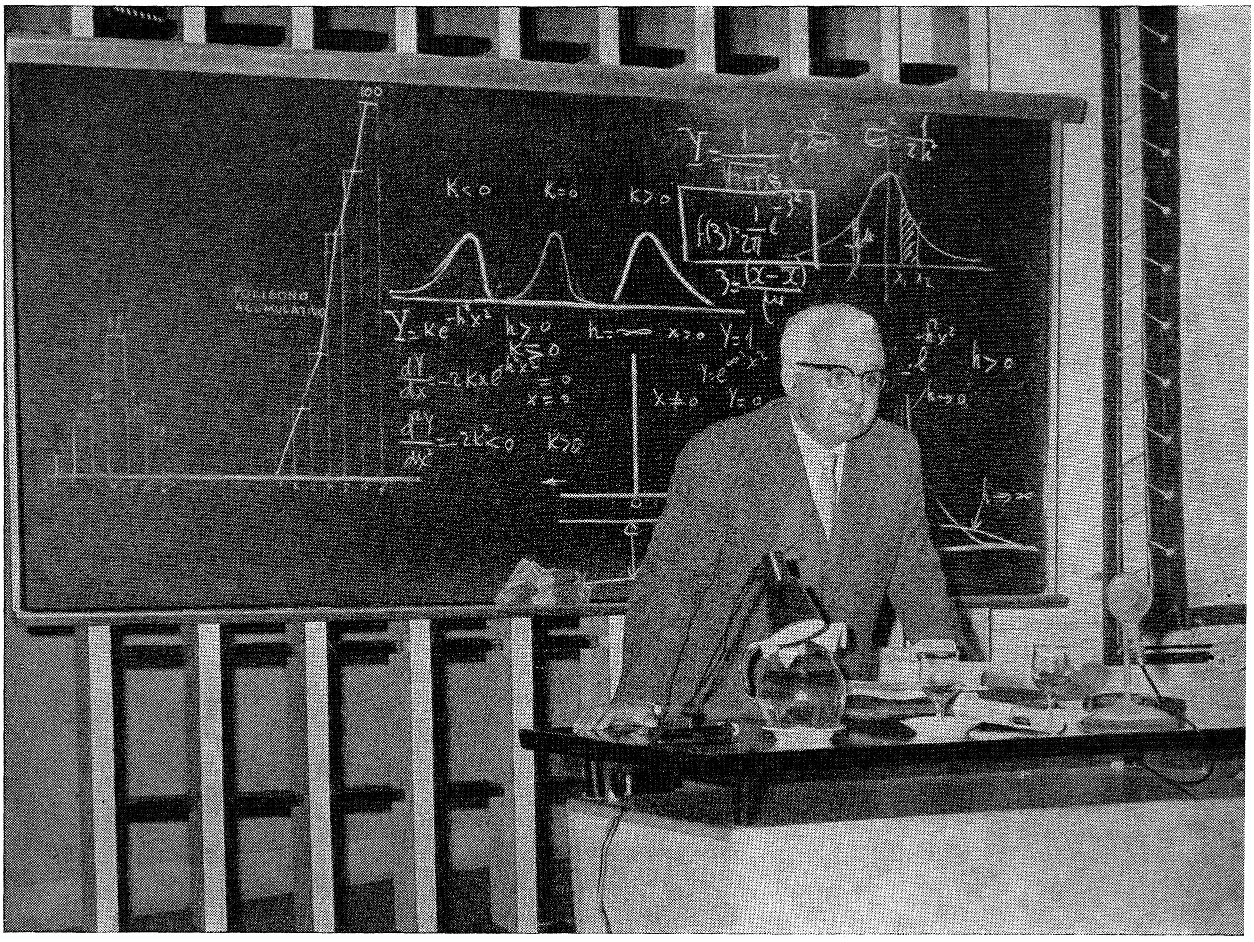
Materiales de Construcción Vol. 15, nº 118-119
Abril-septiembre de 1965



6.3. Matemáticas aplicadas al control de la calidad de los cementos

ANTONIO SARABIA GONZÁLEZ
Dr. Ingeniero Industrial
S. A. Española de Cementos Portland

índice

http://materconstrucc.revistas.csic.es

introducción

Aunque la Academia no haya incorporado a su diccionario la palabra «control», parece fuera de duda que, bien o mal empleada, ha entrado en el caudal lexicológico del castellano y, por tanto, ha merecido la atención de los estudiosos de nuestro idioma.

Así, el filólogo García de Diego afirma, en su «Diccionario etimológico español e hispánico», que control, en castellano, como contrôle, contrôler y controleur, en francés, derivan todos de «contra rotulum» (contra lista), y asigna al sustantivo control las acepciones de: *observación, mando, comprobación,* que convienen perfectamente a las metas perseguidas cuando se aplica aquel sustantivo al concepto «calidad».

Antaño, la palabra castellana «Contralor» designaba al «veedor o interventor de las cuentas de la Casa Real» (García de Diego, o. c.).

El vocablo «calidad» quiere decir, en su amplio significado, el grado que exhibe en su «aptitud funcional» una entidad (p. e. el cemento) para satisfacer una exigencia humana (en nuestro ejemplo, la construcción). Visualizada así la idea «calidad», contiene, implícitamente, la referencia a otra entidad, que llamaremos «EL MODELO», a veces, puramente conceptual, otras, con tangible realidad física, que sirve, por comparación, para determinar el grado de aptitud, aludido antes, que ofrece el objeto cuya «calidad» se pretende establecer.

El «modelo» suele estar definido por el grupo de características, generalmente muy diversas, que satisfacen, en el grado más alto posible, las exigencias funcionales vigentes en una época dada.

Su elevado número y heterogeneidad hacen, a veces, muy difícil la formulación del modelo y de la calidad. Por ello, es obligado recurrir a las llamadas *reglas de calidad,* que son criterios para apreciar y valorar, por comparación, propiedades concretas.

La recogida y análisis de información básica para el establecimiento del «modelo» debe constituir un arduo quehacer. En cierto sentido, las especificaciones contenidas en Normas, Pliegos de Condiciones, etc., constituyen las bases del modelo, y aún hoy, a ciento cuarenta años de Apsdin, no hay coincidencia entre los distintos países acerca de lo que entiende cada uno por CEMENTO PORTLAND, con todas sus consecuencias comerciales, técnicas, económicas y hasta psicológicas.

Citemos, por ejemplo, que la cifra digamos «clave» de la *resistencia mecánica,* a distintas edades, en las diversas especificaciones, puede conducir a equívocos si, al momento de comparar las de dos países, no se toma en consideración la relación agua/cemento del mortero sobre el que han sido realizadas las mediciones para determinar aquélla. (Compárese, por ejemplo, los Pliegos vigentes alemán y español.)

El hecho de poder establecer normas, especificaciones o, tomadas en conjunto, modelos, constituye una clara afirmación de que la tecnología a que se refieren está presentando o puede presentar en mercado, multitud de ejemplares, o cifras inmensas de porciones discretas, obtenidas en proceso continuo que, a lo largo del tiempo de modo constante, satisfacen a aquéllas.

Exigencias funcionales y el modelo de entidad capaz de satisfacerlas son históricos. En cada época, dan testimonio del grado de desarrollo de la tecnología a que se refieren. Nacen y crecen engendrándose en y nutriéndose de los hallazgos de la Ciencia, previamente reelaborados, y hasta de los avances de tecnologías, afines o no, que le resultan asimilables.

Buena prueba de ello la ofrece la historia del desarrollo de las especificaciones, punto de partida para establecer la definición del modelo en el caso del cemento, detalladamente descrita por Mayer en «Normen für die Festigkeitsprüfung von Zemente» (Z.K.G., 1, I. 64).

Las ventajas y hasta la obligación de obtener productos de calidad tan alta como sea posible, en cada época, a nuestro juicio sobrepasa las zonas de la operación y concurrencia comerciales, de modo especial, si se considera socialmente, ya que tiene el significado del máximo aprovechamiento de los materiales primarios, puestos por Dios a nuestro servicio, a través de la técnica más perfecta, creada por la Humanidad a lo largo de los siglos.

Por ello, no es extraño que estimemos la calidad como una meta interesantísima del quehacer productor, para el cual se obtiene entonces, como permanente «leitmotiv», el logro del máximo caudal con la calidad más alta. Así lo viene probando la industria del cemento española, cuando se examinan sus resultados, especialmente en estos últimos años.

Cuando definíamos la calidad, señalábamos ya que el concepto implicaba una comparación con el modelo. El caso ideal sería poder definir un modelo perfecto, y que la realización industrial lo repitiese de modo teóricamente exacto.

