tuberculosis T.T.	Cancer y malignos	Cardiopatías.
Obitos = 2.132 Poblacion Barcelona (p.) 1950 = 2.232.119 hab. Tasa a 100.000 = 95.5	Obitos = 1706 Barcelona 1950 (p). Tasa a 100.000 = 76.4	Obitos = 4396 Barcelona (p.) 1950 Tasa 100.000 = 196.7
Madrid 1950 = 112,80 %ooo Barcelona 1952 = 56,40 %ooo Norte America (1953) = 10,2 %ooo Tasa de enfermos nuevos 55 %ooo 1952 - Norte America	Madrid 1950 (p) Tasa a 100.000 = 88,0 Barcelona 1952 = 128,10 %ooo (con Hodgkin) (capital)	Madrid 1950 (p.) Tasa á 100000 = 153,2 Corazón y vasos. Barcelona (c) 1948 = 203,9.— 1949 = 254,0. 1950 = 242,3.— 1951 = 301,5. 1952 = 260,4.

Mortalidad neonatal (< 30 días)

1949 (EE.UU) - Tasas á 1000 nac. vivos por año.

Los fallecidos el 1er día, darian al año 1143 ‰ n.v.
Los fallecidos la 1ª semana " " " 135,8 ‰ "
Los fallecidos en la 4ª semana " " 34.5 ‰ "

a) Tasa = $\frac{\text{Muertos 1ª semana} \times 52 \text{ semanas}}{\text{num.º total nacidos vivos año}} \times 1000$ =

b) " = $\frac{\text{Fallecidos 1er mes} \times 12 \times 1000}{\text{total anual nacidos vivos}}$ =

Barcelona, tasa b en 1952 = 197,03 ‰.

Infecciosas, excepto "T.T." (mortalidad)

Obitos entre población, por 100.000 hab.
Barcelona (c.) 1951 = 40.7
Barcelona (c.) 1952 = 15,4

Tasas de toda España

- 1950 .

Natalidad = 19,76 ‰
Mortalidad = 10,64 "
Nupcialidad = 7,39 "
Mortalidad < 1 año = 64.20

1951: Natalidad = 19.96
Mortalidad = 11,48
Incremento veget. = 8.47

Con 1) Tabla de concentración por edades (mortalidad) de la tuberculosis.- Barcelona 1945-

– Todas formas de tuberculosis –

Edades = Unidad mínima de grupo = 5 años

Grupos de edades	Obitos o frecuencias	Grupos unitarios	Concentración por grupo unitario	Número de orden	Ordenación de la serie según número	Sumas aditivas	Porcentaje aditivo
0 a 5 años	100 defunc.	1	100	4	167 - 60 á 90 a.	167	10,30 %
5 a 15 años	84 "	2	42	2	84 5 a 15 a.	251	15.48 %
15 a 20 años	157 "	1	157	10	91 55 a 60 a.	342	21.10 %
20 a 25 años	238 "	1	238	12	100 0 a 5 a.	442	27.28 %
25 a 30 años	166 "	1	166	11	112 50 á 55 a.	554	34.19 %
30 a 35 años	125 "	1	125	8	118 45 á 50 a	672	41.40 %
35 a 40 años	143 "	1	143	9	120 40 á 45 a	792	48.85 %
40 a 45 años	120 "	1	120	7	125 30 á 35 a.	917	56.50 %
45 á 50 años	118. "	1	118	6	143 35 á 40 a	1060	65.40 %
00 a 55 años	112 "	1	112	5	157 15 á 20 a	1217	75.50 %
55 a 60 años	91. "	1	91	3	166 25 á 30 a	1383	85.65 %
60 a 90 años	167. "	6	27	1	238 20 a 25 a	1621	100,00 %
	1621 defunc.				1621		

46.)

Errores probables de tasas sanitarias, de coeficientes estadísticos.- Errores ordinarios.-

1º Error probable de tasas demográfico-sanitarias. Desde Poissons se aplica á las tasas la fórmula del Pe. de porcentajes, para darles la variación del azar: $Pe = 0{,}6745\sqrt{\frac{v.m}{n}}$, (v = vivos, m = muertos) - Aunque n, es siempre la cuantía de la muestra, la referencia de las tasas a 1000, 10.000, ó 100.000 habitantes, es número ya alto que diluye la dispersión de las muestras pequeñas y no precisa corrección - En natalidad, m expresa los nacidos; en infecciones los enfermos (morbilidad); en mortalidad, los fallecidos.- V. expresa siempre los habitantes que restan del total n, al deducir m.

Ej: Probable error de una tasa media de mortalidad por todas causas, de valor 18.5 por 1000 hab:

$Pe = 0{,}6745\sqrt{\frac{981{,}5 \times 18{,}5}{1000}} = 2{,}86$; su valor triple 8.58, aplicado en mas y en menos a la tasa media, nos da la variación máxima que puede dar el azar en series amplias: De 9.92 á 27.08 ‰

2º Error probable de los coeficientes y constantes. Deducidos por calculo integral (tablas de Pearson, Fisher, etc). Pueden aplicarse en el triple de su valor, en mas ó en menos, á los coeficientes respectivos, para darles la máxima variabilidad del azar y estimar el valor significativo de estos.

— $Pe.$ de la $Ma = \pm 0{,}6745\sqrt{\frac{\sigma}{n}}$ = "E Ma".

— $Pe.$ de la $\sigma = \pm 0{,}6745\sqrt{\frac{\sigma}{2n}}$ = "E σ".

— Pe del coef. de correl. $= \pm 0{,}6745 \frac{1-r^2}{\sqrt{N}}$ = "Er".

— $Pe.$ de la mediana $= \pm 1{,}25332 \cdot 0{,}6745\sqrt{\frac{\sigma}{n}}$ = "E Md."

— Suma ó diferencia de dos variables independientes:

$Pe.\,dif: \pm 0{,}6745\sqrt{\sigma^2 + \sigma'^2}$; $\sigma\,dif = \pm\sqrt{\sigma^2 + \sigma'^2}$

— Suma ó diferencia de dos variables correlacionadas:

$\pm\, 0{,}6745\sqrt{\sigma^2 \pm 2r\sigma.\sigma' + \sigma'^2}$

— $Pe.$ del coef. de inclinación $= \pm\, 0{,}6745\sqrt{\frac{6}{N}}$ = "Es"

— $Pe.$ del coeficiente de variabilidad

1º: $v < 10$; $\pm\, 0{,}6745 \frac{v}{\sqrt{2N}}$ = "Ev".

2º: $v > 10$; $\pm\, 0{,}6745\sqrt{\frac{1+2\left(\frac{100}{v}\right)^2}{2N}}$ = "Ev".

Errores ordinarios: Previsibles é imprevisibles.-

Modo de corregirlos: 1º Perecuación: Es una forma de interpolación. Consiste en sustituir los grupos irregulares de datos en las series, por el promedio de cada dos contiguos; haciéndolo con tres seguidos, se obtiene con los promedios, una serie mas perfecta que salva errores de la primitiva (censos por edades, etc.)

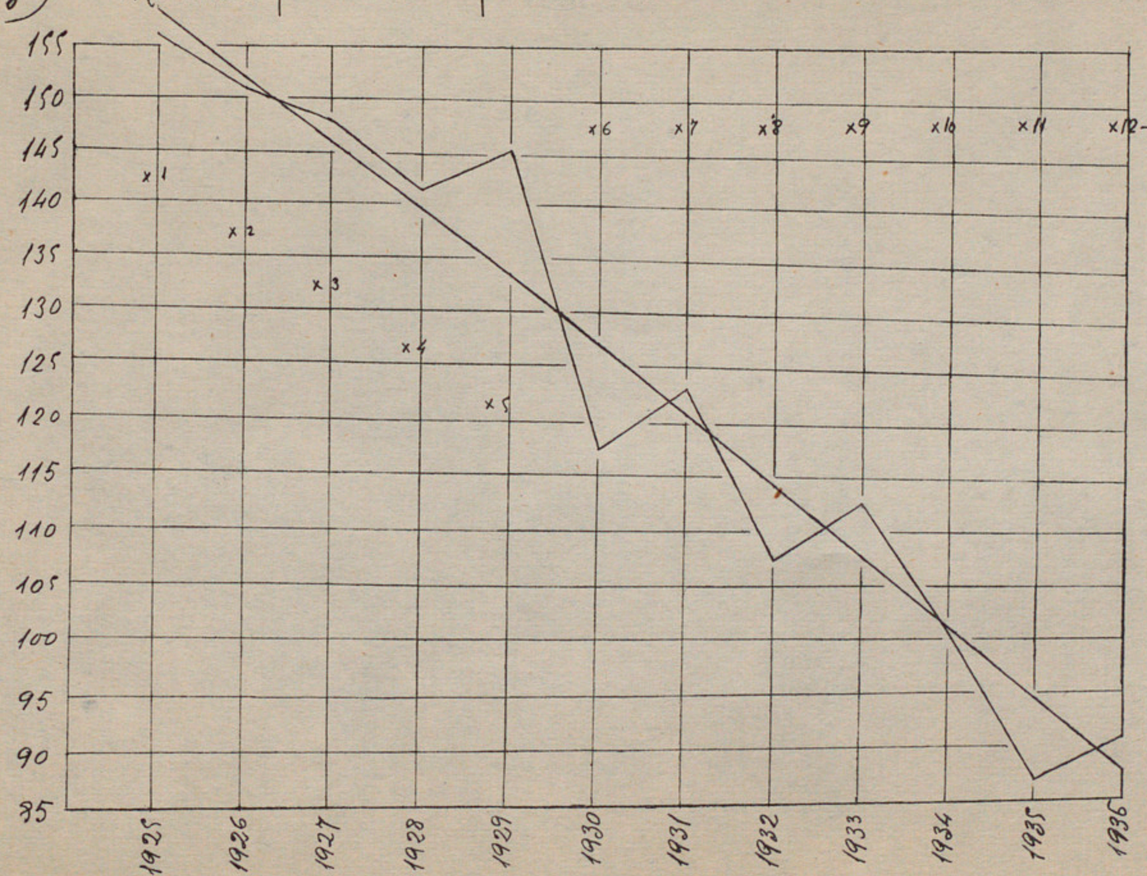
2º Nivelación de diferencias en graduatorias: Cuando un grupo sucesivo de valores (3 o 4) presentan saltos crecientes, disformes, se reparten sus diferencias entre los afectados, buscando un promedio conjunto para anadir a cada uno, de este modo:

Diferencia del primero al último, dividida entre $n-1$ términos (para crecimientos progresivos disformes).

3º Interpolación - Principalmente por diferencias (primera, segunda y tercera); cuando falta un termino, se busca por proporciones entre la diferencia de los próximos existentes, buscando el intermedio proporcional (x) respecto de la situación en que deseamos colocar el valor que falta (V. mediana de edades).

4º Ajustamiento: Ajustando diversas curvas a las series, se obtienen valores intermedios, que siguen las leyes de toda la serie.—

Ag. 3) Mortalidad por Tuberculosis pulmonar en Barcelona de 1925 à 1936. - Su Tendencia recta -



Var.1

Variación en series normales ó gausianas - Desviación standard ó típica.-

1.° Desviación media. Con los promedios se representan protocolos ó estudios individuales de observaciones ó fenómenos.

La necesidad de comparar investigaciones de diversos operadores, que salven errores personales, de tecnica, material, etc, para generalizar deducciones, obteniendo las leyes que regulan los fenómenos en universos amplios (finalidad estadística), obliga á admitir un margen preciso de oscilación en los valores de los datos y promedios, naciendo así el concepto llamado variación, mas importante y representativo que el de los promedios, á los que tienen que acompañar (porcentajes, coeficientes, etc.)

Si se determinan en una serie de datos, las diferencias de cada termino hasta el promedio, puede formarse con ellas, una nueva serie paralela, con valores positivos y negativos, llamada "de Desviaciones", cuya suma (sin signos) dividida entre el número de términos nos dá la desviacion media, primera medida de variacion. Lo mismo que la Ma, se influye mucho por los datos extremos y por eso se prefiere la desviacion standard.- Así pues: Desviacion media $\delta_m = \frac{\Sigma\delta}{n}$.-

2.° Desviación típica, standard ó cuadrática media.- Formando una nueva serie paralela, al elevar al cuadrado todas las desviaciones sencillas (con lo cual se pierde el signo), sumando estos cuadrados, dividiendo entre el número de terminos y hallando la raiz cuadrada del cociente, se obtiene la desviacion standard, que se expresa con la letra sigma minúscula: $\sigma = \sqrt{\frac{\Sigma\delta^2}{n}}$.-

El cociente entre la δ_m y $\sigma = 1{,}253$, si la distribucion de frecuencias en la serie, es normal ó gausiana, cociente que sirve para conocer esta característica de la serie, sin buscar los intervalos ni las frecuencias.- Un valor un poco mas alto ó mas bajo de esta cifra puede denunciarnos una serie gausiana, hiper ó hipobinomial.- Cuando el número de terminos de la serie es menor de 30, pierde sensibilidad representativa la desviación standard, siendo precisa la correccion de Johanssen $= \sigma = \sqrt{\frac{\Sigma\delta^2}{n-1}}$

Para series con gran número de terminos, se facilita el cálculo, ya con tablas de cuadrados, ó siguiendo esta otra norma: a). Ma de los términos de la serie.- b). Cuadrado de todos los datos (tablas, regla de cálculo).- c). Suma de estos cuadrados y determinación de su media aritmética.- d). Sustraer el cuadrado de la media.- e). Dividir el resto entre el número de terminos.- f). Extraer la raiz cuadrada del cociente.- Resulta igualmente el valor de la desviación standard sin utilizar las desviaciones; $\sigma = \sqrt{\left(\frac{\Sigma n^2}{n} - Ma^2\right) : n}$

Cuando se manejan porcentajes aislados, el teorema de distribucion binomial de Bernoilli, nos da la σ de cada uno, por la formula siguiente "error standard de porcentajes" $\sigma = \sqrt{\frac{p \cdot q}{n}}$, en que "p" es el porcentaje; "q", es la diferencia hasta 100 y n el número de elementos sobre el que se ha calculado el porcentaje (razón inversa del tamaño de la muestra).-

3.° - Indice de variabilidad ó de dispersión.- Mide la dispersión de los datos de la serie; cuanto menor es su valor, tanto mas significativo es el promedio; su valor se da por la fórmula de Pearson: $V = \frac{\sigma}{Ma} 100$ -

Todas estas medidas como los promedios, son constantes estadísticas de las series

Aj. 5

Hagamos reduccion entre el sistema I y III, relación 110 a y 11a = 10 por cuya cifra multiplicamos I

$$\left.\begin{array}{l} I^a\ 19.11 = 11a \quad +110c \\ III^a\ 178.90 = 110a \quad +1958c \end{array}\right| \times 10 \left|\begin{array}{l} 191{,}10 = 110a + 1100c \\ \underline{178.90} = \underline{110a} \ \underline{+1958c} \\ 12.20 = 0 \quad + \ -858c \end{array}\right.$$ y restando ; $c = 12.20 : -858 = \boxed{-0.014 = c}$

Aplicando a la I, los valores obtenidos de $\underline{b}$ y $\underline{c}$, obtenemos $\underline{a}$

$19.11 = 11a + (-1.54)$; $20.65 = 11a$; $a = 20.65 : 11$ — — — $\boxed{a = 1.87}$

Formula general de la parábola y valor de sus sucesivos puntos: $y = a + bx + cx^2$ $\boxed{b = 0.027}$