Puede afirmarse que tal situación no es alcanzable. La producción ha de realizarse en dispositivos inmersos en un medio variable, cuya ley de variación no es previsible. Los elementos que constituyen la estructura fabril, aunque cada día más evolucionados, tampoco pueden alcanzar la perfección definitiva, y de este modo, aunque la aproximación teórica al conocimiento del proceso sea muy avanzada, el producto a lo largo del tiempo jamás exhibe, como medida de una propiedad, que forme parte de su definición, un valor rigurosamente constante.

En su lugar, cuando se realizan mediciones sobre los ejemplares producidos, se obtienen grupos de valores encuadrados entre ciertos límites. Sin embargo, la cuantía del intervalo entre ellos ha de ser adecuada a la permisible variabilidad que la exigencia humana, que haya de cubrir el producto, lo defina como «prácticamente aceptable», equivalente a decir que los grupos de valores que para cada ejemplar representan la intensidad de las propiedades correspondientes, tienen que moverse dentro de un campo de variabilidad conocida y medible; es decir, el proceso productor ha de ser controlable.

Equivalente a decir también que, en la tecnología de producción, ha sido alcanzado el grado de madurez que permite «a priori» postular la meta de una producción caracterizada por cierto grupo de propiedades y fijar el *margen de imposibilidad* de conseguirlo de modo teóricamente perfecto.

De forma implícita lo que venimos postulando expresa que, en la tecnología de producción que se considera, es conocida la relación causa-efecto en toda su extensión, es decir, se domina el grupo de causas *asignables,* pero venimos afirmando también que, de modo inevitable, acompañan a aquéllas, otras causas, cuyo dominio no poseemos, en cierto modo ajenas a la esencia del proceso mismo y que engendran la imposibilidad de conseguir la constancia de resultados esperables de una rígida ley de causalidad. Es decir, en el proceso productor actúan también *causas de azar, causas aleatorias,* cuya presencia se traduce en la aparición de un enjambre de valores, digamos orientado, en lugar de uno solo, como expresión de la intensidad de la propiedad deseada, al obtener su medida sobre los diversos ejemplares producidos.

El efecto de las causas de azar es expresable mediante la utilización del concepto de variable aleatoria, cuyos valores representan las desviaciones que, con relación al valor esperado, exhiben los obtenidos.

En las circunstancias supuestas para el proceso productor no es posible predecir el valor alcanzado por una propiedad en un ejemplar discreto; sin embargo, se podrá afirmar que dentro de largas series de producción obtendremos en cada ejemplar desviaciones al valor teórico, contenidas dentro de un intervalo dado de la variable aleatoria.

Así, parece perfectamente claro que al estudio de esos grupos de valores, tantas veces comentado, convenga perfectamente la aplicación de la Matemática de los fenómenos aleatorios o estadísticos, tanto en el sentido descriptivo como en el inductivo; que utilizando como instrumento el Cálculo de Probabilidades permite obtener conclusiones acerca de conjuntos extensos o poblaciones de datos a partir de una muestra o conjunto parcial de los mismos.

Creemos adecuado recordar aquí las siguientes definiciones y postulados, válidos solamente cuando los sucesos están regidos por causas de azar:

a) Se entiende por probabilidad matemática de la realización de un suceso, al cociente:

$$p = \frac{\text{número de casos favorables}}{\text{número total de casos}}$$

$$p = 1 \quad \text{y} \quad p = 0$$

significan, respectivamente, la certeza y la imposibilidad de tal realización.

Es válida la ecuación:

Número de realizaciones favorables + número de realizaciones no favorables =
= número total de realizaciones

de la que se obtiene llamando q a la probabilidad de la realización del caso no favorable:

$$1 = p + q \quad \text{ó} \quad q = 1 - p$$

b) La probabilidad matemática de que se produzca uno u otro de dos sucesos que se excluyen, es igual a la suma de las probabilidades de realización de cada uno de ellos:

$$p = p_1 + p_2$$

c) La probabilidad matemática de realización simultánea de dos sucesos, independientes entre sí, es igual al producto de las probabilidades individuales de cada uno de los sucesos:

$$p = p_1 \cdot p_2$$

d) Ley de Bernouilli.

Si un suceso se realiza s veces en n ensayos, la probabilidad p de aparición del suceso se aproxima al valor:

$$p = \frac{s}{n}$$

tanto más cuanto mayor es n.

e) La probabilidad de que un suceso se produzca r veces en n experimentos vale·

$$P = \binom{n}{r} p^r (1 - p)^{n-r}$$

en la que p es la probabilidad de aparición en un solo ensayo.

f) La probabilidad P de aparición de un suceso *por lo menos* r veces en n ensayos vale:

$$P = \sum_{h=r}^{n} \binom{n}{h} p^h (1 - p)^{n-h}$$

en la que p tiene igual significado que antes.

Parece interesante también dedicar un momento a la sistemática de la descripción de los conjuntos de valores incluidos en una serie estadística o aleatoria.

Suelen aplicarse:

Tablas estadísticas cuantitativas que reúnen todas las mediciones realizadas.

Tablas de frecuencias y su representación gráfica en el *diagrama de barras.*

Tablas de frecuencias relativas en las que para intervalos Δx de la variable aleatoria se anota el valor de la frecuencia relativa:

$$\frac{n_{\Delta x_i}}{\Sigma n_{\Delta x_i}} = f_{\Delta x_i}$$

$n =$ número de veces que se realiza el suceso para los valores de la variable aleatoria contenidos en Δx_i. Es claro que:

$$\sum_{i=1}^{n} f_{\Delta x_i} = \frac{n_{\Delta x_1}}{\Sigma n_{x_n}} + \frac{n_{\Delta x_2}}{\Sigma n_{x_n}} + \ldots \frac{n_{\Delta x_n}}{\Sigma n_{\Delta x_n}} = 1$$

Si cierto intervalo de la variable aleatoria se subdivide en n partes iguales, llamados intervalos de clase, y se anota la frecuencia relativa correspondiente a cada intervalo, se dispone de:

La Tabla de recuento y de *distribución de frecuencias.*

Su representación gráfica, llamada *histograma,* se construye tomando como eje de abscisas el intervalo $x_2 - x_1$ de la variable aleatoria, dividiéndolo en n partes iguales y levantando en el punto medio de cada subdivisión la ordenada con longitud proporcional a la frecuencia relativa, correspondiente al intervalo.

Si a cada intervalo de clase hacemos corresponder el valor de la frecuencia relativa dividida por la magnitud del intervalo mismo como ordenada, se obtiene la tabla de distribución de *densidad de frecuencias* y su correspondiente representación gráfica, el histograma de distribución de densidad de frecuencias. En el histograma de densidad de frecuencias, el área de cada rectángulo

$$A = \frac{x_2 - x_1}{n} \cdot \frac{f_{\Delta x_i}}{\frac{x_2 - x_1}{n}} = f_{\Delta x_i}$$

es igual a la frecuencia relativa correspondiente al intervalo $\left(\frac{x_2 - x_1}{n}\right)_i$ y el área total del histograma de distribución de densidad de frecuencias:

$$\Sigma f_{\Delta x_i} = 1$$

Si manteniendo la misma división de los intervalos de clase que en el histograma de densidad de frecuencia y por el punto medio de aquéllos se levanta una ordenada que represente, a escala, la suma de áreas de los rectángulos parciales del diagrama de densidad correspondientes al intervalo que se considera y a los que le anteceden, se obtiene el *diagrama de frecuencias acumulativo.*

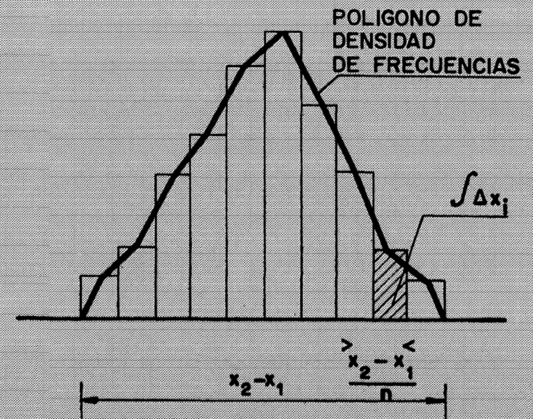


Fig. 1.

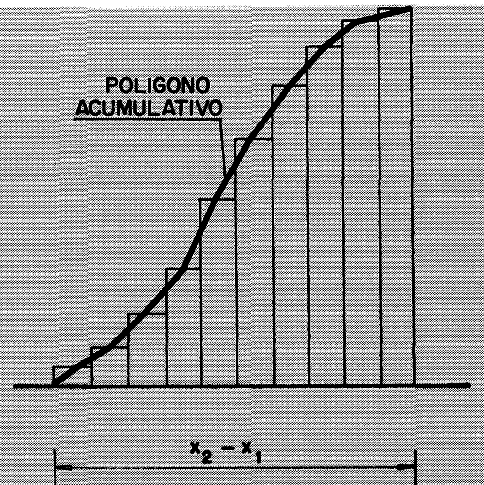


Fig. 2.

Si se unen con una recta los puntos medios de las bases superiores de los rectángulos parciales de ambos diagramas, se obtienen los polígonos de densidad de frecuencias y el polígono acumulativo. (Ver figuras 1 y 2).

Para completar la descripción y, al mismo tiempo, conseguir que sea asequible al cálculo, se utilizan unos números llamados *estadísticos* o *características estadísticas.*

Suelen aplicarse del modo más frecuente dos estadísticos, que se refieren a:

a) *La tendencia central.*—El valor que, por así decir, agrupa el conjunto aleatorio.

b) *La dispersión.*—Número que ha de informar de cómo están distribuidos los distintos valores del conjunto alrededor del valor central.

Un tercer estadístico da cuenta, cuando es necesario, acerca del grado de simetría del universo o población estadística con relación al valor central.

Estadísticos para la tendencia central

El más usado:

La media aritmética: $\bar{X} = \dfrac{x_1 + x_2 + \dots + x_n}{n} = \dfrac{\sum\limits_{i=1}^{n} x_i}{n}$

Si se consideran los n valores x_i como determinaciones de la posición sobre una recta, de n partículas del mismo peso, la media aritmética corresponderá al centro de gravedad del sistema.

La media geométrica: $\sqrt{x_1 \cdot x_2 \cdot \dots \cdot x_n}$

La mediana o valor situado más en el centro de las observaciones; o sea, el valor que divide en partes iguales al número de observaciones, cuando aquéllas se ordenan según sus valores.

La moda, que es el valor que exhibe la frecuencia más alta.

Estadísticos para la dispersión

El más usado es la *desviación típica* o *desviación standard:*

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}} = \sqrt{\frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}{n}}$$

Formalmente, significa el *radio de giro,* medido desde el centro de gravedad de un sistema de n partículas, del mismo peso, que estuvieran dispuestas a lo largo de una línea y a las distancias x_i de aquél.

Otras medidas de dispersión

El coeficiente de variación: $v = 100 \dfrac{\sigma}{\bar{x}}$

Número sin dimensiones. Permite comparar las distribuciones entre sí.

La desviación media: $\sum\limits_{i=1}^{h} \dfrac{(x_i - \bar{x})}{n}$

El recorrido $\bar{R}$, de amplia aplicación en el control de calidad, significa la diferencia entre los valores extremos, máximo y mínimo, de la serie.

La varianza: $\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$

Las tablas de distribución de frecuencias y su representación gráfica obtenidas en la práctica, a veces son simétricas con relación al valor central y otras no.

El estadístico que permite medir la simetría está definido por:

$$K = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{n\sigma^3}$$

Es una magnitud sin dimensiones; puede ser positiva o negativa.

Para distribución simétrica: $K = 0$.

Para $K < 0$ la cola de la distribución está a la izquierda del valor central, y para $K > 0$ a la derecha (en la representación gráfica).

En el campo de las aplicaciones, los estadísticos de mayor interés son

media aritmética, $\bar{x}$,

la desviación típica, σ,

el recorrido, $\bar{R}$,

a los que añadiendo el número de observaciones, n, componen la definición suficiente para una población o universo estadístico.

Si imaginamos que los intervalos de clase del histograma de frecuencia van haciéndose cada vez más pequeños y sustituimos la palabra frecuencia por probabilidad, los polígonos de frecuencias y acumulativo tienden a convertirse en la curva de distribución de densidad de probabilidad y en la curva de probabilidad, respectivamente.

La forma de estas curvas depende del tipo de distribución de probabilidad que presente el universo aleatorio estudiado que, a su vez, será consecuencia de las causas de azar que lo engendran.

Se debe a Gauss la formulación de la ley que relaciona el valor Y de la densidad de probabilidad de un suceso (por ejemplo: un error accidental de magnitud dada, la diferencia del valor deseado al obtenido en la medición de una propiedad) con el valor de la variable aleatoria (error accidental, diferencia entre valor esperado y obtenido).

La ley de Gauss tiene por expresión: $Y = Ke^{-h^2x^2}$ con $\begin{matrix} K > 0 \\ h \gtrless 0 \end{matrix}$

que se conoce como la ley de distribución normal de Gauss, en la que Y es la densidad de probabilidad de aparición del valor x de la variable aleatoria en el universo considerado.

Citaremos, por su importancia, las propiedades de esta ley de distribución de densidad de probabilidad:

a) la función es simétrica; Y es invariable para $\pm x$;

b) posee un máximo, para $x = 0$;

c) presenta dos puntos de inflexión correspondientes a:

$$x = \pm \frac{1}{h\sqrt{2}}$$

Como se ve, simétricos, con relación a $x = 0$, aproximándose a este valor al crecer $\overline{h}$;

d) la curva es simétrica con relación al eje de ordenadas.

K es una constante *no esencial.* Para distintos valores de K (y uno dado para h) se obtiene siempre la misma curva. Basta hacer variar la escala de Y proporcionalmente a K.

h, por el contrario, es una constante esencial, puesto que con

$$K = 1, \text{ si hacemos } h = 0$$

para todo el intervalo $-\infty < x < +\infty$ obtenemos constantemente:

$$Y = 1$$

La curva se ha convertido en una paralela al eje de abscisas. Todos los valores que se tomen para la variable aleatoria convienen con la medición que se realiza, que es como decir que la propiedad de que se trata no es medible.

$$\text{Si } h \to \infty, \text{ obtendríamos: } \begin{array}{ll} \text{para } x = 0 & Y = 1 \\ \text{para } x \neq 0 & Y = 0 \end{array}$$

La curva de Gauss, en este caso límite, se transforma en la normal al eje de abscisas para $x = 0$, y coincide con ese mismo eje para todo punto comprendido en los intervalos:

$$-\infty \to 0 \to +\infty$$

Esta sería la representación del resultado de mediciones de propiedades sobre objetos obtenidos por un proceso regido por una ley de causalidad pura.

Entre los dos límites correspondientes a $h = 0$ y $h = \infty$ discurre la curva de Gauss. Su forma de campana resulta aplanada, con puntos de inflexión alejados del punto $x = 0$ para $h = 0$, y con forma más aguda y puntos de inflexión en la proximidad de $x = 0$ cuando $h \to \infty$.

La constante h es una medida de la precisión de las mediciones que originan el universo aleatorio.

La figura 3 reproduce una familia de curvas de distribución de densidad de probabilidad siguiendo la ley de Gauss para distintos valores de h.

Por su carácter de curva de distribución y de densidad de probabilidad el área del rectángulo elemental comprendida en el intervalo $x + dx$ (fig. 4), el eje de abscisas y la curva, mide la probabilidad de que aparezca un valor de la variable aleatoria comprendido entre x y $x + dx$, o sea:

$$dP = Y \cdot dx$$

y, por tanto, la probabilidad de que obtengamos para la variable aleatoria un valor que esté comprendido entre x_2 y x_1, será:

$$P = \int_{x_1}^{x_2} Y \cdot dx = K \int_{x_1}^{x_2} e^{-h^2x^2 . dx} \qquad \text{(I)}$$

o en palabras, la probabilidad de obtener un valor de la variable aleatoria comprendido entre dos valores de x_2 y x_1, de ésta, puede medirse por el área limitada por las ordenadas $y_2 - y_1$, la curva de distribución y el eje de las x.

La función o curva de Gauss, por utilización de la propiedad de que su área, como curva de densidad, es siempre la unidad, y mediante la aplicación del cálculo, toma la forma de:

$$Y = \frac{h}{\sqrt{\pi}} e^{-h^2x^2} \text{ (II)} \quad \text{con } h > 0$$

que depende ya solamente de h.