Valores de y en los respectivos de x

Para	$x = -5$	$y = 1.870 + (-0.135) + (-0.35)$	—	$y = 1.385$
	$x = -4$	$y = 1.870 + (-0.108) + (-0.24)$		$y = 1.542$
	$x = -3$	$y = 1.870 + (-0.081) + (-0.126)$		$y = 1.663$
	$x = -2$	$y = 1.870 + (-0.054) + (-0.056)$		$y = 1.760$
	$x = -1$	$y = 1.870 + (-0.027) + (-0.014)$		$y = 1.829$
	$x = 0$	$y =$		$y = 1.870$
	$x = +1$	$y = 1.870 + 0.027 + (-0.014)$		$y = 1.883$
	$x = +2$	$y = 1.870 + 0.054 + (-0.056)$		$y = 1.868$
	$x = +3$	$y = 1.870 + 0.081 + (-0.126)$		$y = 1.825$
	$x = +4$	$y = 1.870 + 0.108 + (-0.224)$		$y = 1.754$
	$x = +5$	$y = 1.870 + 0.135 + (-0.350)$		$y = 1.655$

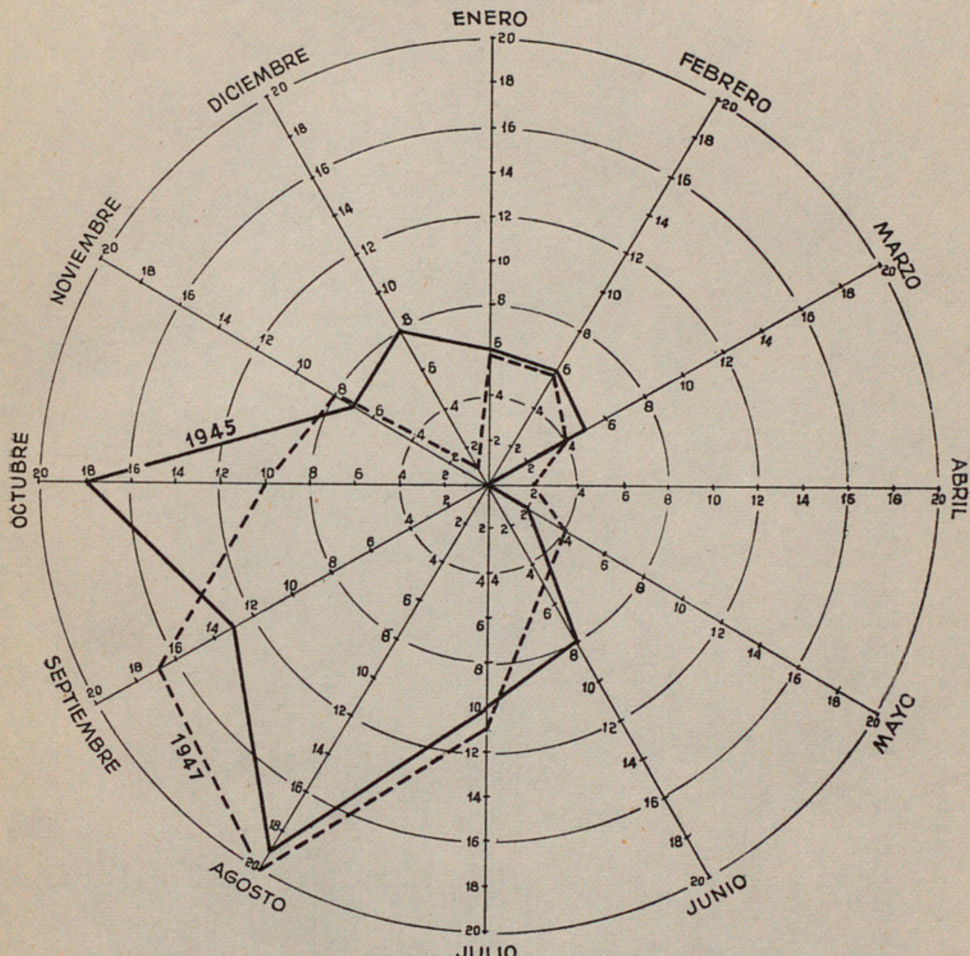
I)

— Definiciones —

Los elementos numéricos con que iniciamos el estudio estadístico de un fenómeno, se llaman - hechos, datos, variables.
Los cocientes de cada grupo análogo de datos, referidos á cada 100 del total ó á cada 1000, se llaman porcentajes ó tantos por 1000 (%, ‰).
Su proporcionalidad al total, les hace seguir reglas de la distribución binomial -
El cociente ó relación entre dos valores, es un índice. El coeficiente es una expresión matemática que concreta una gama de variaciones de un fenómeno estadístico. A veces es factor de una expresión algébrica (como en el binomio de Newton).
Tasa es la relación ó cociente de un grupo de hechos demográficos ó epidemiológicos, á una cuantía de población (á 1000; á 10.000; á 100.000 habitantes), así: tasa de natalidad, de mortalidad, de mortalidad por tuberculosis, por cancer, etc. -
Los hechos, tasas, índices, etc, presentados en grupos distintos, de un mismo fenómeno, forman las series estadísticas. Si en ellas los hechos distintos se suceden con relación al tiempo, se llaman históricas ó dinámicas; si expresan datos recogidos en un solo momento ó plazo, se llaman estáticas ó de frecuencias (natalidad en las provincias en 1953). Las estáticas admiten la ponderación; las dinámicas admiten la tendencia estadística. -

Los datos deben agruparse según divisiones de ordenación lógica; los grupos fundamentales se llaman atributos ó clases; suelen subdividirse dicotómicamente, asignando á cada grupo las variables ó datos correspondientes (en censos: varones, hembras; matrimonios con hijos y sin hijos; estado: solteros, casados, viudos, etc)
Las series estáticas se ordenan en sus datos ya en norma progresiva de sus valores (graduatoria creciente), o regresiva (graduatoria decreciente); sus extremos son límites de la serie, cuya diferencia en valores constituye la medida llamada rango ó range, variación máxima posible entre sus datos. - Los datos de las series estáticas se reunen en grupos menores, á modo de medidas de igual diferencia, creciente ó decreciente, según sea la serie, quedando dividida en intervalos, dentro de cuyas medidas iguales quedan situados los datos; el número de datos de cada intervalo se llaman frecuencias, con cuyas cifras se forma una segunda serie condensada, llamada de frecuencias. Se vé de este modo la zona de la serie donde se concentran los datos, ó distribución de frecuencias (centro, extremos, irregular, etc)
El conjunto de los datos recogidos en el estudio de un fenómeno, se llama protocolo. - El total de datos que pueden obtenerse por muy diversos investigadores sobre el mismo fenómeno, se llama universo estadístico
Los datos similares que pueden formar serie se buscan tambien de las observaciones genéricas recogidas en las fichas de exploraciones sistemáticas; se formalizan de todos los casos en laboratorios, dispensarios, etc, cuyo conjunto forma los ficheros, que deben separarse y numerarse por años ó periodos de tiempo convenientes para sumar mayor número de observaciones de iguales circunstancias. El protocolo se refiere sólo a observaciones de pocas particularidades estudiadas sobre los casos; el fichero reune de modo rutinario una amplia gama de particularidades y del mismo se puede hacer protocolo de alguno ó varios datos. -
La representación de toda una serie, mediante una de sus cifras ó próxima á ella, se llama promedio. Sin los límites de la serie conocidos, un promedio, no tiene representación estimable; no se conoce su error probable y no tiene precisión; el porcentaje en cambio nos da una relación proporcionada. Ambos necesitan como todas las constantes estadísticas un margen de oscilación que les hace adaptables á la realidad; tal es el concepto de la variación que estudiaremos y que juega hoy un papel importante en la aplicación de la estadística al campo de la investigación científica. El promedio y la variación son las constantes estadísticas fundamentales.
Los índices, coeficientes, promedios, etc, se expresan generalmente ya con letras del alfabeto griego ó con otras de convenio universal; pasan á formar elementos de ecuaciones ó expresiones para el cálculo. Se utilizan tambien en estadística los signos matemáticos y curvas geométricas.

CICLO ESTACIONAL DE LA MORTALIDAD POR TIFOIDEA EN BARCELONA • 1945 - 1947



	1945	1947		1945	1947
ENERO	6	6	JULIO	10	11
FEBRERO	6	6	AGOSTO	19	20
MARZO	5	4	SEPTIEMBRE	13	17
ABRIL	0	2	OCTUBRE	18	10
MAYO	2	4	NOVIEMBRE	7	8
JUNIO	8	5	DICIEMBRE	8	1

41) – Curva normal de Gauss-Laplace, de frecuencias del azar, de errores. Significación analítica elemental.–

1º Es el límite á que tiende la distribución binomial (de frecuencias finitas á las infinitas del azar)

Valor de la ordenada máxima ó Ma, en curvas perfectas, modo ó mediana $= \dfrac{1}{\sigma\sqrt{2\pi}}$ (1-2)

Punto cuspidal (2), es el origen normal de las $\bar{x}$ (ó desviaciones). En este punto: $\bar{x} = 0$

– Valor "y" en el punto de inflexión (3 - - 4) $= \bar{x} = \sigma = y = \pm \dfrac{Ma}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}}$

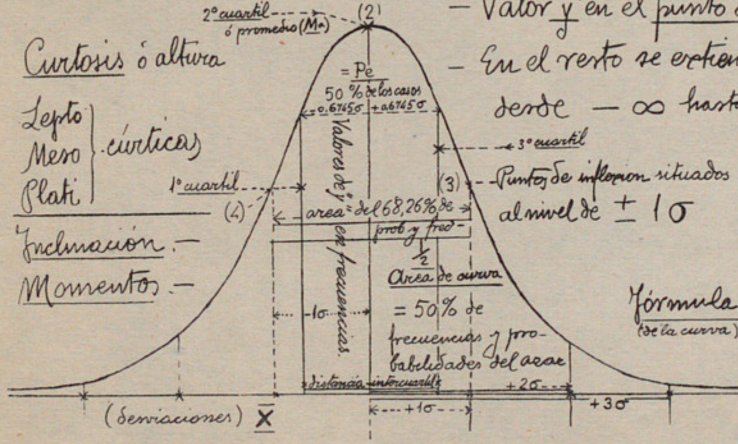
– En el resto se extiende sobre el eje de abscisas desde $-\infty$ hasta $+\infty$; es pues asintótica

Curtosis ó altura

Lepto, Meso, Plati } cúrticas

Inclinación.–

Momentos.–



Siendo toda el area incluida, el total de probabilidades = 1, la fórmula de cada uno de sus puntos (valores de $\pm\bar{x}$), para un fenómeno dado, es

Fórmula 1ª (de la curva) - - - $y = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\bar{x}}{2\sigma^2}}$

(valor de las ordenadas hasta la curva en cada desviación)

($e = 2{,}7183\ldots$ es el límite a que tiende $\left(1+\frac{1}{n}\right)^n$ base de logaritmos Neper)

Si no hay desviación, $\bar{x} = 0$, el segundo miembro desaparece; la curva se abre y se hace asintótica en sentido vertical: todas las probabilidades son de la media aritmética (V. su valor, arriba).

La altura y velocidad de las vertientes se dá por el factor de "precisión" $= h$ " $\dfrac{1}{\sigma\sqrt{2}}$ (a) $= \dfrac{1}{2\sigma^2}$ (b) sustituyendo en la fórmula de la curva (primer miembro)

$$\text{(a)}\quad \frac{1}{\sigma\sqrt{2\pi}}\ ;\ \frac{1}{\sigma\sqrt{2}}\cdot\frac{1}{\sqrt{\pi}}\ ;\ = \sqrt{\frac{h}{\pi}}$$

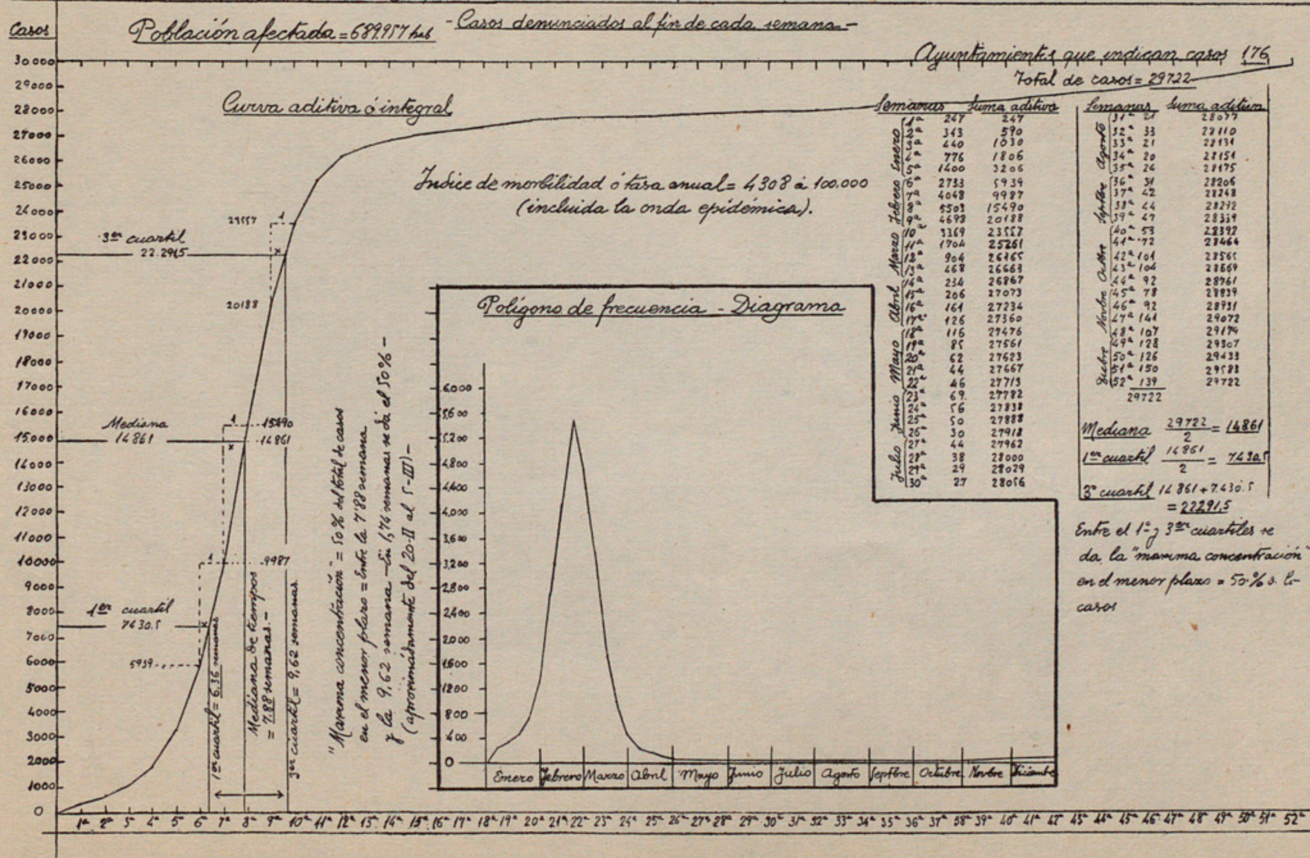
y el exponente: (b) $-\dfrac{\bar{x}^2}{2\sigma^2} = -\left(\dfrac{1}{2\sigma^2}\cdot\dfrac{\bar{x}^2}{1}\right) = -h\bar{x}^2$

Fórmula 2ª de la curva, con el factor h:

$$y = \sqrt{\frac{h}{\pi}}\, e^{-h\bar{x}^2}$$

Al nivel de $\pm 4{,}24\sigma$, el valor de la ordenada (y) es nulo; desde ese punto hasta la ordenada central (Ma), se incluye el 50% del área de probabilidades y á ambos lados = 1, el total.–

Fig. 3 Marcha cronológica de la gripe en Barcelona (sin capital) durante el año 1953.- Constantes estadísticas.-



	Semanas		Suma aditiva
Enero	1ª	247	247
	2ª	343	590
	3ª	440	1030
	4ª	776	1806
	5ª	1400	3206
Febrero	6ª	2733	5939
	7ª	4048	9987
	8ª	5503	15490
	9ª	4698	20188
Marzo	10ª	3369	23557
	11ª	1704	25261
	12ª	904	26165
	13ª	468	26663
Abril	14ª	234	26867
	15ª	206	27073
	16ª	161	27234
	17ª	126	27360
Mayo	18ª	116	27476
	19ª	85	27561
	20ª	62	27623
	21ª	44	27667
	22ª	46	27713
Junio	23ª	69	27782
	24ª	56	27838
	25ª	50	27888
	26ª	30	27918
Julio	27ª	44	27962
	28ª	38	28000
	29ª	29	28029
	30ª	27	28056

	Semanas		Suma aditiva
Agosto	31ª	21	28077
	32ª	33	28110
	33ª	21	28131
	34ª	20	28151
	35ª	24	28175
Septbre	36ª	31	28206
	37ª	42	28248
	38ª	44	28292
	39ª	47	28339
Octubre	40ª	53	28392
	41ª	72	28464
	42ª	101	28565
	43ª	104	28669
	44ª	92	28761
Novbre	45ª	78	28839
	46ª	92	28931
	47ª	141	29072
	48ª	107	29179
Dicbre	49ª	128	29307
	50ª	126	29433
	51ª	150	29583
	52ª	139	29722
		29722	

Mediana $\frac{29722}{2} = 14861$

1er cuartil $\frac{14861}{2} = 7430.5$

3º cuartil $14861 + 7430.5 = 22291.5$

Entre el 1º y 3er cuartiles se da la "maxima concentración" en el menor plazo = 50% de los casos

Hacinamiento e infección tuberculosa – Correlación lineal – Indice de Pearson – ("r.")