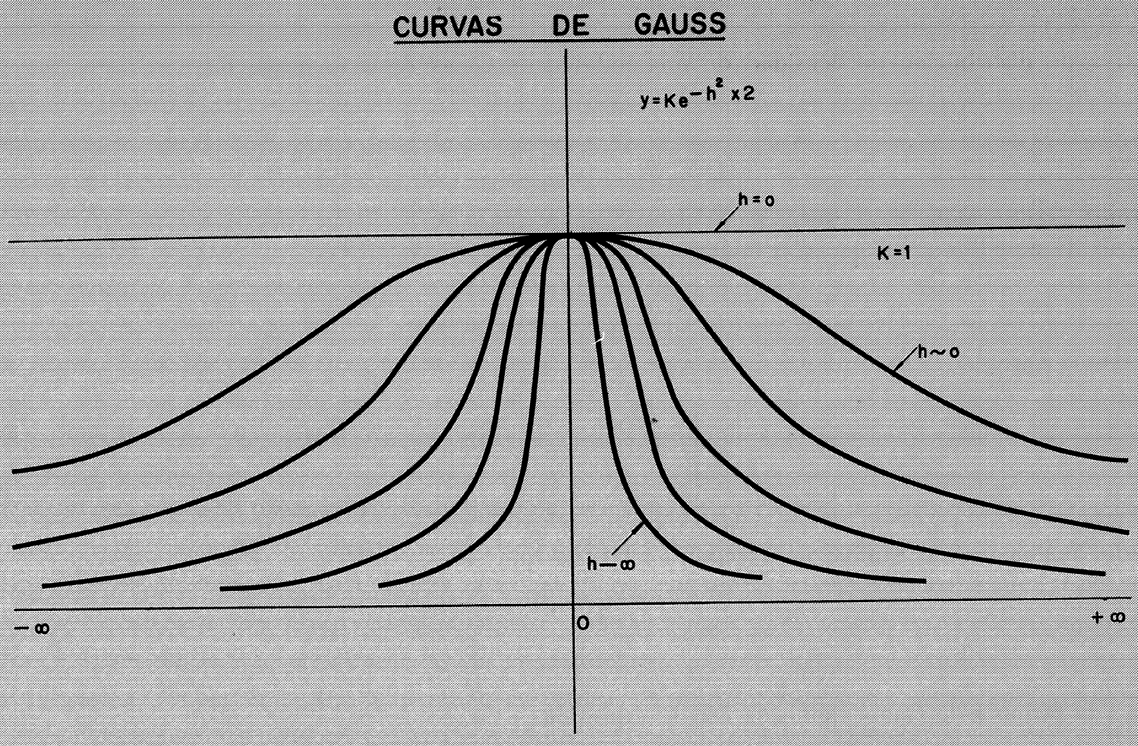


Fig. 3.

Si se hace el cálculo para distintos valores de h, se obtiene un haz de curvas del tipo que reproduce la figura 5, en la que también, a medida que h crece, la curva se aproxima a la normal Y en $x = 0$.

Aplicando a la fórmula anterior consideraciones probabilísticas y de la teoría de máximos y mínimos, y tomando para dispersión a

$$\sigma = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n}}$$

se obtiene que:

$$\sigma^2 = \frac{1}{2\,h^2}$$

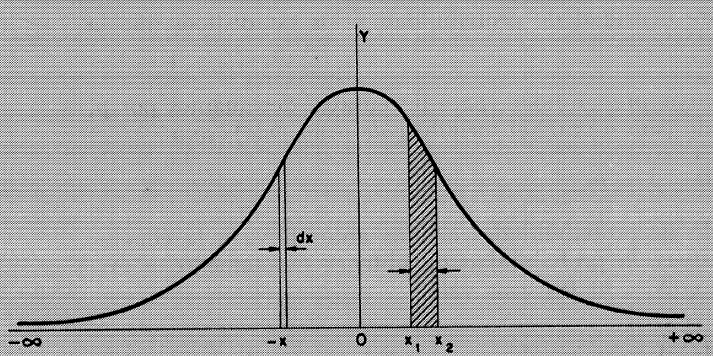


Fig. 4.

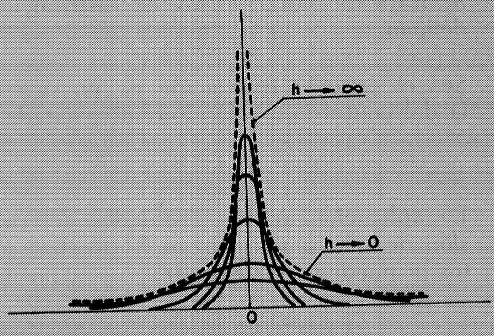


Fig. 5.

y la curva de distribución de densidad de probabilidad de Gauss toma la forma de:

$$Y = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{x^2}{2\sigma^2}}$$

Por último, identificando a x, desviaciones de los valores obtenidos al valor verdadero, con las diferencias de aquéllos a su media aritmética

$$x - \bar{x}$$

y, por tanto, con

$$\sigma = \sqrt{\frac{\sum_{i}^{n} (x_i - x)^2}{n}}$$

que ya hemos denominado desviación típica o standard del universo de valores, que se estudia realizando, finalmente, un cambio de variable en III, tal que

$$Z = \frac{x - \bar{x}}{\sigma}$$

se llega a la fórmula de Gauss tipificada con la expresión:

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

representada simbólicamente por $N(0 - 1)$, pues puede decirse que la variable Z está distribuida normalmente con la media 0 y varianza 1.

La figura 6 reproduce la curva normal tipificada, en la que se han indicado las áreas delimitadas entre $z = 1$; $z = 2$, y $z = 3$.

Existen tablas que dan las áreas definidas entre la ordenada de $z = 0$ y cualquier valor positivo de z; el área correspondiente a dos ordenadas cualesquiera se obtiene fácilmente utilizando la simetría de la curva y sumando.

Otra distribución de densidad de probabilidad es la conocida como *distribución binomial.*

Imaginemos una población o universo de valores regido por leyes de azar. Sea la variable aleatoria el número de éxitos, es decir, el número de veces que en N ensayos ocurre el fenómeno esperado, en la alternativa de que éste se realice o que se realice el fenómeno contrario. Por ejemplo, salir cara o cruz al tirar una moneda; que un tornillo sea aceptable o no en un recuento para determinar su recepción.

Se trata de hallar su ley de distribución de densidad de probabilidad y los estadísticos que la definan.

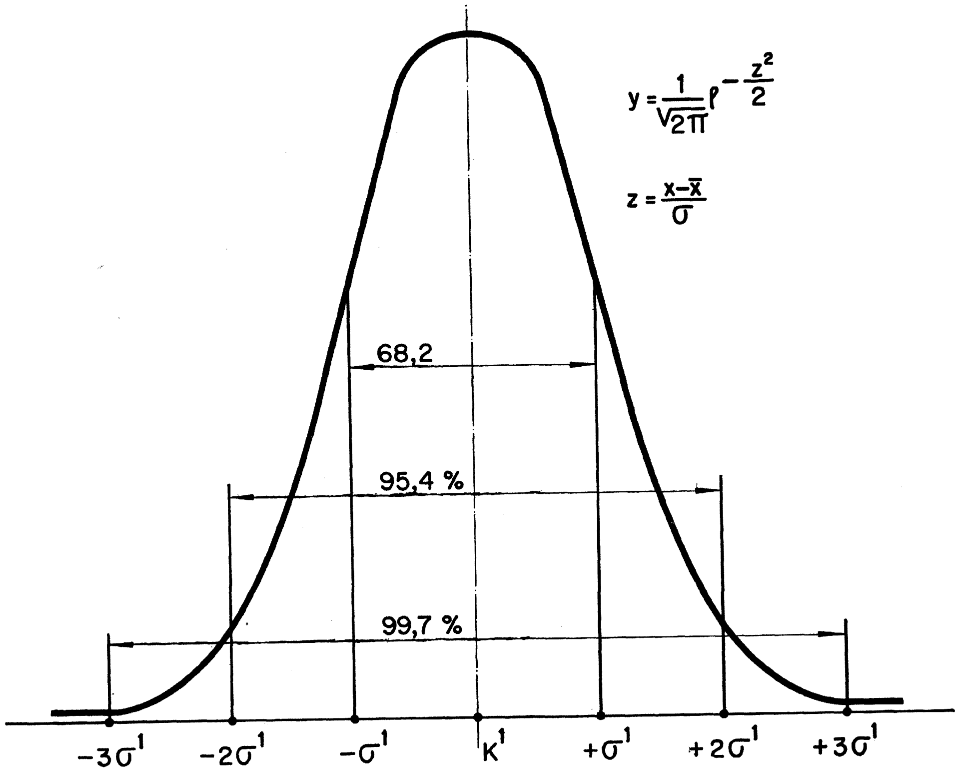
Sea el éxito la realización de un suceso cuya probabilidad, para un ensayo, designamos por p, la del contrario, en la alternativa citada (sale cara o cruz, el tornillo *vale o no vale*), será

$$q = 1 - p.$$

La aplicación de los postulados del cálculo de probabilidades citados antes lleva a la conclusión de que si se realizan N pruebas sucesivas, la probabilidad de obtener exactamente n éxitos (n puede variar entre 0 y N), y, por tanto, $(N - n)$ fracasos vale:

$$P_n = \binom{N}{n} p^n q^{N-n}$$

DISTRIBUCION NORMAL
Curva tipificada



y entonces podemos obtener el siguiente cuadro:

variable n	0	1	2		N
P	$\frac{N!}{0!N!} p^0 q^N$	$\frac{N!}{1!(N-1)!} pq^{(N-1)}$	$\frac{N!}{2!(N-2)!} p^2 q^{(n-2)}$		$\frac{N!}{N!0!} p^N q^0$

que, como puede observarse, contiene los términos del desarrollo del binomio:

$$(p+q)^N = \sum_{r=o}^{N} \binom{N}{r} p^r \cdot q^{N-r}$$

La probabilidad de que el suceso se produzca *por lo menos n* veces con N ensayos será:

$$P = \sum_{n}^{N} \binom{N}{n} p^n (1-p)^{N-n}$$

Los estadísticos para la distribución binomial valen:

la media aritmética: $\bar{x} = N \cdot p$

la varianza: $\sigma^2 = N \cdot p \cdot q$

la desviación típica: $\sigma = \sqrt{N \cdot p \cdot q}$

el coeficiente para simetría: $\alpha_3 = \frac{q - p}{N \cdot p \cdot q}$

Relación entre la distribución binomial y la distribución normal de Gauss

Cuando N es grande y si ni p ni q están muy próximas a 0 la distribución binomial se aproxima a una distribución normal cuya variable tipificada estuviese dada por:

$$Z = \frac{x - N \cdot p}{\sqrt{N \cdot p \cdot q}}$$

Creciendo N la aproximación es mejor y en el límite es coincidente. En la práctica la aproximación es muy buena cuando $N \cdot p$ y $N \cdot q$ son mayores que 5.

A la distribución binomial se le conoce también por distribución de Bernouilli, en honor de James Bernouilli que la postuló en el siglo XVII.

La distribución de Poisson

Cuando p es muy pequeño, para que la distribución binomial se aproxime a la normal es preciso que N sea muy grande. En el caso de que $p < 0{,}1$ y $N \cdot p < 5$ es muy útil, para el cálculo aproximado de probabilidades binomiales, la ley de distribución de Poisson:

$$P_r(x = r) = \frac{\lambda^r}{r!} e^{-\lambda}$$

fórmula límite de las propiedades binomiales cuando $N = \infty$ y con $\lambda = K = n\,p$.

La representación gráfica ofrece claramente la semejanza con la curva de distribución normal cuando $\lambda = n\,p = 5$:

la media en esta distribución es:	$\bar{x} = \lambda$
y la varianza:	$\sigma^2 = \lambda$

Se aplica en los casos en que la probabilidad de realización de un suceso en un ensayo es muy pequeña y se quiere predecir la realización de K sucesos en un número muy elevado de observaciones. Se aplica, en algunos casos particulares del control de calidad.

Las muestras

De modo implícito venimos llamando colectivo, universo o población a un conjunto numérico que expresa la intensidad individual con que n elementos poseen una característica común.

Una parte de tal conjunto se llama muestra.

Las poblaciones son finitas (por ejemplo, los habitantes de Madrid), o infinitas (por ejemplo, los resultados de lanzar una moneda al aire indefinido número de veces).

El estudio directo de un colectivo es a veces difícil o costoso. Se acude entonces a considerar muestras de dicho colectivo. Y, a partir de las propiedades de éstas, se inducen las características del colectivo a que pertenecen.

Este tema es la finalidad de la Teoría de las muestras, importantísimo capítulo de la Matemática estadística.

La muestra ha de ser representativa del colectivo. El estudio de los métodos de tomar muestras y los problemas que surgen de ello se denomina proyecto para el experimento o proceso aleatorio.

Una forma de obtener una muestra representativa es el sistema del muestreo al azar, de acuerdo con el cual cada elemento del colectivo tiene la misma probabilidad de ser incluido en la muestra. Una técnica para obtener muestras al azar es asignar un número a cada elemento, escribir éstos en un papel, mezclarlos cuidadosamente, y obtener así la muestra. La operación puede sustituirse por la aplicación de las Tablas de números de azar.

Distribuciones muestrales

Consideremos todas las muestras posibles de tamaño N que pueden extraerse, con o sin reemplazo, de una población dada. Podemos asociar a la muestra diversos estadísticos y obtener las curvas de distribución correspondientes a las poblaciones que aquéllos forman. Citaremos por su interés:

Distribución de las medias muestrales

Sea una muestra de n elementos tomados de un universo de n_1 elementos ($n_1 > n$), llamando:

$\mu_{\bar{x}}$ = media de la distribución muestral;

μ = media del universo-madre;

$\sigma_{\bar{x}}$ = desviación típica de la distribución muestral;

σ = desviación típica del universo;

cuando n_1 es finito y la muestra se toma *sin reemplazo* (no vuelve a incorporarse a n_1) se obtiene que:

$$\mu_{\bar{x}} = \mu \quad ; \quad \sigma_{\bar{x}} = \sqrt{\frac{n_1 - n}{n_1 - 1}} \quad \frac{\sigma}{\sqrt{n}}$$

Si las muestras corresponden a una población infinita o aquéllas son extraídas *con reemplazo* (vuelve a formar parte del universo, para una nueva toma de muestra), resulta que:

$$\mu_{\bar{x}} = \mu \quad ; \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Cuando n es grande ($n > 30$), la distribución muestral de las medias es aproximadamente una *distribución normal* con $\sigma_{\bar{x}}$ y $\mu_{\bar{x}}$ independientes de n_1 en tanto que el tamaño de la población cumpla la condición

$$n_1 > 2n$$

y que su media sea finita.

En suma, que si la variable aleatoria es normal con $N(\mu, \sigma)$, la ley de distribución de las medias de las infinitas muestras que pueden extraerse del universo n_1 exhibirá, también, distribución normal con $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. Equivalente a decir que si aceptamos que una población es normal definida por $N(\mu, \sigma)$, conocemos inmediatamente la ley de distribución de la población constituida por las muestras de n elementos tomados de $N(\mu, \sigma)$.

Así, estamos capacitados, partiendo de mediciones realizadas sobre una muestra concreta, para afirmar que ésta, con un grado de certidumbre establecido «a priori», puede ser considerada como perteneciente a la población $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ y, recíprocamente.

Realizadas en distintas modalidades las operaciones que anteceden constituyen, a nuestro juicio, el control de calidad.

La certidumbre aludida, diferente según la precisión que se desee o que se pueda conseguir en el problema planteado, se expresa mediante los conceptos de nivel de confianza y de su complemento a 100, llamado nivel de significación.

Refiriéndonos, por ejemplo, a la distribución muestral de las medias, si ésta es normal y la muestra tiene $n > 30$ elementos, utilizando las tablas de probabilidad para la función normal tipificada, y fijamos simétricamente a μ dos valores

$$\mu \pm K\sigma_{\bar{x}} \quad ; \quad \sigma_{\bar{x}} = \frac{\sigma \text{ (de la población)}}{\sqrt{n}}$$

tales que el intervalo que definen (intervalo de aceptación) ofrezca probabilidad de 99 por 100, podremos decir que la de obtener una muestra con una media que caiga fuera de aquel intervalo es de 1 por 100.

De modo general al intervalo $\mu + K\sigma$, $\mu - K\sigma$ se denomina *intervalo de confianza.*

A los valores concretos $(\mu + K\sigma)$ y $(\mu - K\sigma)$ se les designa como *límites de confianza;* los valores de K se denominan coeficientes de confianza. Damos a continuación (cuadro de valores) los niveles de confianza y los coeficientes de confianza definidos por

$$C = \mu \pm K\sigma$$

frecuentemente empleados.

Nivel de confianza.	99,73 %	99 %	98 %	96 %	95,45 %	95 %	90 %	80 %	68,27 %	50 %
K	3,00	2,58	2,33	2,05	2,00	1,96	1,645	1,28	1,00	0,6745

Designando por α el *nivel de confianza* $(100 - \alpha)$ es por definición el *nivel de significación,* y mide la probabilidad de que la media de la muestra concreta considerada caiga en la llamada región crítica, o sea, fuera del intervalo de confianza.