Barcelona 1945.

(a) Hacinamiento en distritos: Indices de concentración urbana – b. Infección. Positividades al Mantoux de 0 a 14 años

Distritos	Indices (a)	Desviac.	Desv. cuadr.
I	3.87	+0.86	0.77
II	0.56	−2.43	5.90
III	0.37	−2.62	6.86
IV	4.59	+1.60	2.56
V	9.47	+6.48	41.49
VI	4.31	+1.32	1.76
VII	2.90	−0.09	0.81
VIII	2.18	−0.81	0.65
IX	0.67	−2.32	5.38
X	0.99	−2.00	4.00
	29.91		70.18

Media ar. = 2.99

Desv. σ

$$\sigma = \sqrt{\frac{\Sigma \delta^2}{n}}$$

$$\sigma = \sqrt{7.02}$$

$$\sigma = 2{,}64$$

	Positivid.% (b)	Desviac.	Desvia. cuad.
I	69.60	+ 4.4	19.36
II	63.10	− 2.1	4.44
III	54.70	−10.5	110.25
IV	67.1	+ 1.9	3.61
V	78.0	+12.8	163.84
VI	63.6	− 1.6	2.56
VII	67.0	+ 1.8	3,24
VIII	62.5	− 2.7	7.29
IX	60.7	− 4.5	20.25
X	66.3	+ 1.1	1,21
	652.6		336,05

Media arit. = 65.2

Desvia. σ

$$\sigma' = \sqrt{\frac{\Sigma \delta^2}{n}}$$

$$\sigma' = \sqrt{33{,}60}$$

$$\sigma' = 5.79$$

Indice de correlación de Pearson: (r.)

Fórmula "r" $= (\Sigma\, \delta . \delta') : (n \sigma \sigma')$

+0,86	×	+ 4,4	=	+ 3,87	
− 2,43	×	− 2,1	=	+ 5,10	
− 2,62	×	−10,5	=	+27,51	
+ 1,60	×	+ 1,9	=	+ 3,04	
+ 6,48	×	+12,8	=	+82,94	
+ 1,32	×	− 1,6	=	......	−2,11
− 0,09	×	+ 1,8	=	......	−1,62
− 0,81	×	− 2,7	=	+ 2,18	
− 2,32	×	− 4,5	=	+10,44	
− 2,00	×	+ 1,1	=	......	−2,20
				+135,08	−5,93

$\Sigma\, \delta . \delta' = 129{,}15$

$n \sigma \sigma' = 10 \cdot 2{,}64 \cdot 5{,}79 = 152.8$

$r = (\Sigma \delta\delta') : (n\sigma\sigma') = 129{,}15 : 152{,}8 = \boxed{r = 0{,}84}$

Pe de r $= 0{,}6745 \dfrac{1-r^2}{\sqrt{n}} = P_e = 0{,}0627$

r = muy significativo (hacinamiento infección)

Error standard de porcentajes y promedios.- Estimación de diferencias.-

Hemos dicho que dos errores standard son igual a tres errores probables prácticamente: $2\sigma = 3\,Pe$.- Constituye su cuantía, el margen máximo convenido, de la influencia del azar, en los datos de los fenómenos gausianos.— Por bajo de este límite de variación "<u>no son significativos</u>" los valores hallados, diferencias, etc.

1º <u>Error standard de porcentajes</u>: El teorema de Bernouilli de la distribución binomial de frecuencias nos da valor de $\sigma = \sqrt{n.p.q}$; como esta distribución es el paso á la curva de Gauss, cuando es infinito el número de términos (azar, universos generalizados), cualquier estudio en fenómenos gausianos lleva en si esta idea de amplitud y la muestra que se estudia, debe considerarse como una parte del total, tomada al azar o "por sondeo".- En tal muestra se tienen p. ej: <u>n</u> individuos, de los que hay <u>x</u> con cierto caracter y en cada uno hay probabilidad $x : n = p'$; frecuencia relativa respecto de la probabilidad total.- Esta probabilidad total $= p$, es empíricamente $= p'$; siendo en las muestras ó ensayos, despreciable su cuantía, respecto del total; el factor de corrección <u>n: universo completo</u>, no se tiene en cuenta. Por esto la fórmula binomial $\sigma = \sqrt{n.p.q}$, se convierte en fenomenos gausianos en $\sigma p' = \sqrt{\frac{p.q}{n}}$, que sirve muy bien para hallar el error standard de porcentajes, tasas, etc. que lleven en si idea de proporcionalidad en sus elementos. No puede en cambio aplicarse á promedios, pues les falta aquella limitación de variación (100); el σ standard ó probable, tiene que deducirse por la técnica general de la desviación standard en toda la serie.- En los porcentajes, $\sigma = \sqrt{\frac{p.q}{n}}$, en que "p" es el valor del tanto por ciento; "q", es el resto hasta 100; "n" el número de términos de que ha sido deducido (tamaño de la muestra, protocolo, etc). En las fórmulas del error standard, siempre figura "n" en el denominador; un teorema de muestras nos indica que la dispersión de los datos respecto de la <u>Ma</u>, crece en razón inversa de la raíz cuadrada de la amplitud del número de términos, siendo por tanto la variación (σ), mayor ó menor, en los protocolos segun el número de sus datos, aun tratándose del mismo fenómeno.-

Si tenemos dos porcentajes cuyas diferencias queremos estimar (técnicas distintas en intervenciones, etc), pulsaremos su diferencia mediante el error standard de la diferencia de porcentajes "<u>sigma diferencia</u>" $(\sigma\ dif.) = \sqrt{\sigma. + \sigma'}$, cuyo valor se utiliza para medir si la diferencia que vemos entre los porcentajes no es significativa (puede darla el simple azar), ó es débilmente ó fuertemente significativa (específica, estimable). Ej: En una investigación, con una técnica encontramos un resultado un 20 % de las veces; con otra hallamos un 10 %; siempre hemos utilizado 50 pruebas ó datos, etc (tamaño de la muestra).- σ' de 20 % = <u>5.6</u>; σ' de 10 % = <u>4.2</u>; σ dif. de ambos $= \sqrt{5.6 + 4.2} = 3.1$ —— La diferencia de porcentajes es <u>10</u>.- Como esta diferencia es mayor de 2 veces su σ dif $(10 > 6.2)$, la diferencia en la investigación sale de la esfera del azar $(2\sigma ; 3\,Pe)$ y es por tanto "significativa".- Si la diferencia es $= \sigma$ dif, no es significativa.- Si la diferencia es $= 1.5\,\sigma$ dif, será probáblemente significativa.-

2º <u>Error probable, ó standard de promedios</u>: El error standard de la <u>Ma</u> (V. σ de constantes) es: $\frac{\sigma}{\sqrt{n}}$.- Si la <u>Ma</u> es $> 2\frac{\sigma}{\sqrt{n}}$, se puede utilizar como representativa de su serie y solo en este caso, puede compararse con la Ma' tambien representativa para medir si sus diferencias son ó no significativas por el "<u>σ dif de promedios</u>" de $Ma - Ma' = \sqrt{(\sigma Ma)^2 + (\sigma' Ma')^2}$. Si la dif $Ma - Ma'$ es $=$ ó $> 2\sigma$ dif, la diferencia es "<u>significativa</u>".-

8ª/ Probabilidades en cálculo combinatorio y en funciones –

1º **Permutaciones**. Cambios indistintos del total de elementos. $P_m = m!$

En ciclo cerrado $P_m = (m-1)!$

Ej: Cambios con 6 elementos $P_6 = 6.5.4.3.2.1 = 720$

En ciclo cerrado $= 5.4.3.2.1 = 120$

Probabilidad de una posición determinada $= 1:720 = 0.00138$ (0.138%)

$1:120 = 0.0083$ (0.83%)

2º **Variaciones**: Cambios en grupos del total

$$V_{m,n} = m! : (m-n)!$$

Ej: Seis términos de dos en dos $V_{6,2}$ –

$6.5.4.3.2.1 : 4.3.2.1$

$720 : 24 = 30$

Probabilidad de una variación dada $\frac{1}{30} = 0.033$ (3,3%)

3º **Combinaciones**. Variaciones que no aceptan cambios de posición en los grupos –

$$C_{m,n} = m! : n!\,(m-n)!$$

Ej: Seis términos de 2 en 2 sin invertir los grupos

$C_{6,2} = 6.5.4.3.2.1 : 2.1\,(4.3.2.1)$

$720 : 48 = 15$

Probabilidad de una de ellas prefijada $= 1:15 = 0.066$ (6.66%)

Funciones. Si en un fenómeno influyen factores que en su desarrollo tienen funciones geométricas, la probabilidad lleva como constante, un elemento de aquella función –

Ej: Tiramos al aire, á cierta altura, un número de agujas de gramola que caen sobre un papel rayado á la distancia de su longitud

Probabilidad de que las agujas queden cruzando rayas: La aguja gira en el aire: aparece la constante K, integral del seno, que marca la probabilidad, $K = \frac{2}{\pi} = 0,63$

De 100 agujas, 63 agujas cortarán rayas – 37 no las cortarán

Si son cuadrículas á esa distancia, son probabilidades independientes $= (p+q)^2 = p^2 + 2pq + q^2$; $(63+37)^2$ –: de 100 agujas:

Resultado: 39.69, cortan dos líneas; 46.6 una sola; 13.7 ninguna

Permutaciones con repetición – Volteo de 4 monedas:

Casos posibles = 16.

1º = C C C C

2º:
C C C $\oplus$
C C $\oplus$ C
C $\oplus$ C C
$\oplus$ C C C

3º:
C C $\oplus$ $\oplus$
C $\oplus$ C $\oplus$
$\oplus$ C $\oplus$ C
$\oplus$ $\oplus$ C C
C $\oplus$ $\oplus$ C
$\oplus$ C C $\oplus$

4º:
C $\oplus$ $\oplus$ $\oplus$
$\oplus$ C $\oplus$ $\oplus$
$\oplus$ $\oplus$ C $\oplus$
$\oplus$ $\oplus$ $\oplus$ C

5º $\oplus$ $\oplus$ $\oplus$ $\oplus$

Probabilidad del

1º = 1:16; 6'25%
2º = 4:16; 25%
3º = 6:16; 37'5%
4º = 4:16; 25%
5º = 1:16; 6'25%

100.00

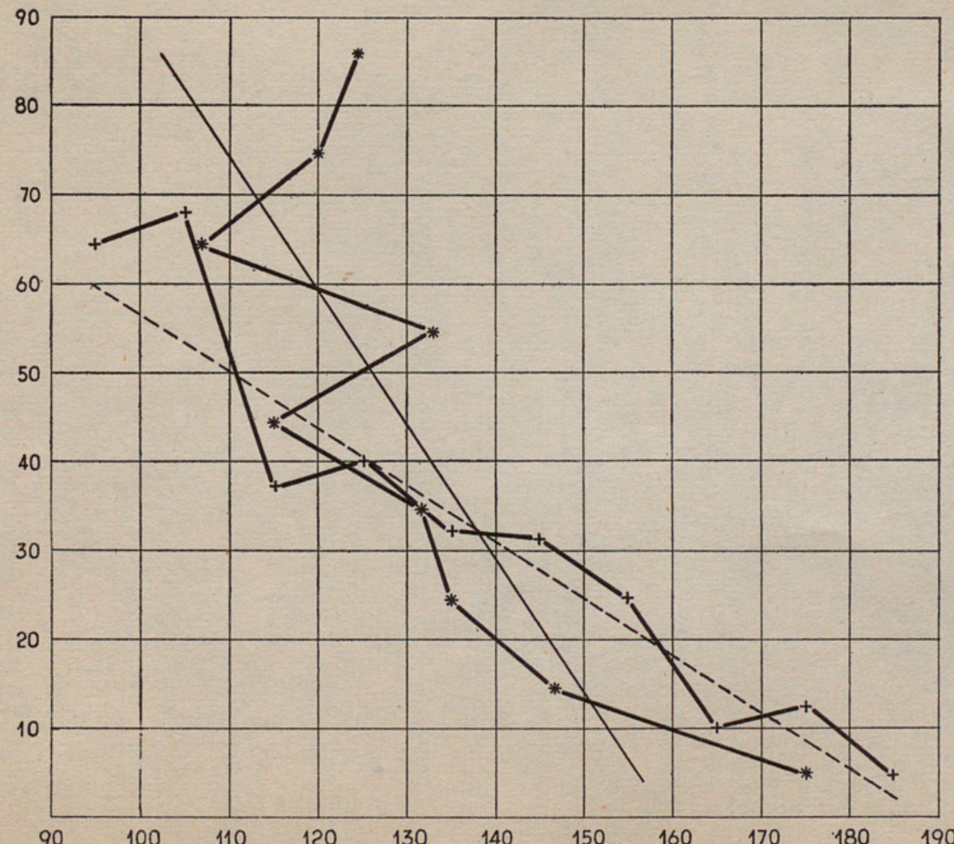
Resumen posible en 16 tiradas:

1 grupo de 4 caras; 4 grupos de 3 caras y 1 cruz; 6 grupos de dos caras y dos cruces; 4 de una cara 3 cruces; 1 grupo de cuatro cruces:

$$1c^4 + 4c^3\oplus + 6c^2\oplus^2 + 4c\oplus^3 + 1\oplus^4$$

que son los coeficientes y exponentes del binomio de Newton

CORRELACION ENTRE MORTALIDAD POR TIFOIDEA Y SUMINISTRO DE AGUA EN BARCELONA · 1910-1948
2º LINEAS DE REGRESION Y TENDENCIA



SIGNOS : ✱ LITROS DE AGUA + TASAS DE MORTALIDAD

9.3./ Valores atípicos en los protocolos. Estimación de diferencias. Precisión. Imprecisión.

1.- Valores atípicos. En una serie destacan á simple vista algunos valores como distanciados de los demas. Se buscará la probabilidad entre las desviaciones, hallando su desviación reducida $= \frac{\delta}{\sigma} = z$ y se acude a la tabla de Sheppard para hallar su probabilidad por ciento o por mil, restando hasta 50 la probabilidad hallada en la tabla y multiplicando por 100 ó por 1000. Si el resto de los valores de la serie tienen probabilidades de 400 ó 600 por 1000 y el analizado solo tiene 2 ó 3, se calificará como atípico, marcándole con un círculo rojo, ya para no contabilizarle en los cálculos ó para buscar la causa determinante (finalidad de la estadística).-

2. Estimación de diferencias entre varias investigaciones (por tablas de probabilidades) se halla la desviación reducida de ambas series con la fórmula siguiente: $\frac{\text{Diferencia de ambas medias aritméticas}}{\text{Suma de los cuadrados de las desviaciones de ambas series, por su número de términos y raíz cuadrada de esta suma = "sigma diferencia" (V. error standard)}} = \frac{Ma - M'a}{\sqrt{\delta^2 n + \delta'^2 n'}}$

El valor hallado se lleva á la tabla de Sheppard para ver la probabilidad, ya a media área de la curva ó para el total (resto a 0,5 ó á 1,00), por cuyo valor vemos el % ó ‰ en que pueden presentarse (menos de 5% ó 50‰ no es significativa), en norma del azar, diferencias iguales o mayores.-

3º Precisión é imprecisión - En las curvas achatadas "platicúrticas", se dispersan los datos, desde la media aritmética, aumentan las desviaciones y se pierde la precisión de la Ma. y de σ. - Se buscan por las tablas, los coeficientes de precisión "K" y de imprecisión $\lambda = a$). La precisión (con relación a σ) figura en la 2ª fórmula de la curva, como valor "h" $= \frac{1}{\sigma\sqrt{2}} = \frac{1}{1{,}414\sigma} = \boxed{\frac{0.707}{\sigma} = K_\sigma}$ — Con relación a la desviación media, se demuestra (cálculo integ.) que la relación entre la σ y la δ_m es igual a la que guardan $\sqrt{\pi} : \sqrt{2} = 1{,}2533$ y hallando en la proporción de ambas el valor de $\delta_m = \frac{\sigma\sqrt{2}}{\sqrt{\pi}}$ ($\sqrt{2} = 1.414$, $\sqrt{\pi} = 1.772$; $\frac{1}{\sqrt{\pi}} = 0.5642$) - Si introducimos $h = (K)$ en la fórmula de $\delta m = \frac{\sigma\sqrt{2}}{\sqrt{\pi}}$; $\delta_m = \frac{1}{K\sqrt{\pi}}$; $K = \frac{1}{\delta_m\sqrt{\pi}}$, $K = \frac{1}{\delta_m} \cdot \frac{1}{\sqrt{\pi}}$ $\boxed{K_\delta = \frac{0.5642}{\delta m}}$ -

b) La imprecisión = $\lambda_\delta = \delta_m\sqrt{\pi} = \delta_\times 1{,}722$ (b)

$\lambda_\sigma = \sigma\sqrt{2} = \sigma_\times 1{,}414$ (a)

Precisión máxima ó imprecisión máxima K ó $\lambda = \infty$.

Precisión mínima ó imprecisión mínima K ó $\lambda = 0$

(los valores intermedios definen la precisión de la serie)

Ma 4)

Mediana de edades

Cálculo de la mediana de edades en la mortalidad por todas tuberculosis. - Barcelona 1945.-

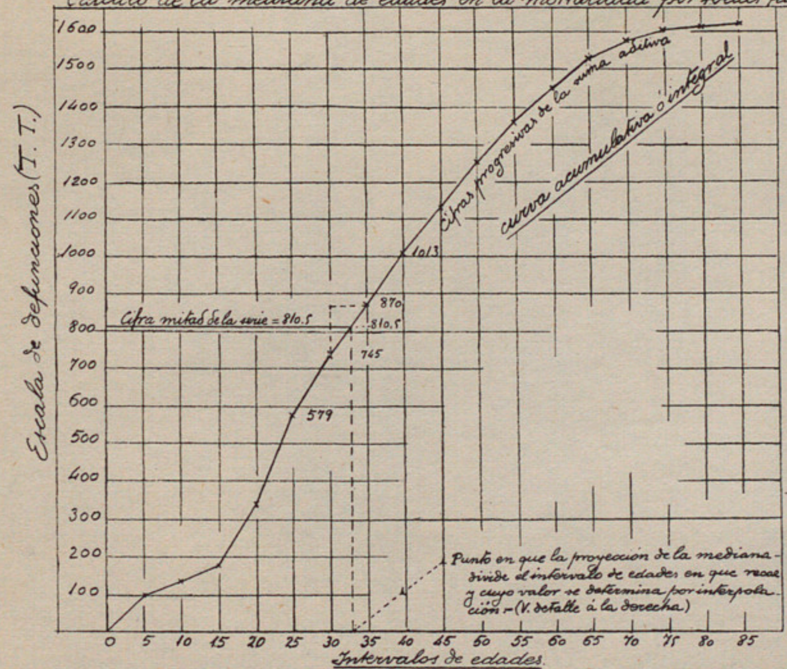


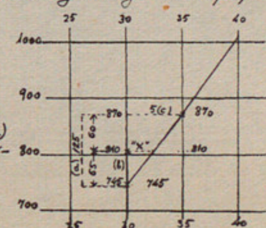
Tabla de distribuciones

Grupos de edades	Defunciones	Suma aditiva
De 0 a 4 años	100	100
5 a 9 "	35	135
10 á 14 "	49	184
15 a 19 "	157	341
20 a 24 "	238	579
25 a 29 "	166	745
30 a 34 "	125	870
35 á 39 "	143	1013
40 á 44 "	120	1133
45 á 49 "	118	1251
50 á 54 "	112	1363
55 á 59 "	91	1454
60 a 64 "	71	1525
65 á 69 "	59	1584
70 a 74 "	18	1602
75 a 79 "	10	1612
80 á 84 "	9	1621 : 2 = 810,5
85 y mas "	0	
No consta edad	15	

Valor de posición que corresponde á la Mediana = 810,5 =

Detalle geométrico de la interpolacion en el intervalo de 30 a 35 años.-

Dos triángulos rectángulos que tienen un ángulo comun, son semejantes y sus lados proporcionales:



El valor medio de la serie aditiva (810), incide en la curva entre las edades del intervalo 30 á 35 años en el que hay que hallar el punto exacto en que recae la mediana = De los dos triángulos semejantes que se ven en el "detalle", se conocen: (c), valor 5 entre las ordenadas 30 y 35; el lado del triángulo mayor de valor 125 (a) entre las abscisas de 870 - 745; el lado (b) del triángulo menor, entre abscisas 745 y 810 = 65.- Es incógnito el valor "X", al nivel de la abscisa 810, pero siendo semejantes se tiene la proporción siguiente:

a : b :: c : x ; 125 : 65 :: 5 : x ; x = (5 × 65) : 125 ; x = 325 : 125 = 2.6, valor que añadido al de 30 años menor del intervalo, nos dá la Mediana de edades buscada = 32,60 años

Aj. 4)

Ajustamiento de una parábola - Tendencia.-

Mortalidad por tuberculosis pulmonar: Edades de 0 y 4 años de edad - Barcelona 1940-1950.-

1º Tasas - Valores de X — Y - Ecuaciones generales de la parábola de 2º grado

Años	Tasas Y	Valores de X	Val. de X^2	Prod. XY	Prod x^2y	Valor X^3	Val. X^4
1940	1.60	-5	25	-8.40	42.00	-125	625
1941	1.28	-4	16	-5.12	20.48	-64	256
1942	1.46	-3	9	-4.38	13.14	-27	81
1943	1.98	-2	4	-3.96	7.92	-8	16
1944	1.63	-1	1	-1.63	1.63	-1	1
1945	1.57	0	0	0,00	0.00	0	0
1946	2.60	+1	1	+2,60	2.60	+1	1
1947	1.72	+2	4	+3.44	6.88	+8	16
1948	1.82	+3	9	+5.46	16.38	+27	81
1949	1.82	+4	16	+7.28	29.12	+64	256
1950	1.55	+5	25	+7.75	38.75	+125	625
n=11	19.11	0	110	+3.04	178.90	0	1958
	Σy	Σx	Σx^2	Σxy	Σx^2y	Σx^3	Σx^4

2º Ecuaciones y valores:

$$I = \Sigma y = na + \underbrace{\Sigma x b}_{0} + \Sigma x^2 c \;;\; 19.11 = 11a + 110c$$

$$II;\ \Sigma xy = \underbrace{\Sigma x a}_{0} + \Sigma x^2 b + \underbrace{\Sigma x^3 c}_{0} \;;\; 3.04 = 110b \;\ldots\; \boxed{b = 0{,}027}$$

$$III;\ \Sigma x^2 y = \Sigma x^2 a + \Sigma x^3 b + \Sigma x^4 c \;;\; 178{,}9 = 110a + 1958c$$

RG 1

- Coordenadas. - Curvas y ecuaciones. -

Cuadro de coordenadas cartesianas u ortogonales (Descartes).

+(y)

Cuadrante de los valores negativos de "x", positivos de "y".

Cuadrante de valores positivos de "x", de "y", el más conocido y frecuente en Sanidad.

-(x)

+(x)

0 (centro de coordenadas Valor = 0).

eje de abscisas ó de los valores "x" (tiempos; edades. intervalos.)

Cuadrante de los valores negativos de "x" é "y".

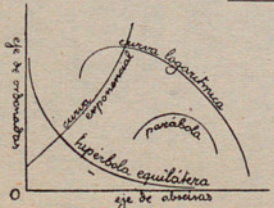
Cuadrante de valores positivos de "x", negativos de "y".

Eje de ordenadas, de valores "y", escala sucesiva numerica de hechos, tasas, etc. de las series.

-"y"

Los valores de "x" é "y", se deducen de las series que se estudian; son las variables (x = variable independiente). Proyectando sobre el campo del cuadrante las ordenadas y abscisas desde la medida correspondiente de cada eje, se obtienen los puntos de encuentro, que unidos por rectas sucesivamente, dibujan el "polígono de frecuencia" ó diagrama mas elemental. Mediante cálculo se simplifica su forma irregular y se trazan rectas ó curvas ajustadas, de tipo geométrico, cuyo estudio analítico, da leyes ó características del fenómeno representado. Las ecuaciones de estas curvas constan de valores de las variables indicadas y de elementos reguladores, de valor fijo ("constantes"), llamados parámetros, que determinan la forma de la curva; intervienen tambien los coeficientes, exponentes, funciones trigonométricas (sen. cos. tg. ctg. etc):

Ecuaciones de las curvas principales



Recta; $y = a + bx$ (función lineal)

Parábola; $y = a + bx + cx^2$ (2º grado, canónica o equilátera).

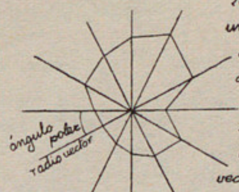
Id. cubica; $y = a + bx + cx^2 + dx^3$

Curva exponencial; $y = ab^x$; (la variable independiente, pasa á exponente (tiempo).

Curva hiperbolica; $y = x^{-1}$; $y = \frac{1}{x}$.

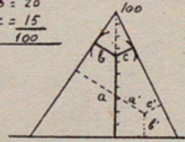
Curva logarítmica; $y = a + bx + c\log x$

Coordenadas polares



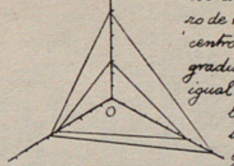
Para representar un fenómeno cíclico con tantas ordenadas como períodos se establezcan en su estudio. Las ordenadas ó radios vectores, se graduan en partes iguales desde el centro, marcando en cada radio, el valor correspondiente de las variables (datos ó tasas de una enfermedad en los doce meses del año) uniendo todos estos puntos en un polígono.-

Coordenadas triangulares



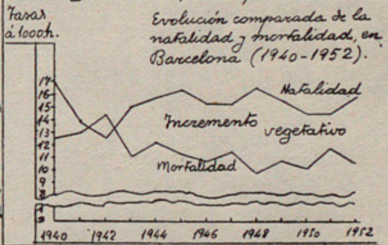
Para representar tres variables cuyos valores oscilan dentro de un margen fijo total que es su suma (composición de productos en tres elementos para 100 total). Dividimos una de dos perpendiculares en 100 partes; se toma la variable mayor en una de ellas y en su final se forma el centro de coordenadas; desde este punto, con radios iguales a cada una de las otras dos variables, se trazan arcos y desde el punto 100, las tangentes a cada arco, forman con la otra perpendicular no graduada, un triángulo. Cuando se desean triángulos iguales para comparar fenómenos distintos, el centro de coordenadas puede caer fuera de la altura; en cualquier punto del área de un triángulo, la suma de las perpendiculares á los lados, es igual á la altura del mismo

Diagrama de conjugadas



Coordenadas en numero de tres, radiando del centro en ángulos de 120 graduados en partes a igual escala. Para solo tres variables, una en cada eje; los extremos se unen por rectas que dibujan un triángulo. Comparando dos o más fenómenos ó protocolos, ó los mismos en momentos distintos, puede objetivarse su variación.-

Polígono múltiple de frecuencias:-

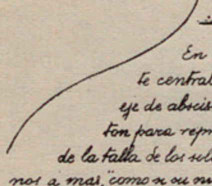


Papel aritmético y logarítmico

Los diagramas en que se representen valores, de uno ó varios fenomenos, proximos, pueden trazarse en papel milimétrico (aritmético). Si la variacion en el valor sucesivo de los datos es grande, ó bien cuando se comparan dos fenómenos en un solo diagrama, con valores de cuantía muy distinta, ha de utilizarse el papel logarítmico simple (en ordenadas), pues solo en él, son proporcionales las variaciones en todas sus zonas. El doble logarítmico es muy util en el cálculo de exponentes, convertido en el de lineas rectas

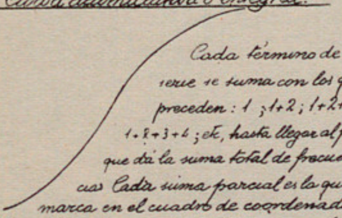
RG 4/ Curvas estadísticas - Areas comparadas - Mapas estadísticos - Estereogramas.- Curvas de nivel - Nomogramas.-

Curva de Galton.-



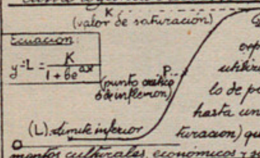
En series grandes, la parte central se hace paralela al eje de abcisas. La introdujo Galton para representar las frecuencias de la talla de los soldados, ordenados de menos á más, "como si su nivel se viera en un espejo. Poco utilizada.-

Curva acumulativa ó integral.-



Cada término de la serie se suma con los que preceden: 1; 1+2; 1+2+3; 1+2+3+4; etc, hasta llegar al final que dá la suma total de frecuencias. Cada suma parcial es la que se marca en el cuadro de coordenadas. Muy utilizada en medidas de frecuencias (mediana, cuartiles, decentiles, etc. Invirtiendo la segunda mitad, se obtiene la curva de Gauss.

Curva logística ó del crecimiento (exponencial)



Ecuación:

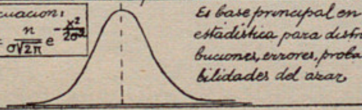
$$y - L = \frac{K}{1 + be^{ax}}$$

Para fenómenos que expresan crecimiento; utilizada para el calculo de población futura, hasta un tope máximo (saturación) que marcan los elementos culturales, económicos y sociales.

Curva normal de frecuencias, de errores. (Gauss)

Ecuación:

$$Y = \frac{n}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$



Es base principal en estadística para distribuciones, errores, probabilidades del azar.

Representación en áreas proporcionales á los valores

a). Figuras regulares. (Fórmula del área de cada una).