Es corriente tomar el 5 por 100 como nivel *casi significativo,* 1 por 100 como *nivel significativo* y 0,1 por 100 como *muy significativo.*

Inversamente, si a partir de la información ofrecida por la distribución muestral queremos comprobar si la hipótesis aceptada $N(\mu, \sigma)$ es correcta, podremos, utilizando los datos anteriores, esperar que μ debe de encontrarse dentro del intervalo,

$$\bar{x} \pm \frac{\sigma}{\sqrt{n}}$$

simétrico alrededor del valor $\bar{x}$ de una muestra concreta.

Intervalos de confianza para la desviación típica

Los límites de confianza para σ, en una población normal, estimada a partir de una muestra con desviación típica S, están dados por:

$$S \pm K \frac{\sigma}{\sqrt{2n}}$$

Error probable

Los límites para el nivel de confianza al 50 por 100 están dados, para un estadístico e, por:

$$e \pm 0{,}6745\sigma_e$$

Errores de muestreo y sesgo

Existen errores característicos de muestreo, y otros que pueden presentarse en toda investigación estadística y que reciben el nombre de sesgo. Como ejemplo, puede darse el fenómeno que representa la distribución de disparos sobre un blanco cuando los impactos están distribuidos alrededor de un punto que no coincide con la diana. La distancia entre ambos mide el sesgo, y la amplitud de la agrupación alrededor del punto central mide el error de muestreo.

Distribución de t y de χ^2 (chi cuadrado)

Cuando el número de elementos que constituyen la muestra es superior a 30, las distribuciones muestrales de sus estadísticos son aproximadamente normales y, tanto más próximas a tal distribución, cuanto mayor es el número de elementos N de la muestra.

Para $N < 30$, la aproximación ya no es buena y empeora al decrecer n. Por ello, ha sido elaborada la llamada teoría de las muestras pequeñas, a las que se refieren las distribuciones que vamos a considerar.

Distribución t de Student

Si se define un estadístico t por:

$$t = \frac{\bar{x} - \mu}{S} \sqrt{n - 1}$$

$\bar{x}$ = media de la muestra;
μ = media de la población;
S = desviación típica de la muestra;
n = número de elementos de la muestra.

Considerando las infinitas muestras posibles extraídas de un universo normal y si para cada muestra calculáramos t, utilizando la media de las desviaciones típicas de cada muestra, se obtendría la distribución muestral de t según:

$$Y = \frac{Y_o}{\left(1 + \frac{t^2}{n - 1}\right)^{\frac{n}{2}}} = \frac{Y_o}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu}{2}}}$$

en la que Y_o es una constante que depende de n y que responde a la condición de que el área bajo la curva sea la unidad, y $\nu = n - 1$, llamado número de grados de libertad, representa el número de observaciones independientes ofrecidas por la muestra. (Si, por ejemplo, la media del universo μ se sustituye por el $\bar{x}$ de la muestra, las n observaciones no ofrecen ya más que $n - 1$ observaciones independientes y, por tanto, $\nu = n - 1$).

Para valores grandes de n ($n > 30$) o de ν, las curvas se aproximan a la curva normal tipificada de ecuación:

$$Y = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} t^2}$$

Reproducimos en la figura 7 la forma de las curvas.

En los tratados de estadística están tabulados los valores de la probabilidad según la ley de distribución de densidad de probabilidad para la t de Student y, por tanto, con ella se puede determinar la probabilidad de un suceso, límites de confianza, etc., a partir de los estadísticos de muestras pequeñas. La ley de Student puede hallar aplicaciones útiles en el control de calidad, especialmente como «test» de hipótesis.

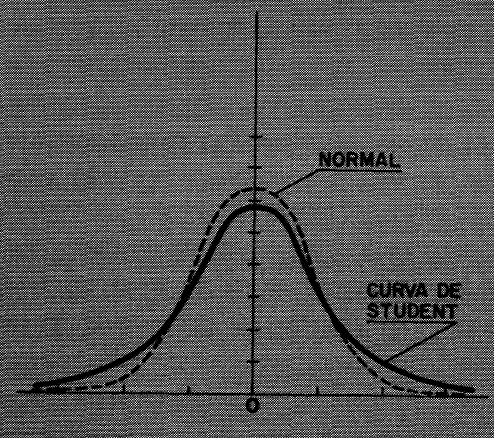


Fig. 7.

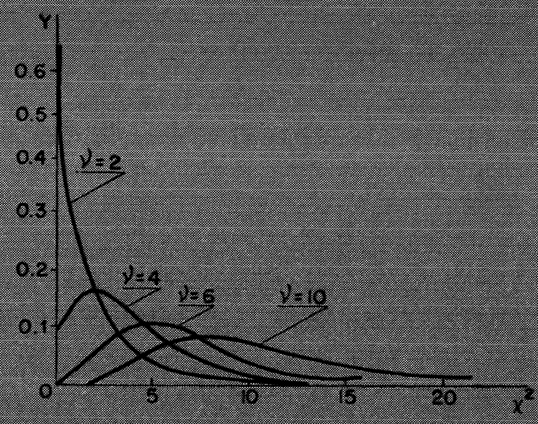


Fig. 8.

Ley de distribución de Pearson o del χ^2 (chi cuadrado)

Definiendo el estadístico:

$$\chi^2 = \frac{n\,S^2}{\sigma^2} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{\sigma^2}$$

$\sigma =$ desviación típica de la población normal, de donde se supone tomada la muestra;

$S^2 =$ varianza de la muestra (de n elementos).

Considerando las muestras de tamaño n extraídas de ese universo, calculando la función χ^2 para cada una de ellas, se obtendría una distribución de densidad de probabilidad para χ^2, cuya ecuación es:

$$Y = Y_o (\chi^2)^{1/2\,\nu-2}\, e^{-\frac{1}{2}\chi^2} = Y_o \chi^{\nu-2}\, e^{-\frac{1}{2}\chi^2}$$

en la que $\nu = n - 1$ es el número de grados de libertad, e Y_o es una constante que depende de ν de tal modo que satisfaga a la condición de que el área bajo la curva sea, como en toda ley de distribución de densidad, igual a la unidad.

La figura 8 reproduce la representación gráfica de la χ^2 para distintos valores de ν.

Los valores de Y para intervalos de χ^2 han sido calculados y, por tanto, es fácil la aplicación de esta ley de distribución en el control de hipótesis estadísticas, que pueden originarse en los problemas planteados por la realización del control de calidad.

Como complemento final de la exposición de las distribuciones, y por su importancia en la Teoría de muestras, cuyos resultados hemos pretendido esbozar en lo que antecede, citaremos el hallazgo de la Matemática estadística que, basado en el teorema del cálculo de probabilidades, llamado teorema central del límite, permite afirmar:

Si la variable aleatoria x tiene una función de distribución densidad cualquiera, con media μ y desviación típica σ finitas, la función de distribución de densidad de probabilidad de la nueva variable

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

es la normal (0, 1) cuando $n \to \infty$.

Control de calidad

Ya al comienzo nos hemos referido a lo que se entiende por control de calidad. En forma distinta se dice que «un fenómeno está bajo control, para nuestros fines, cuando, basándose en experiencias del pasado, se puede predecir dentro de qué límites se espera que varíe dicho fenómeno en el futuro».

Predecir «dentro de límites» significa que, al menos aproximadamente, podemos afirmar la posibilidad de que un fenómeno (por ejemplo, la producción en serie de determinados objetos) permanecerá dentro de las tolerancias fijadas.

La posibilidad de predicción es cualitativamente distinta si se comparan, por ejemplo, la de la aparición de un eclipse con la de estimar la esperanza de sobrevivir para una persona en un instante dado o la de la velocidad de una molécula.

La predicción que interesa al control de calidad quiere apoyarse en bases científicas y se refiere, como ya hemos indicado, a las posibles variaciones engendradas por causas de azar, entendiendo por tales a toda causa desconocida que pueda actuar sobre un fenómeno.

El control de calidad se apoya, teóricamente, en los dos postulados enunciados por Shewhart, como consecuencia de sus estudios:

Primer postulado: En sentido de predecir el futuro en términos del pasado, no son equivalentes todos los sistemas de causas de azar. Por ello, si por ejemplo queremos determinar la calidad de un producto, definiéndola mediante el intervalo de valores fijado a alguna propiedad, es preciso establecer un criterio que permita asegurar cuando el sistema de causas aleatorias permite hacer la predicción.

Tales criterios están expresados en formas de funciones de distribución de probabilidad, a las que ya nos hemos referido.