Norma general. Se ordenan de mayor á menor los valores á representar gráficamente y se forman las siguientes proporciones:

I $\frac{\text{Valor del mayor (1°)}}{\text{Valor del menor (2°)}} = \frac{\text{área para el mayor (arbitraria) 1°.}}{\text{área para el menor (incógnita) 2°.}}$

II $\frac{\text{Valor del 1°}}{\text{Valor del 3°}} = \frac{\text{área del 1°}}{\text{área del 3° (incógnita)}}$

y así sucesivamente si hay más valores.-

1° Círculos Con la fórmula del área, el segundo miembro de la igualdad es $\frac{\cancel{\pi} r^2}{\cancel{\pi} r'^2} = \frac{r^2}{r'^2}$ y como tenemos conocido r que tomamos arbitrariamente, deducimos el radio para el segundo círculo:

$$\frac{\text{valor del mayor (1°)}}{\text{valor del menor (2°)}} = \frac{r^2}{r'^2}; \quad r' = \sqrt{\frac{r^2 \cdot \text{valor del 2°}}{\text{valor del 1°}}}$$

Del mismo modo se procede para obtener el radio de los demas círculos, que se sitúan y trazan como el arte del calculador aconseje, separados, concentricos, tangentes, siendo forma muy expresiva, representar por un círculo grande, la suma total de valores, incluyendo en su interior círculos representativos de los valores parciales.-

2°. Cuadrados. ($a = l^2$).- Del mismo se calculan los lados de los sucesivos cuadrados, partiendo de la medida arbitraria del mayor:

$$l' = \sqrt{\frac{l^2 \cdot \text{valor del 2°}}{\text{valor del 1°}}}$$

3°.- Triángulos. ($a = \frac{b \cdot a}{2}$). Hallada la relación entre los valores de los datos se dibuja un triángulo arbitrario que represente al mayor, ó á la suma de todos y se deducen las areas que en proporción a los datos, deben tener los menores, tanteando, por la base y altura, sus medidas.-

4° Sectores.- Trazado un círculo mayor suma de los datos y en proporción a estos, la parte alícuota de 360° de cada uno, se dibuja otro círculo proporcional (1°) para la otra serie que se compara, refiriendo también á grados, los respectivos sectores.-

b) Figuras irregulares. Dibujada en cuadrícula milimétrica la figura, se traza otra cuadrícula, que guarde la proporción que se desea, donde se traza el problema.-

Mapas estadísticos.

Las frecuencias obtenidas de cada región, provincia, etc, se ordenan progresivamente, se clasifican en intervalos de grupos y se establece una clave de signos que se dibuja al margen del mapa, con sus respectivos valores, que sirven para marcarlos en las provincias del mapa (colores diversos, punteado, rayado, cuadrícula, etc.

Curvas de nivel. Lo mismo que en topografía, pueden aplicarse á la estadística, para representar en diagrama, el juego de tres valores correlacionados, uniendo sus puntos mediante curvas que siguen á través del mismo las oscilaciones progresivas del tercer valor, si puede así sistematizarse.

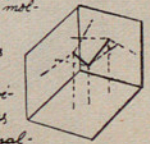
Nomogramas. Variación del anterior en que se coordinan por líneas regulares, en norma de coordenadas, tres clases de medidas, que varian todas proporcionalmente, dependiendo unas de otras.

Variación de la superficie del cuerpo en función de la talla y peso (Du Bois para el metabolismo básico); probabilidades del azar, en función de los valores del χ^2 y del grado de libertad en una asociación, etc. Permiten deducir el tercer valor con los otros dos conocidos

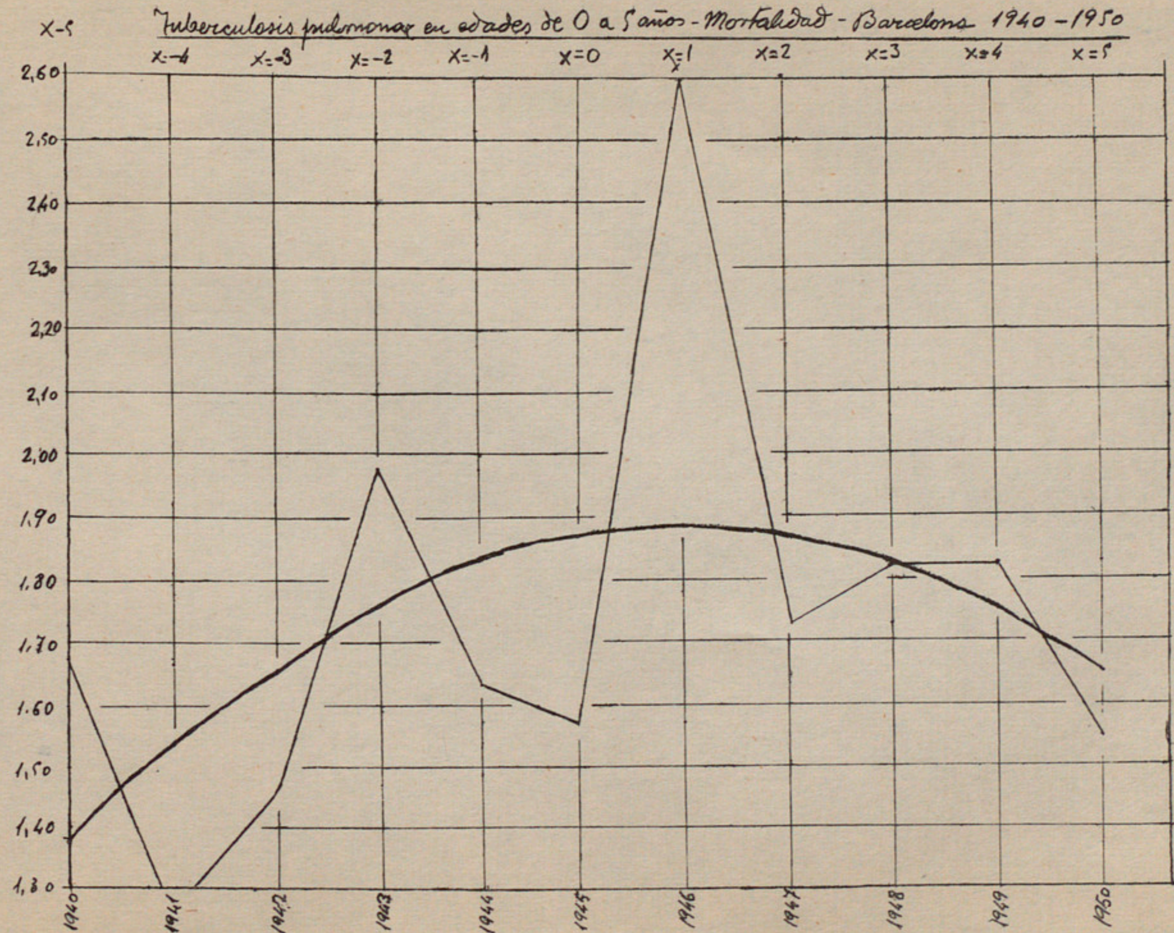
Representaciones en el espacio. Ya mediante el dibujo de proyección ó construyendo modelos

a). Coordenadas en el espacio. Dos valores se sitúan en un plano y el tercero en su proyección correspondiente á escala (superficies de correlación) con sus ondulaciones.

b). Estereogramas. Tanto moldeando, como dibujando bien se obtienen relieves estadísticos, en que sobre la base de planos cuadriculados de coordenadas se levantan proporcionalmente los grupos de frecuencias como acúmulos piramidales, truncados, etc., con la mayor impresión objetiva.-



Aj. 6)



Tuberculosis pulmonar en edades de 0 a 5 años - Mortalidad - Barcelona 1940 - 1950
X-5
x=-4
x=-3
x=-2
x=-1
x=0
x=1
x=2
x=3
x=4
x=5
2,60
2,50
2,40
2,30
2,20
2,10
2,00
1,90
1,80
1,70
1,60
1,50
1,40
1,30
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950

Gr. 2/ <u>Hacinamiento y morbilidad por tuberculosis pulmonar - Correlación - Indice de Pearson ("r")</u>

<u>Barcelona 1945-1948</u>

(a) Hacinamiento en barrios del distrito quinto (Viviendas por 100 m²) (b) Morbilidad (T.P.) sobre su población a 1000 hab. (1945).

Movimiento de enfermos del Dispensario Central (1948).-

Barrios	(a) Indice de viviendas.		Desviaciones	Id. cuadrado	(b). Indices de morb.		Desviac.	Desv. al cuadr.
Teatro	2,5 p. 100 m²	Media aritm. = 2.81	− 0,31	0.09	2.74	Media aritm. = 2.51	+ 0.21	0.04
Liceo	1.7 "		− 1,11	1.22	2.07		− 0.44	0.19
San Pablo	1.6 "		− 1,21	1.46	2.74		+ 0.23	0.05
San Agustín	2,5 "		− 0.31	0.09	2.55		+ 0.04	0.001
Padró	4.0 "		+ 1,19	1.41	1.79		− 0,72	0,52
EE. Pías	3.8 "		+ 0,99	0.98	2.82		+ 0.31	0.09
San Lázaro	3.7 "		+ 0,89	0.79	3.52		+ 1.01	1.02
Carcel	2.7 "		− 0.11	0.01	1.86		− 0,65	0,42
	22,5		δ	6.05	20.09		δ'	2.331 $\Sigma\delta^2$

$$\sigma = \sqrt{\frac{6.05}{8}} = 0{,}869$$

$$\sigma' = \sqrt{\frac{\Sigma\delta^2}{n}}$$

$$\sigma' = \sqrt{\frac{2{,}331}{8}}$$

$$\sigma' = 0{,}539$$

<u>Indice de correlación de Pearson</u> - "r" $= (\Sigma\delta.\delta') : (n.\sigma.\sigma')$

δ		δ'		+	−
− 0.31	×	+ 0,21	=	- - - -	− 0,065
− 1,11	×	− 0,44	=	+ 0,49	
− 1,21	×	+ 0.23	=	- - - -	− 0.28
− 0,31	×	+ 0,04	=	- - - -	− 0,01
+ 1,19	×	− 0,72	=	- - - -	− 0.85
+ 0,99	×	+ 0,31	=	+ 0,31	
+ 0,89	×	+ 1,01	=	+ 0,90	
− 0.11	×	− 0,65	=	+ 0,71	
				+ 2,41	− 1,205

$$\Sigma\delta\delta' = 1{,}205$$

$n\sigma\sigma' = 8 \times 0{,}539 \times 0{,}869 = 3.747$

$r = (\Sigma\delta\delta') : (n\sigma\sigma') = 1{,}205 : 3{,}747 = \boxed{r = 0.348}$

<u>Pe de r</u> $= 0{,}6745 \; \dfrac{1-r^2}{\sqrt{n}} = 0{,}21$

(r = no es significativo)

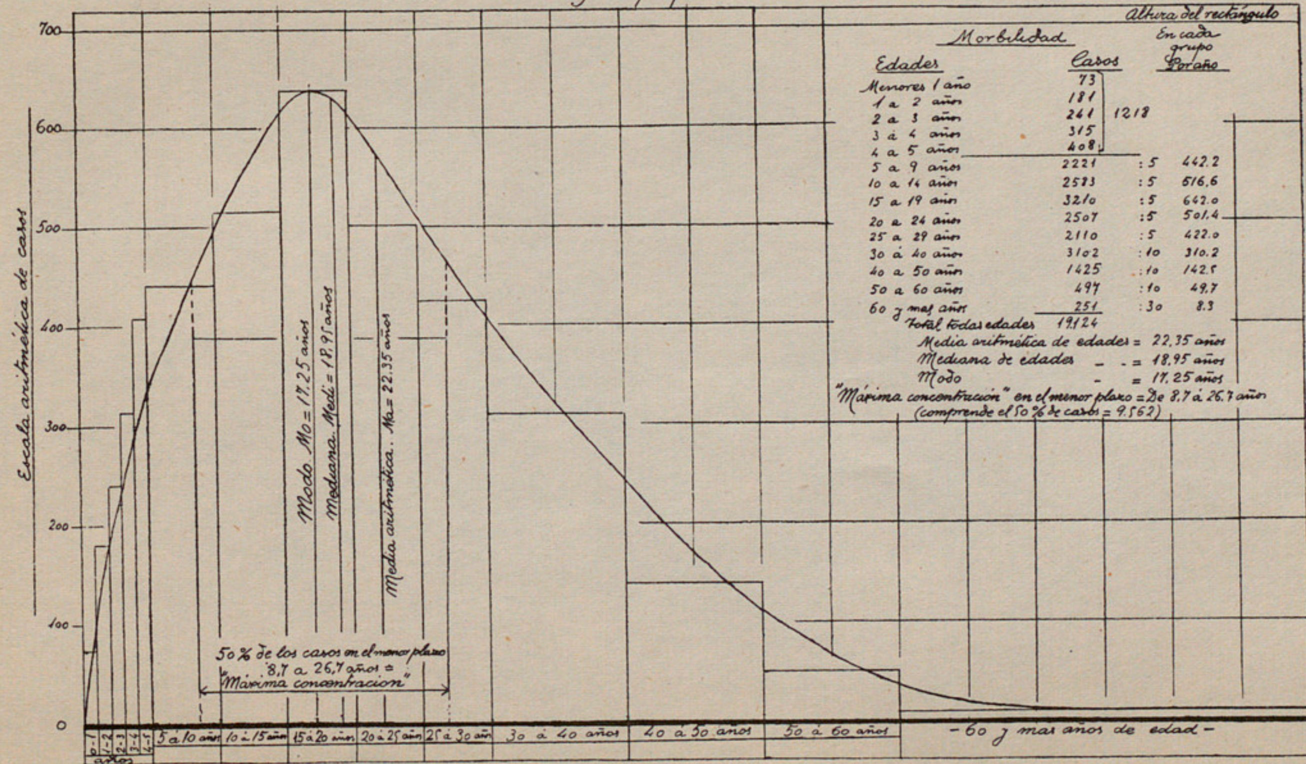
(hacinamiento morbilidad T.P.)

R. Gr. 2

Distribución por edades de 19.124 casos de gripe. Promedios y concentración (Maryland. Dep. de Higiene. EE. UU.)

– Morbilidad específica – Histograma de Pearson –

Areas de los rectángulos proporcionales al número de casos



Edades	Morbilidad Casos			Altura del rectángulo En cada grupo Por año
Menores 1 año	73	1218		
1 a 2 años	181			
2 a 3 años	241			
3 á 4 años	315			
4 a 5 años	408			
5 a 9 años	2221		: 5	442.2
10 a 14 años	2583		: 5	516.6
15 a 19 años	3210		: 5	642.0
20 a 24 años	2507		: 5	501.4
25 a 29 años	2110		: 5	422.0
30 á 40 años	3102		: 10	310.2
40 a 50 años	1425		: 10	142.5
50 a 60 años	497		: 10	49.7
60 y mas años	251		: 30	8.3
Total todas edades	19124			

Media aritmética de edades = 22,35 años
Mediana de edades – – = 18.95 años
Modo – = 17.25 años
"Máxima concentración" en el menor plazo = De 8.7 á 26.7 años
(comprende el 50 % de casos = 9.562)

P-1) Probabilidades de promedios y desviaciones – Series estáticas – Mortalidad por cardiopatías 1950 – España

Media aritmét. = 181,27

Intervalos de tasas en las provincias –	Promedio del intervalo	Frecuencias	Probabilidad del intervalo	Probabil. × prom.	Desviac. a la Ma	D. por probabilidad	Cuadrado desviaciones	D. por la probabilidad
+100 a 120 –	110	3	3:50 = 0.06	6.60	71.27	4.27	5079.41	304.76
+120 a 140 –	130	3	3:50 = 0.06	7.80	51.27	3.07	2628.61	157.71
+140 à 160 –	150	9	9:50 = 0.18	27.00	31.27	5.63	977.81	176.00
+160 a 180 –	170	10	10:50 = 0.20	34.00	11.27	2.25	127.01	25.40
+180 a 200 –	190	11	11:50 = 0.22	41.80	8.73	1.92	16.21	3.56
+200 a 220 –	210	5	5:50 = 0.10	21.00	28.73	2.87	825.41	82.54
+220 a 240 –	230	4	4:50 = 0.08	18.40	48.73	3.90	2374.61	189.97
+240 à 260 –	250	3	3:50 = 0.06	15.00	68.73	4.12	4723.81	283.43
+260 a 280 –	270	2	2:50 = 0.04	10.80	88.73	3.55	7873.01	314.92
			$\Sigma p = 1{,}00$	183,40 PMa		31.58 PDm		1538,29

$$P\sigma = \sqrt{1538{,}29} = 39.2$$

Probabilidad $P = \dfrac{p \text{ (casos favorables)}}{q \text{ (casos posibles)}}$; siendo $q = \infty$; $P = 0$; solo los intervalos pueden tener probabilidad que es lo que ocurre en el azar (curva de Gauss) (áreas)

Si $p = q$, $P = 1$ (certeza). – $p = 2$, $q = 4$, $P = 0.5$ (50%) = incertidumbre. –

La suma de las probabilidades parciales es la unidad – (100%). – Las probabilidades parciales que han de agruparse respecto del total, pueden sumarse de sus elementos – (Total 1.)

Las probabilidades de dos hechos independientes, sucesivos, de igual naturaleza, se multiplican para deducir el total. –

Ma 3

— Media aritmética de edades —

Es un promedio sobre datos agrupados, que determina el lugar exacto en que incide sobre los intervalos de edades, la media aritmética de una serie de hechos clasificados por edades, ordenadas en graduatoria.- Es muy utilizado en estadística sanitaria.-

Se recogen los datos de las defunciones en los servicios de Estadística, donde figuran clasificados por grupos de edades.- Se busca arbitrariamente una cifra de edad que corresponda a los grupos donde se concentran los óbitos valor que se toma a priori como Ma de edades, provisional-

Mortalidad por todas tuberculosis en Barcelona año 1945.

Grupos de edades	Promedio del grupo	Obitos ó frecuencias (f)		Desviaciones ("d") del promedio á la Ma. arbitraria	Productos "df"		Prueba: Restando los obitos de cada grupo (df-f)	Sumando f
0 á 4 años	2.5 años	100 defunciones	Media arbitraria de edades = 32 años	– 29.5	– 2950		– 2850	
5 a 9 "	7,5 "	35 "		– 24,5	– 857,5		– 822.5	
10 á 14 "	12.5 "	49 "		– 19,5	– 955.5		– 906.5	
15 a 19 "	17,5 "	157 "		– 14.5	– 2276.5		– 2119.5	
20 á 24 "	22.5 "	238 "		– 9.5	– 2261.0	Negativos	– 2023	
25 á 29 "	27.5 "	166 "		– 4.5	– 749.0	– 10049.5	– 583	
							– 9304.5	
30 á 34 "	32.5 "	125 "		+ 0.5	+ 62.5			+ 187.5
35 á 39 "	37.5 "	143 "		+ 5,5	+ 786.5			+ 929.5
40 á 44 "	42.5 "	120 "		+ 10.5	+ 1260			+ 1380.0
45 a 49 "	47.5 "	118 "		+ 15.5	+ 1829			+ 1947.0
50 á 54 "	52.5 "	112 "		+ 20.5	+ 2296			+ 2408.0
55 a 59 "	57.5 "	91 "		+ 25,5	+ 2320			+ 2411.0
60 á 64 "	62.5 "	71 "		+ 30.5	+ 2165			+ 2236.0
65 a 69 "	67.5 "	59 "		+ 35.5	+ 2094			+ 2153.0
70 a 74 "	72,5 "	18 "		+ 40.5	+ 729			+ 747.0
75 á 79 "	77,5 "	10 "		+ 45.5	+ 455	Positivos		+ 465.0
80 á 90 "	85.0 "	9 "		+ 53.0	+ 477	+ 14474		+ 486.0
		1621 "			Suma algébrica	+ 4424.5		+ 15350.0

No consta (se prescinde) - - - 15

Suma algébrica + 15350.0 − 9304.5 = 6045.5

Diferencia entre ambas sumas igual á número de frecuencias: 6045.5 − 4424.5 = 1621,0

En los productos "df" hay un exceso positivo de 4424,5 que nos muestra que el promedio arbitrario elegido es mas bajo de la Ma real - dividiendo este exceso de 4424,5 entre el total de óbitos = 1621 (f), obtenemos el factor de corrección que se debe añadir á la media arbitraria para obtener la Media de edades de dicha mortalidad = 32 + 2,72 = 34,72 años

Corr. 3/

Hacinamiento y mortalidad por tuberculosis pulmonar - Correlación - Indice de Pearson (r.)

Barcelona 1945

(a) Hacinamiento en barrios del distrito quinto (Viviendas por 100 m²). - (b). Mortalidad (T.P.) sobre su población á 100.000 hab. (1945) Datos del Registro civil, refiriendo los de nosocomios á sus domicilios (1945.).-

Barrios	Viviendas: 100m² (a)	Desviaciones	Id. al cuadrado	(b) Tasas de mort.	Desviaciones	Id. al cuadrado
Teatro	2.5	-0.31	0.09	128.5	+ 4.41	19.44
Luco	1.7	-1.11	1.22	75.3	-48.79	2380.46
San Pablo	1.6	-1.21	1.46	198.0	+73.91	5462.69
San Agustín	2.5	-0.31	0.09	71.0	-53.09	2818.55
Padró	4.0	+1.19	1.41	80.1	-43.99	1935.12
E.E. Pías	3.8	+0.99	0.98	135.3	+ 9.21	84,82
San Lazaro	3.7	+ 0.89	0.79	176.0	+51,91	2694,65
Carcel	2.7	- 0.11	0.01	130.5	+ 6,41	41.08
	22.5	δ	6.05	992.7	δ'	15436,81

Media aritm. = 2.81

Media aritm. = 124,09

$$\sigma = \sqrt{\frac{6.05}{8}} = 0,869$$

$$\sigma' = \sqrt{\frac{\Sigma\delta^2}{n}} = \sqrt{1929.50}$$

$$\sigma' = 43.9$$

Indice de correlación de Pearson

$$r = (\Sigma\delta.\delta') : (n.\sigma\sigma')$$

δ		δ'		+	-
- 0.31	×	+ 4.41	=	-----	- 1.37
- 1.11	×	-48.79	=	+ 54.16	
- 1.21	×	+ 73.91	=	------	-89.43
- 0.31	×	- 53.09	=	+ 16.46	
+ 1.19	×	- 43.99	=	-----	- 52.35
+ 0.99	×	+ 9.21	=	+ 9.12	
+ 0.89	×	+ 51.91	=	+ 46.20	
- 0.11	×	+ 6.41	=	-----	- 0.70
				+125.94	-143.85

$$\Sigma\delta\delta' = -17,91$$

$$n\sigma\sigma' = 8 \times 0,869 \times 43,9 = 302,03$$

$$r = (\Sigma\delta\delta') : (n\sigma\sigma') = -17,91 : 302,03 = \boxed{-0,059}$$

(r = ausencia de correlación)

(hacinamiento-mortalidad)

Aj. 1)

Ajustamiento de una recta. - Tendencia -

Mortalidad por tuberculosis pulmonar - Barcelona 1925-1936 -

1° Determinacion de las tasas y valores x é y. - Ecuaciones generales de la recta

$n = 12$

Año	Tasas	Valores x	Valores de x^2	Productos de xy
1925	156.25	1	1	156.25
1926	151.05	2	4	302.10
1927	148.09	3	9	444.27
1928	141.50	4	16	566.00
1929	145.18	5	25	725.90
1930	117.54	6	36	705.24
1931	123.33	7	49	863.31
1932	107.73	8	64	861.84
1933	112.73	9	81	1014.57
1934	100.59	10	100	1005.90
1935	86.92	11	121	956.12
1936	90.90	12	144	1090.80
	1481.82	78	650	8692.31
	Σy	Σx	Σx^2	Σxy

Ecuaciones

1° De la recta (Geom. Anali.)

$$y = a + bx$$

2° De la recta que pasa por dos puntos dados (Calculo diferencial) -

$$\Sigma y = na + b\Sigma x \quad (\mathrm{I})$$

$$\Sigma xy = \Sigma x a + b\Sigma x^2 \quad (\mathrm{II})$$

2° Aplicacion de las ecuaciones I y II -

$$\mathrm{I}. - 1481.82 = 12a + 78b$$

$$\mathrm{II} - 8692.31 = 78a + 650b$$

Sistema de dos ecuaciones con dos incognitas

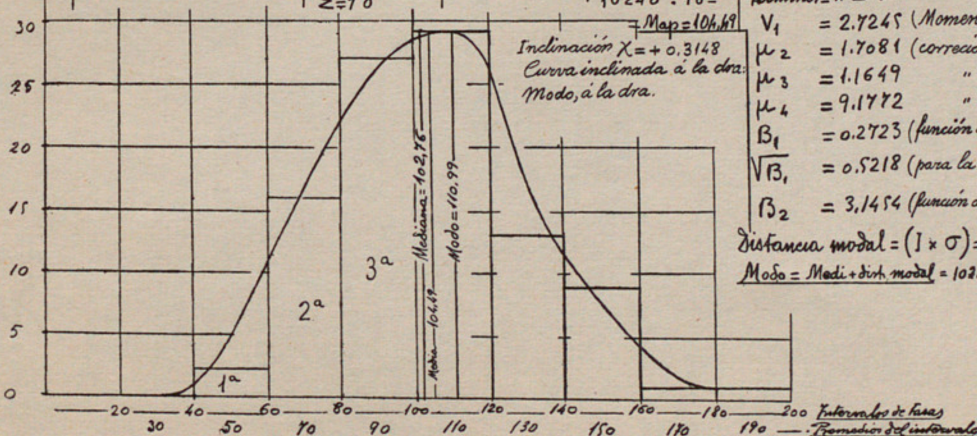
M.I.V. *Estudio técnico-estadístico de la mortalidad infantil de una nación.–*

Tasas de <1 año a 1000 nacidos vivos en 98 ciudades de un país, de más de 25.000 hab. cada una.

Frecuencias
Número de ciudades que dan las tasas

Escala aritmética del eje de ordenadas.

Intervalos de tasas	Nº de ciudades	Promedio intervalo.	Ponderación.
40 á 59 ‰	2	por - - - 50	100
60 á 79	16	" - - - 70	1120
80 á 99	27	" - - - 90	2430
100 á 119	29	" - - 110	3190
120 á 139	13	" - - 130	1690
140 á 159	9	" - - 150	1350
160 á 179	1	" - - 170	170
180 á 199	1	" - - 190	190
	Σ = 98		10240 : 98 =
			Map = 104.49



Constantes de la serie.

Media = 104.49
Modo = 110.99
Mediana = 102.76
$\sigma = 26.14 \pm P_e\, 1.26$
Coefic. de variabilidad = $25.02 \left(\frac{\sigma \cdot 100}{Map}\right)$
Inclinación = + 0,3148 = "I" ó χ
Curtosis = 0,145 $\pm P_e$ 0.491
Términos = N = 98
$V_1 = 2.7245$ (Momento 1º)
$\mu_2 = 1.7081$ (corrección de id.)
$\mu_3 = 1.1649$ "
$\mu_4 = 9.1772$ "
$B_1 = 0.2723$ (función del momento)
$\sqrt{B_1} = 0.5218$ (para la inclinación)
$B_2 = 3.1454$ (función del momento "V_2")

Distancia modal = $(I \times \sigma) = 8.23\,(+)$
Modo = Medi + dist. modal = 102.76 + 8.23 = 110,99

Area del histograma

Σ. de frecuencias por valor del intervalo = 98 × 20 = 1960
Su mitad (área) = 980
Area de la 1ª barra = 2 × 20 = 40.–
Area de la 2ª barra 16 × 20 = 320.–
Area de la 3ª barra 27 × 20 = 540
Las 3 barras = Area = 900 –
El resto hasta 980 (mitad del area total es una fraccion de la 4ª barra, cuya area parcial es 29 × 20 = 580
Esta fracción puede referirse á fraccion de intervalo y nos dá la adicion para obtener la mediana
(580 : 20 :: 80 : α)
α = 2.76

Mediana del área, expresada en valores de intervalo:

$$\left.\begin{array}{l} 1^a \text{ barra} = 40 \\ 2^a \text{ barra} = 320 \\ 3^a \text{ barra} = 540 \\ \text{fracción } 4^a \text{ barra} = 80 \end{array}\right\} 980 = 102.76$$

8.3/ Binomio de Newton – Probabilidades binomiales – (términos limitados, finitos.)

1.° Las permutaciones posibles de darse un fenómeno de dos modos distintos (cara + cruz), en números dados (finitos) de unidades, estaban dadas por el desarrollo del binomio de Newton y de ello se deducía la probabilidad de cada uno. El binomio por eso, es base de distribución de frecuencias y probabilidades, con sus coeficientes y exponentes. Los coeficientes marcan los grupos de permutaciones; los exponentes marcan el modo ó forma del fenómeno. – Conociendo el desarrollo del binomio, se saben las probabilidades de dos contingencias en número variable de casos, — Expresión general: $(a+b)^n$ —

2.° Desarrollo del binomio

a) Coeficientes: Triangulo de Pascal ó de Tartaglia

		número de casos
1 . . . 1 ———	Potencia	1.ª
1 . . . 2 . . . 1 ———	id	2.ª
1 . . . 3 . . . 3 . . . 1 ———	id	3.ª
1 . . . 4 . . . 6 . . . 4 . . . 1 ———	id	4.ª
1 . . 5 . . . 10 . . . 10 . . . 5 . . . 1 ———	id	5.ª
1 . . . 6 . . 15 . . . 20 . . 15 . . . 6 . . 1 ———	id	6.ª

etc — Son los términos del desarrollo

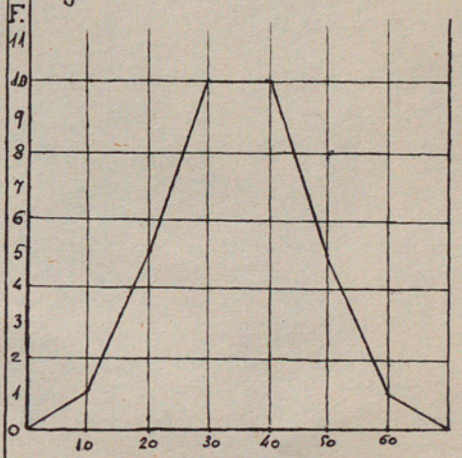
b). Exponentes de los dos modos ó contingencias (a . b)

- 1.er término: El 1.er modo (a) solo, con la potencia máxima.
- siguientes. a; decrece su exponente y b, acrece, de 1 en 1.
- último : El 2.° modo (b) solo, con la potencia máxima. –

Ej: $(a+b)^4 = \underline{a^4} + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$
(son productos en cada término)

Serie	Intervalos	Frecuencias
0.08	1 a 10	1
0.11, 0.13, 0.15, 0.17, 0.19	+ 10 à 20	5
0.21, 0.22, 0.23, 0.24, 0.25, 0.26, 0.27, 0.28, 0.29, 0.30	+ 20 a 30	10
0.31, 0.32, 0.33, 0.34, 0.35, 0.36, 0.37, 0.38, 0.39, 0.40	+ 30 a 40	10
0.42, 0.43, 0.45, 0.47, 0.49	+ 40 à 50	5
0.53	+ 50 à 60	1
		$\Sigma f = 32$

Diagrama binomial de frecuencias. –



Del diagrama binomial de frecuencias e intervalos limitados y finitos, pasando á un universo estadístico de frecuencias é intervalos infinitos, fenómenos biológicos, del azar, etc, se constituye la curva especial de frecuencias y probabilidades, llamada de Gauss. –

Problema = Probabilidades de vida y muerte (a y b) en grupo de 3 enfermos. $(v+m)^3$.
Sobre 3 casos de tifoidea, sin antibióticos. Letalidad = 10% = Binomio $(0{,}9+0{,}1)^3$

Desarrollo – $0{,}9^3 + 3 \cdot 0{,}9^2 \cdot 0{,}1 + 3 \cdot 0{,}9 \cdot 0{,}1^2 + 0{,}1^3 = \underline{0{,}729 + 0{,}243 + 0{,}027 + 0{,}001}$