Segundo postulado: «En la naturaleza existen sistemas constantes de causas de azar».

Dos cuestiones suscita el postulado: Comprobar la existencia de sistemas constantes de azar y, a los efectos del control de calidad, determinar su existencia en un proceso de fabricación.

La experiencia prueba que, en numerosas aplicaciones del *control de calidad,* se registran causas de variabilidad del producto, desconocidas y, sin embargo, estas causas no pertenecen a un sistema de azar constante. El fin práctico y útil, entre otros, del control de calidad, es localizar esas causas asignables imputables o accidentales y eliminarlas.

Alguien ha afirmado, además:

«La calidad no se inspecciona, sino que se crea por medio de la inspección».

El control de calidad «admite la variabilidad, pero dentro de ciertos límites».

Que son repetición de afirmaciones deducibles, con carácter general, de la consideración de la posibilidad de someter a control un proceso.

El control de calidad puede practicarse según dos modalidades.

Control de calidad por variables

La calidad de un producto se evalúa practicando mediciones de ciertas características exigidas por las reglas de calidad, sobre una muestra, que permiten determinar los estadísticos del universo de la muestra y, a partir de él, señalar unos límites de confianza, por ejemplo, a la media. Las mediciones que se realicen sobre ejemplares que caigan fuera de esos límites señalados al universo, se declaran defectuosas. La base matemática para este control es la distribución normal.

En el control por atributos las unidades sometidas a control son examinadas para comprobar si satisfacen o no las condiciones prefijadas. Suministra información, por tanto, del total de defectuosos a partir de los ofrecidos por una muestra.

Se basa en la distribución binomial.

El control de calidad utiliza un medio peculiar de exposición de datos llamados «gráficos de control» que, en definitiva, son la representación gráfica de los resultados de mediciones realizadas sobre muestras de igual tamaño tomadas a lo largo del tiempo.

Se parte de aceptar que como resultado de la medición, muy reiterada, de ciertas propiedades, la producción puede ser representada por un universo normal $N(0, 1)$, y se pretende conseguir que la producción futura se ajuste a ese universo.

El curso de la comparación entre lo pretendido y lo obtenido en la producción a lo largo del tiempo se realiza mediante el *gráfico de control*, para cuya confección se señalan a lo largo de un eje (eje de abscisas), puntos dispuestos a intervalos determinados que pueden ser, por ejemplo, proporcionales a los intervalos de tiempo de toma de muestra. O bien sobre tal eje, a intervalos constantes se numeran puntos que siguen el mismo orden que el de la toma de muestras. Por cada punto representativo se levanta una perpendicular cuya longitud es proporcional a la medición realizada sobre la muestra. Se establece, además, la recta paralela al eje x, que representa a μ y las rectas que delimitan el intervalo de aceptación o de confianza, dadas, por ejemplo:

$$y = \mu \pm 1{,}96\, \sigma$$

que en este caso indican que sólo el 5 por 100 de los puntos puede estar fuera de la banda limitada por ellas.

Si las rectas de control estuvieran definidas por

$$y = \mu \pm 3{,}09\, \sigma$$

sólo el 0,2 por 100 de los puntos puede estar fuera de esta banda.

La figura 9 es un ejemplo de gráfico de control. Los datos originales corresponden a la producción de una fábrica de cemento española.

Las operaciones necesarias para establecer el gráfico de control son facilitadas mediante tablas que, para cada tamaño de muestra, dan la correlación entre los estadísticos correspondientes a la muestra y los de la población madre de aquéllos.

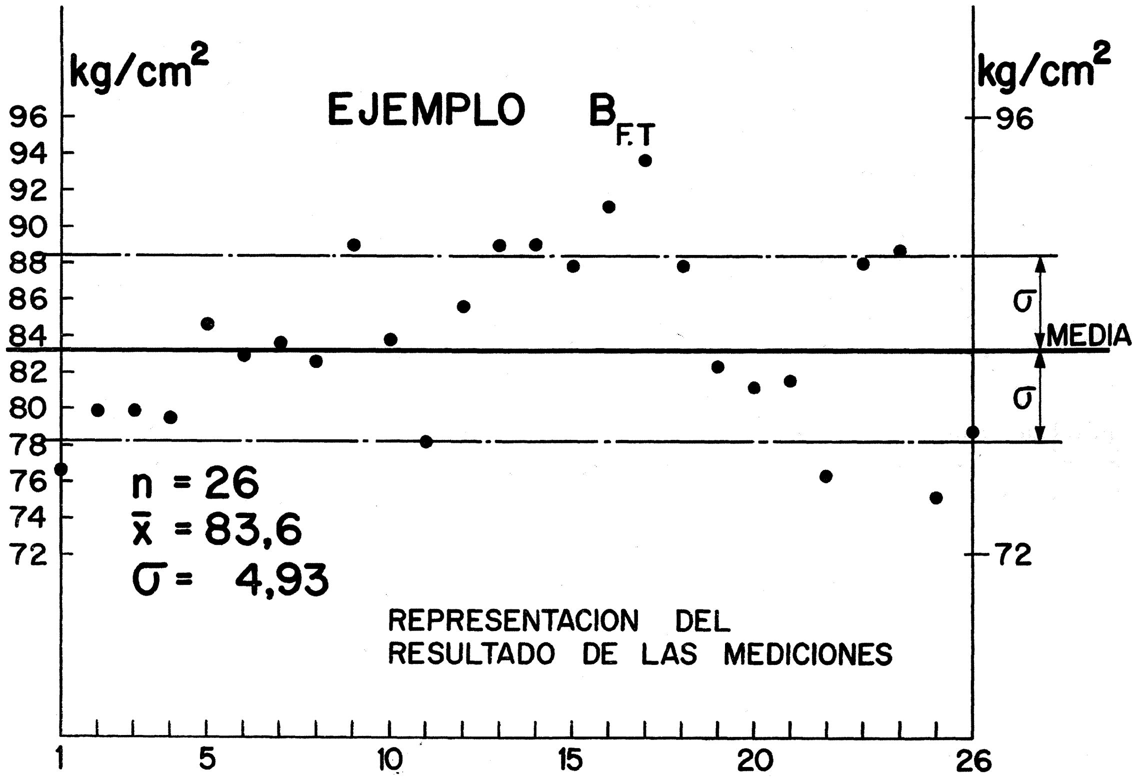
Por su gran comodidad, suele emplearse como estadístico para la medida de la dispersión

el recorrido R,

en lugar de la desviación típica σ, de elaboración de cálculo más difícil. Las relaciones numéricas entre ambos estadísticos están tabuladas también gracias a los trabajos experimentales de Shewhart para la distribución normal y a los de Galton para la binomial.

Como ejemplo de aplicación vamos a interpretar el significado de las condiciones exigidas por la Dirección General de Industrias de la Construcción para obtener la «marca de Calidad».

En el control de *la media* se exige que la correspondiente a la población aleatoria representada por los resultados de la medición de una propiedad sobre la producción de una fábrica dada (pues lo más probable será que para cada fábrica sea diferente de las demás) alcance tal valor que su límite crítico inferior (cuando la estipulación obliga a obtener valores superiores a uno dado, por ejemplo, en la resistencia mecánica), frontera entre región de confianza y región crítica, ha de quedar a la distancia de de 1,45 R del límite de especificación.


kg/cm2
EJEMPLO $B_{F.T}$
kg/cm2
96
94
92
90
88
86
84
82
80
78
76
74
72
96
72
σ
MEDIA
σ
n = 26
x̄ = 83,6
σ = 4,93
REPRESENTACION DEL
RESULTADO DE LAS MEDICIONES
1
5
10
15
20
26