Probabilidad de que vivan los tres	=	72,9 %
Probabilidad de que vivan 2 y muera 1	=	24,3 %
Probabilidad de que viva 1 y mueran 2	=	2,7 %
Probabilidad de que mueran los 3	=	0,1 %
		100,0

a) j. II) Eliminación y valores de $\underline{a}$ • $\underline{b}$

a - Regla determinantes Cramer:

$$a = \frac{(\Sigma y . \Sigma x^2) - (\Sigma x . \Sigma xy)}{n \Sigma x^2 - (\Sigma x)^2} ; \underline{a} = 166,19 \quad \Bigg| \quad b = \frac{n \Sigma xy - (\Sigma x . \Sigma y)}{n \Sigma x^2 - (\Sigma x)^2} , \underline{b} = -6,57$$

b - Reducción.

$$\left.\begin{array}{l} \text{I} - 1481,82 = \underline{12} a + 78 b \\ \text{II} - 8692,31 = \underline{78} a + 650 b \end{array}\right\}$$

Relación entre 78 y 12 (coefic. de $\underline{a}$ en ambas) $= 78:12 = 6,5$

multiplicando los dos miembros de I por 6,5, la ecuación no varía:

$$\begin{array}{lrcl} \text{I}' & 9631,83 & = & 78 a + 507 b \\ \underline{\text{II}'} & \underline{8692,31} & \underline{=} & \underline{78 a + 650 b} \\ & 939,52 & = & 0 \quad + -143 b \end{array}$$

y restando ambas

$; \underline{b} = 939,52 : -143 = \underline{-6,57}$

Aplicando el valor de $\underline{b}$ a las ecuaciones I y II, obtenemos el de $\underline{a}$

$1481,82 = 12 a + 78(-6,57)$; $1481,82 = 12 a + (-512,46)$

; $1994,28 = 12 a$; $\underline{a} = 1994,28 : 12 = \underline{166,19}$

3º Fórmula de la recta y trazado:

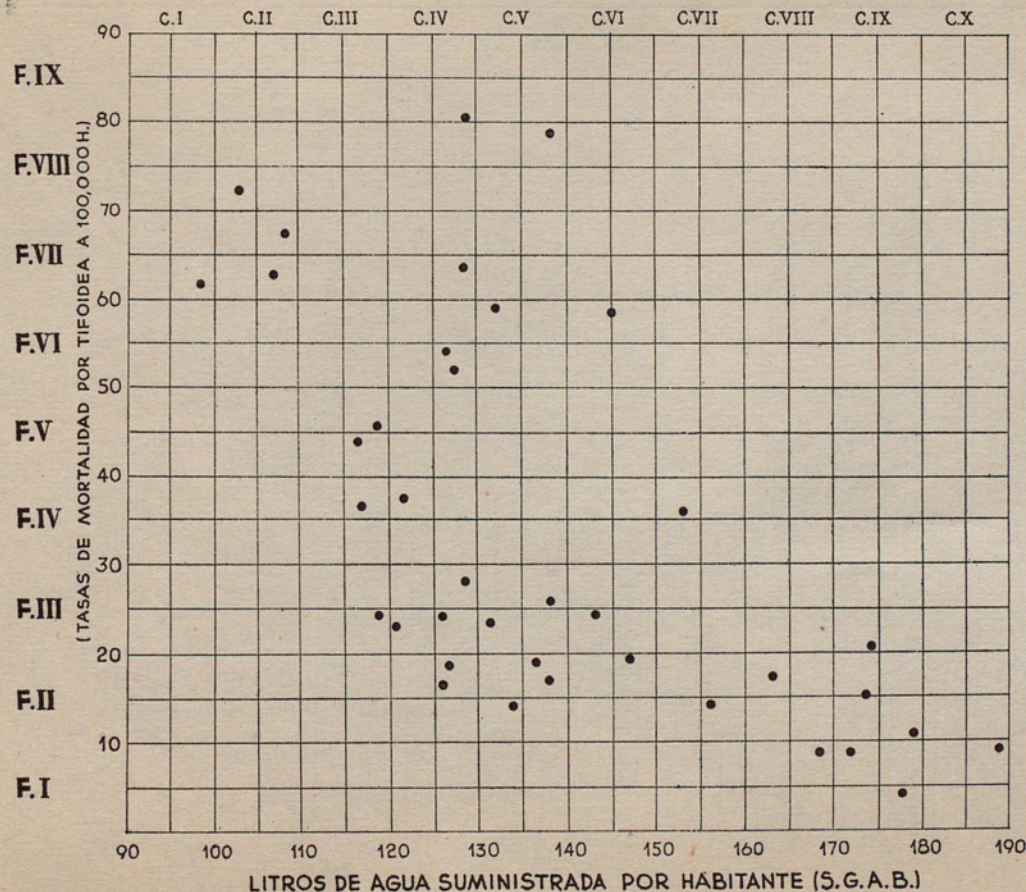
$y = a + bx$; En el 1er punto de la recta $x = 1$; $y = a + b = 159,62$

En el 12 punto de la recta $x = 12$; $y = a + 12b$; $= 87,35$

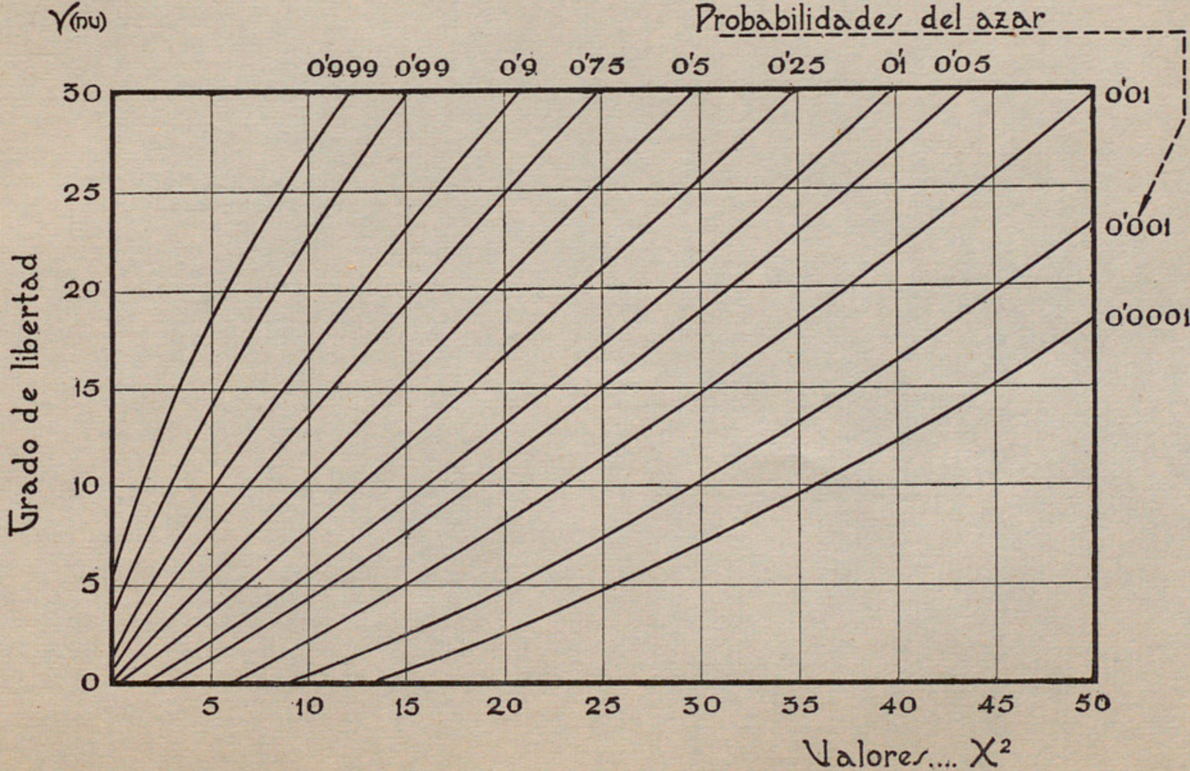
con cuyos dos puntos se traza la línea intercalada en el polígono de frecuencia

CORRELACION ENTRE MORTALIDAD POR TIFOIDEA Y SUMINISTRO DE AGUA EN BARCELONA · 1910 - 1948

1º TABLA DE CORRELACION



Nomograma de los grados de libertad ν en función de χ^2 para las probabilidades Φ



Ma/

Promedios.

Promedio es el valor de una serie que indica una cuantía de equilibrio entre sus datos y puede re- -presentarles á todos.- Se utilizan especialmente la Media, la Mediana y el Modo ó Valor típico.

Media aritmética = Cociente entre la suma de los datos de una serie y el número de sus términos: Ma.-

Puede determinarse por el llamado procedimiento arbitrario, eligiendo cualquiera de sus datos centrales y anotando en otra columna las diferencias cuantitativas y de signo de cada dato hasta el valor arbitrario; se suman algébricamente estas diferencias, se divide entre el número de términos; el cociente positivo será añadido y el negativo restado, del valor elegido arbitrariamente, obteniendo mediante ello, la Ma exacta.

Propiedades y defectos: No coincide con ninguno de los términos; no refleja ninguna característica de ellos; se influye mucho por los valores extremos. Es el término de mayor probabilidad; hace nula la suma algébrica de las diferencias hasta los datos de toda la serie; es mínima la suma de los cuadrados de esas diferencias. La Ma de varias series de igual número de términos, es la Ma de las medias de todas ellas. Es el baricentro ó centro de suspensión de la serie.

$$Ma = (a + b + c + \ldots + n) : n$$

Media geométrica. Es el producto de los términos de la serie del que se obtiene despues la raiz de grado del número de sus datos.- Con logaritmos hay que operar desde número superior a tres términos. Se utiliza como promedio de elección para términos que son proporcionales á la suma de la serie, promedios de cocientes, de porcentajes, series de números índices. La media geométrica de varias series es el producto de las geométricas de todas ellas.

Propiedades. Es siempre menor que la Ma.- Si algun término es cero, el producto hace nula la Mg.- Si son valores negativos la Mg es imaginaria

$$Mg = \sqrt[n]{a.b.c \ldots n}$$

Media potencial. Se elevan á la potencia m los términos de la serie; se suman; se divide entre el número de términos y se extrae la raiz de grado m:

$$\sqrt{\frac{a^2 + b^2 + c^2 + \ldots + n^2}{n}} \; ; \; \sqrt[3]{\frac{a^3 + b^3 + c^3 + \ldots + n^3}{n}} \; ; \; \sqrt[4]{\frac{a^4 + b^4 + c^4 + \ldots + n^4}{n}}$$

media cuadratica — media cúbica — media bicuadrática

Media ajustada. Serie ordenada en graduatoria; es la Ma de los términos centrales

Media armónica. El recíproco de la Ma de los recíprocos de sus términos

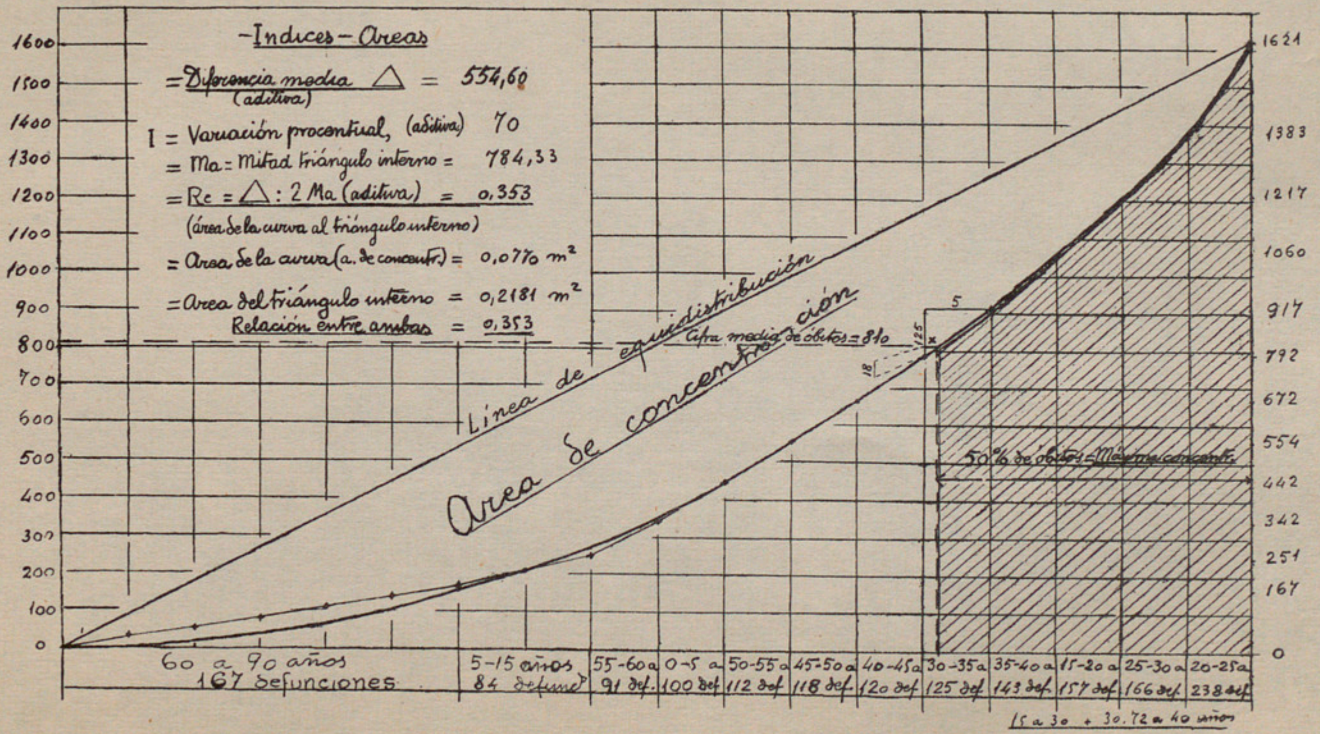
$$\frac{1}{\dfrac{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} + \ldots + \frac{1}{n}}{n}}$$

El mayor valor es el de la media cuadrática y siguen descendiendo la aritmética, la geometrica y la armónica.-

Con. 2

Curva de concentración de Gini (Variación no gaussiana)

Mortalidad por tuberculosis todas formas - Barcelona (ciudad) 1945



6.2/ Desviaciones reducidas. - Universalización de la curva. Tablas de probabilidades. -

- De valores concretos de cada serie por sus desviaciones $\bar{x}$, a los valores abstractos y universales:
  - Desviación reducida "Z" = $\delta : \sigma$. - Sus posibles fracciones en la curva de Gauss, desde 0,01 hasta 4,5 Z. - Idénticos valores de frecuencias y probabilidades, pero universalizados para todas las series -

Tabla de probabilidades de "Sheppard"

Areas medidas por integrales

Valor "Z"	"Probabilidad"	Valores "a"
0.00	0.000	
0.10	0.0398	
0.20	0.0793	
0.30	0.1179	
0.40	0.1554	
0.50	0.1915	
1.00	0.3413 *	± 0.6826
1.50	0.4332	
2.00	0.4772 *	± 0.9544
2.50	0.4938	
3.00	0.49865 *	± 0.99730
3.50	0.4998	
4.00	0.49996	
4.50	0.499997	(media curva de Gauss) = 50%
5.00	-----	

Las cifras de probabilidad se aplican á los protocolos ó series.

Marcan la probabilidad de que un dato nuevo pueda darse con una desviacion igual o mayor que el considerado, ó duplicándola para ±

Las tablas especiales dan resueltos los valores integrales siguientes:

=1º Valor de la ordenada hasta cortar la curva para cada dato $\frac{\delta}{\sigma}$ problema

=2º Proporción del area que queda á la izqda de la ordenada correspondiente á la desviación problema $\frac{\delta}{\sigma}$.

Ej: para $\frac{\delta}{\sigma} = 0$ (media aritmética; área izqda = A = 0,500) 50 %

=3º Probabilidad P, de que un dato se encuentre fuera del margen ó límites de $\pm\ \bar{x} : \sigma$ en la curva

Si con una σ a cada lado de la media se incluyen el 68,26 % de las pruebas, la probabilidad de un valor mayor es solo del 31,74 %

Una desviacion igual ó superior a 2σ de la media solo puede darse (normalmente) en el 4,55 % de las pruebas (no significativo).

Una desviación igual ó superior a 3σ (a cada lado) de la media no alcanza la probabilidad del 1 % (0,27 %) prácticamente nula. (complemento de los valores "a", que marcan las probabilidades de los datos con desviación igual ó menor. Solo pueden estimarse como significativas hasta el tope mínimo de 5 % ($2\sigma = 3Pe$).

Ar 2./

— Medidas utilizadas para conocer la cuantía de la asociación - Coeficiente de asociación - Otros coeficientes.—

La fórmula para deducir el χ^2 de las tablas es según descripción: $\chi^2 = \frac{(\text{Valor real} - \text{valor teórico})^2}{\text{Valor teórico}}$, debiendo ser minuendo del numerador, el término mayor.— La suma del χ^2 de todas las casillas, da el χ^2 total de la tabla.

Procedimiento directo en la tabla tetracórica (sin cálculo de valores teóricos): $\chi^2 = \frac{[(ad)-(bc)]^2 . N}{(a+b).(c+d).(a+c).(b+d)}$ (minuendo, el mayor).

Cuando el valor en una casilla es menor de 5 se debe obtener con la corrección de Yates: $\chi^2 = \frac{[(ad-bc)-(\frac{1}{2}a+b+c+d)]^2.(a+b+c+d)}{(a+b).(c+d).(a+c).(b+d)}$

Medidas de la asociación:

1º. Coeficiente de asociación en las tablas tetracóricas: $\gamma = \frac{(ad)-(bc)}{(ad)+(bc)}$; su valor de máxima asoción es = 1.— De máxima disociación = −1.— Independencia = 0

2º. Por probabilidades del azar con el χ^2.— Determinamos la probabilidad de que por el simple azar puedan darse las diferencias halladas entre los valores reales y teóricos, deduciendo con ello la significación que podemos dar á la confluencia de los caracteres asociados en las tablas.

— Ademas del valor del χ^2 necesitamos conocer el llamado "grado de libertad" de la tabla ó valor nu (ν). Este el número de casillas interiores mínimo necesario de valores reales, para deducir por diferencia hasta los totales de columnas (C) é hileras (H), los valores de las demas casillas; la fórmula de Fisher $\nu = (H-1).(C-1)$, nos dice que $\nu = 1$, en la tabla tetracórica; una sola casilla interior conocida, nos da con las sumas, el valor de las restantes. (Pearson agregaba (+1) en la asociación múltiple, pero los tablas actuales de prob. estan hechas segun la formula de Fisher).— Con los valores de χ^2 y de ν, acudimos á las tablas especiales que nos dan el valor de las probabilidades unitarias (P. total = 1) que podemos referir á tantos por ciento ó por mil, de la influencia del azar, en la clasificación tabulada y segun su cuantía, deducir el valor significativo de la asociación de caracteres.— (Tablas para valores altos de χ^2, para valores bajos, nomogramas, etc, deducidos por cálculo integral).—

Probabilidad de 2,7 ‰ ó menor, es influencia muy baja del azar y por tanto valor "significativo" de la asociacion estudiada.—

Entre 2,7 ‰ y 12 ‰, es "probáblemente significativo"; desde 13 ‰ ó mayor valor, se estima "no significativa" la asociación.—

Otros coeficientes de asociación.

a). Contingencia cuadrática media, φ^2 (psi²) $= \frac{\chi^2}{N}$.

b). Coeficiente de contingencia cuadrática media, $C = \sqrt{\frac{\chi^2}{\chi^2+N}} = \sqrt{\frac{\varphi^2}{\varphi^2+1}}$; sus valores numéricos crecen hasta 1 (Pearson).—

c). Coeficiente de Tschuprow, $T^2 = \frac{\chi^2}{\sqrt{N(C-1).(H-1)}} = \frac{\varphi^2}{\sqrt{\nu}}$; varía entre 0 y 1.— Es el más perfecto y corrige defectos del χ^2.—

As.1/

Asociación cualitativa.- Cuadrado de la contingencia de Pearson ó "test X^2", de la bondad de un ajustamiento.-

Si en el desarrollo de un fenómeno intervienen varios caracteres cualitativos y asociamos sus resultados numéricos, en vez de por porcentajes, con los hechos mismos, pueden influirse los resultados, en cuantía superior á la puramente proporcional (asociación), ó en cuantía opuesta (disociación), ó no variar su proporcionalidad (independencia), variaciones á veces fuertes que llevan á descubrir influencias de causalidad; el resultado de estas relaciones cualitativas, se llama "contingencia".- La asociación de dos caracteres sobre otros dos, generalmente dicotómicos, es la mas elemental ó simple, pero puede tambien establecerse entre varios caracteres sobre otros varios formando la asociación múltiple; para todas ellas, se llevan los caracteres y sus valores resultantes a tablas de asociación, para medir si la influencia de ambos grupos es puramente proporcional ó del azar, ó son significativas.

1° Tabla de los cuatro grupos, tetracórica, dos por dos, four for table, etc.- Se forma lo mismo con la asociación múltiple, con la siguiente distribución y notación, en columnas (C), é hileras (H), cuya suma coincidente de valores, se llama N.

Caracteres	-B- (+) Atacados	-b- (-) No atacados	Totales de hileras
A. Inoculados (+)	FAB. (a) 56	FAb. (c) 10.322	FA (a+c) 10.378
a. No inoculados (-)	FaB. (b) 272	Fab. (d) 8.664	Fa (b+d) 8.936
Totales de columnas	328 FB. (a+b)	18.986 Fb (a+d)	19.314 N (a+b+c+d)

- Tabla tetracórica de valores reales.-

Conociendo las sumas, los valores de todas las casillas pueden deducirse en norma proporcional. Si para 19314 (N), hay una suma de hileras (FA) de 10378, para 328 (FB) habrá en FAB = 176 y para 18986 (Fb) habrá en FAb = 10202; del mismo modo, si para 19314 (N) hay una suma (Fa) de hileras de 8936, para la suma 328 (FB), habrá en FaB = 152 y para 18986, habrá en Fab = 8784, de donde:

$FAB = \frac{FA \cdot FB}{N}$; $FaB = \frac{Fa \cdot FB}{N}$; $FAb = \frac{FA \cdot Fb}{N}$; $Fab = \frac{Fa \cdot Fb}{N}$. Con estos resultados de cada casilla, se forma (lo mismo en la asociacion multiple), la tabla de valores del azar, proporcionales, teóricos ó independientes, que serian para dicho ejemplo, con las mismas sumas: →

Valores teóricos de la tabla:

176 (a)	10.202 (c)	10378 (a+c)
152 (b)	8784 (d)	8936 (b+d)
328 (a+b)	18986 (a+d)	19314 = N (a+b+c+d)

- Los valores teóricos tienen el caracter de probabilidades del azar. Las diferencias entre el valor real y el teórico (sin signos), en cada casilla, elevadas al cuadrado (para que los calculos finales sean positivos) y divididas entre los valores teóricos propios, miden la diferencia entre el azar y la especificidad, relación que se llama X^2, y sumados estos valores de todas las casillas, forman el "test X^2 de toda la clasificación. En cada casilla un valor real superior al teórico expresa asociacion; si es menor, disociación; si son iguales independencia. Cuando el X^2 total es cero, son iguales los valores reales y teóricos, no hay asociación; no hay significación en los valores reales obtenidos de un protocolo; cuanto mas alto es el X^2, tanto mas significativa, especifica, que llega a veces a la causalidad y apartada de la influencia del azar, es la asociacion de los caracteres; su medida, se hace mas ó menos significativa, segun veremos despues.-

El X^2, no puede aplicarse a porcentajes ni proporciones, solo á datos absolutos de los hechos de observacion; su aplicacion al campo de la investigación, hace hoy dia imprescindible su determinación para dar valor significativo á los resultados obtenidos.

Cuando el X^2 es la unidad, deben reputarse los datos como falsos.- Si el X^2 sale con valor muy distinto en varias observaciones de un mismo fenómeno, las técnicas de los datos son defectuosas; sino se encuentra el defecto, súmense los datos en una sola tabla para sensibilizar el X^2.-

En la asociación múltiple, se utilizan varios caracteres, color variado del pelo, segun razas; resultados de los tratamientos segun diversas técnicas, etc. siendo su determinación la misma.-

Ma 2)

Mediana - Modo.-

Mediana. Es el término que ocupa el lugar central en la serie de datos ó variables cuando está ordenada en graduatoria - Si el número de términos es par, la mediana es la Ma, de los dos términos centrales.

Propiedades = Es el promedio mas usado en demografía y epidemología.- La suma aritmética de las desviaciones de los términos hasta la mediana es un mínimum, respecto de la que pueda haber á otro término de la serie. En series cíclicas puede haber varias medianas, como varias medias, segun el punto de partida

Mo.- Modo, valor modal, valor cuspidal, valor típico. Es la variable que más se repite dentro de una serie ordenada en graduatoria; manifiesta la mayor concentración de datos y en las curvas de frecuencia, es el punto mas alto (cuspidal). En la distribución normal ó gausiana, es el promedio más próximo a la Ma, separados por la distancia modal; en curvas regulares ó simétricas, no inclinadas, la tangente á la curva, paralela al eje de abscisas, marca un punto cuya ordenada es el Mo.

Determinación: Pearson da como valor aproximado: 3 veces la mediana menos 2 veces la media aritmética. En cuanto no son bien regulares las series (gausianas), esta fórmula no es aplicable.- Puede haber dos ó tres modos (plurimodales), cuando las frecuencias se concentran en varias zonas de la serie; en las series distribuidas con intervalos y frecuencias, el modo se busca por interpolación entre los datos de mayor concentración; para ello se aplica la formula siguiente

$$Mo = L + \frac{CF}{F+f};$$

Ejemplo:

Intervalos	Frecuencias
2.50 á 2.99	2
3.00 á 3.49	8
3.50 á 3.99	23
4.00 á 4.49	30 *
4.50 á 4.99	20
5.00 á 5.49	13
5.50 á 5.99	6
6.00 á 6.49	6
	108

$$Mo = 4.00 + \frac{0.50 \times 20}{20 + 23} = 4.23$$

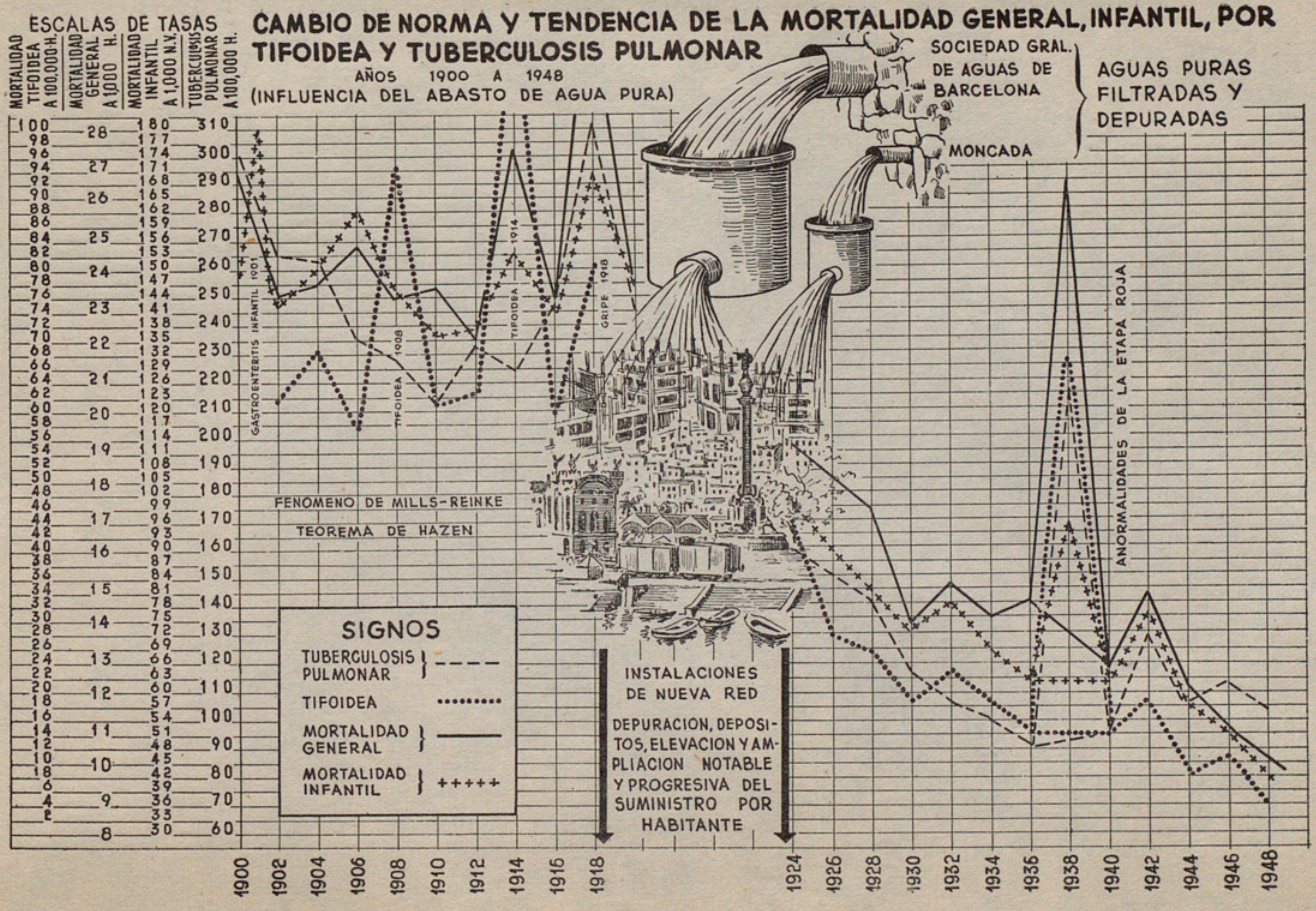
4.23, punto de la serie de intervalos en que incide el modo

El grupo ó intervalo de 4,00 á 4,49, contiene el mayor número de frecuencias, e incluye el valor modal de la serie.

En distribuciones gausianas no simétricas el Mo, se determina por el indice inclinación (χ).- Es el producto de este índice por la desviación standard de la serie ($\chi\sigma$).- El valor del índice de inclinación se dá en la fórmula

$$\chi = \sqrt{\frac{B_1\ (B_2+3)}{2\,(5B_2 - 6B_1 - 9)}}$$

(V. momentos en la curva de Gauss y la σ en el capítulo de "Variación").-


CAMBIO DE NORMA Y TENDENCIA DE LA MORTALIDAD GENERAL, INFANTIL, POR TIFOIDEA Y TUBERCULOSIS PULMONAR
AÑOS 1900 A 1948
(INFLUENCIA DEL ABASTO DE AGUA PURA)
ESCALAS DE TASAS
MORTALIDAD TIFOIDEA A 100.000 H.
MORTALIDAD GENERAL A 1.000 H.
MORTALIDAD INFANTIL A 1.000 N.V.
TUBERCULOSIS PULMONAR A 100.000 H.
SOCIEDAD GRAL. DE AGUAS DE BARCELONA
MONCADA
AGUAS PURAS FILTRADAS Y DEPURADAS
GASTROENTERITIS INFANTIL 1901
TIFOIDEA 1908
TIFOIDEA 1914
GRIPE 1918
FENOMENO DE MILLS-REINKE
TEOREMA DE HAZEN
ANORMALIDADES DE LA ETAPA ROJA
SIGNOS
TUBERCULOSIS PULMONAR
TIFOIDEA
MORTALIDAD GENERAL
MORTALIDAD INFANTIL
INSTALACIONES DE NUEVA RED
DEPURACION, DEPOSITOS, ELEVACION Y AMPLIACION NOTABLE Y PROGRESIVA DEL SUMINISTRO POR HABITANTE
1900 1902 1904 1906 1908 1910 1912 1914 1916 1918 1924 1926 1928 1930 1932 1934 1936 1938 1940 1942 1944 1946 1948

~~44 pg~~

44-1 = 4r

Ex libris