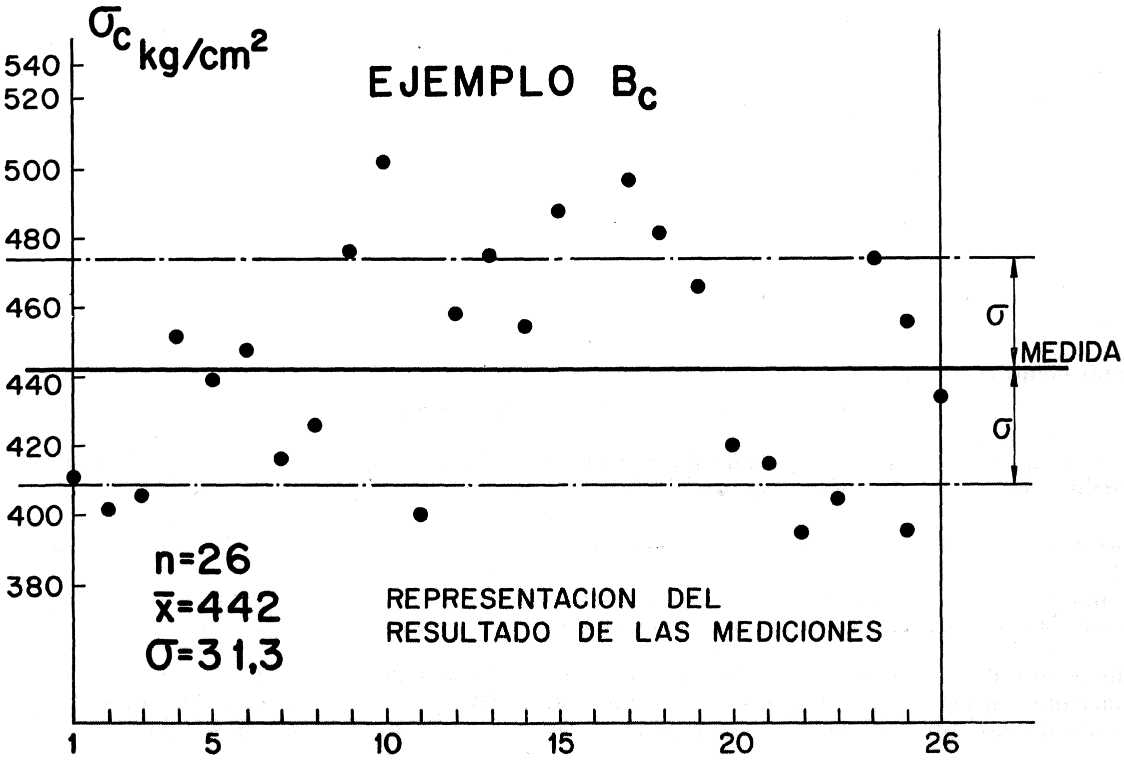

σc kg/cm2
EJEMPLO B_C
540
520
500
480
460
440
420
400
380
σ
MEDIDA
σ
n=26
x̄=442
σ=31,3
REPRESENTACION DEL
RESULTADO DE LAS MEDICIONES
1
5
10
15
20
26

Fig. 9.

En las aplicaciones del control de calidad es aceptado de modo general que los límites críticos estén definidos por $\mu \pm 3\,\sigma$ (99,73 % conf.).

Podemos escribir por lo tanto,

$$\text{Límite de especificación} + 1{,}45\,R = \overline{X} - 3{,}00\,\sigma.$$

y como para $n = 2$, $R = 1{,}128\,\sigma$ (ver tablas para aplicación del control de calidad) se obtiene:

$$\text{Límite de especificación} + 1{,}45 \times 1{,}128\,\sigma = \overline{X} + 3{,}00\,\sigma$$

$$\text{o} \quad \overline{X} = \text{Lím. de esp.} + 1{,}45 \times 1{,}128\,\sigma + 3{,}00\,\sigma = \text{Lím. de esp.} + 4{,}635\,\sigma$$

equivalente a decir que para conseguir la «marca de calidad» el valor central de la población aleatoria representada por la producción, si el recorrido promedio obtenido en el ensayo exigido hubiera sido 15, y si la marca de calidad se refiriera a un P-350, la media $\overline{X}$ para Comp_{28} deberá alcanzar el valor:

$$\overline{X} = 350 + 4{,}635 \;\frac{15}{1{,}128} = \sim 410.$$

Si el promedio de R hubiera sido 7,5, la media deberá alcanzar el valor

$$\overline{X} \sim 350 + 30 = 380$$

solamente.

En ambos casos queda estadísticamente asegurado que la probabilidad de obtener muestras cuya media llegue a ser igual al valor de especificación, vale:

$$0{,}138\,\%$$

[Ver tablas de probabilidad de $N(0.\ 1)$].

La faja comprendida entre el límite crítico y el límite de especificación (el no man's land) como humorísticamente oí denominarla al distinguido colega señor Peña, del Instituto de Racionalización del Trabajo, representa un intervalo de probabilidad igual a la diferencia de los que corresponden a $P \pm 3{,}00\,\sigma$ y $P \pm 4{,}635\,\sigma$ (0,27 por 100 y 0,138 por 100, respectivamente).

El gráfico de regularidad de la calidad se apoya en consideraciones semejantes.

Las relaciones numéricas para el límite de control son

$$L \cdot C = \overline{R} + 3\,\frac{d^3}{d^2}\,\overline{R} = \overline{R}\left(1 + 3\,\frac{0{,}853}{1{,}123}\right) = (1 + 2{,}267)R = 3{,}267R$$

equivalente a

$$L \cdot C = 3{,}267\overline{R} \times 1{,}128 = 3{,}685\,\sigma$$

d^3 y d^2 están tabulados. Véase Tabla B.2 y Tabla II en las páginas 115 de A.S.T.M. «Manual on Quality control of Materials», January 1951.

Las figuras 10 y 11 traducen gráficamente la interpretación que antecede.

Como final de esta larga exposición quisiera expresar mi convicción de que el control de calidad puede ofrecer una posibilidad de aplicación que para mí es absolutamente nueva.

Me refiero al control estadístico del proceso fabril. Actualmente, a lo largo del mismo existen suficientes elementos de información que suministran datos (por ejemplo, en el caso de la fabricación de cemento), relativos a temperaturas, presiones, caudales de gases, de sólidos, etc., que,

EJEMPLO DE CONDICIONES PARA LA MARCA DE CALIDAD

(REGULARIDAD)

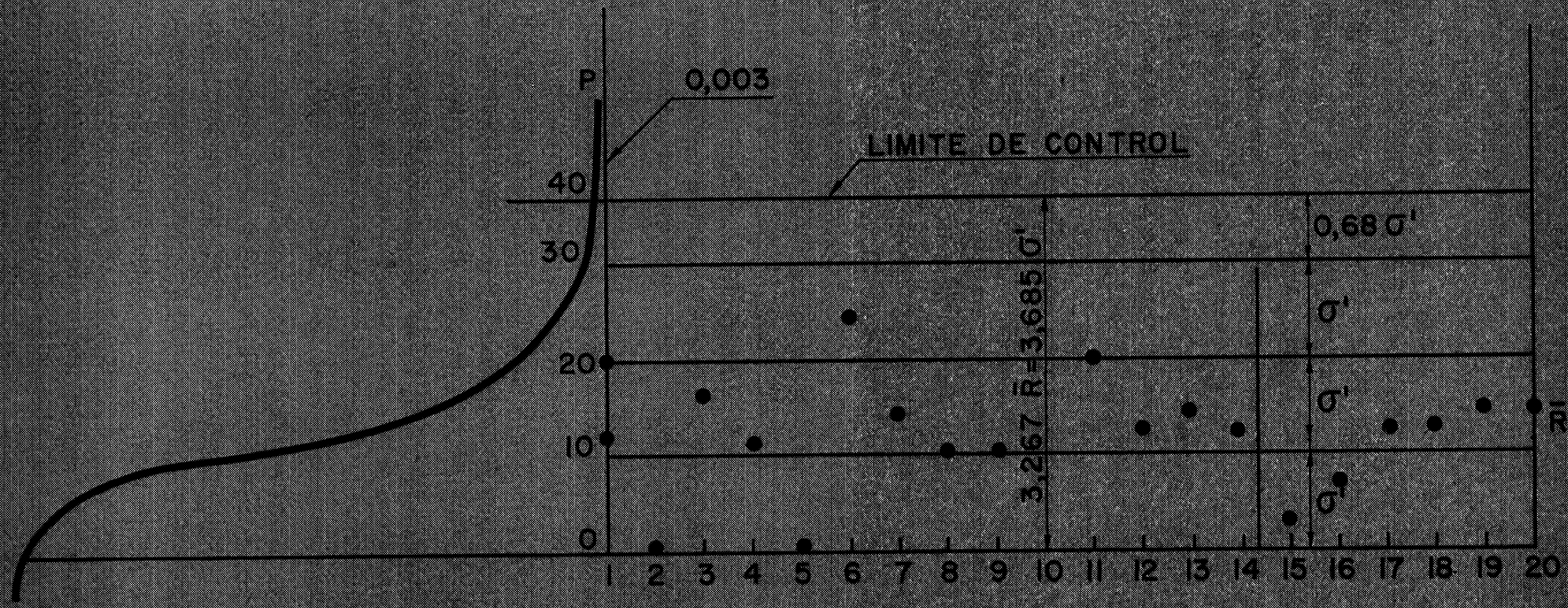


Fig. 18.

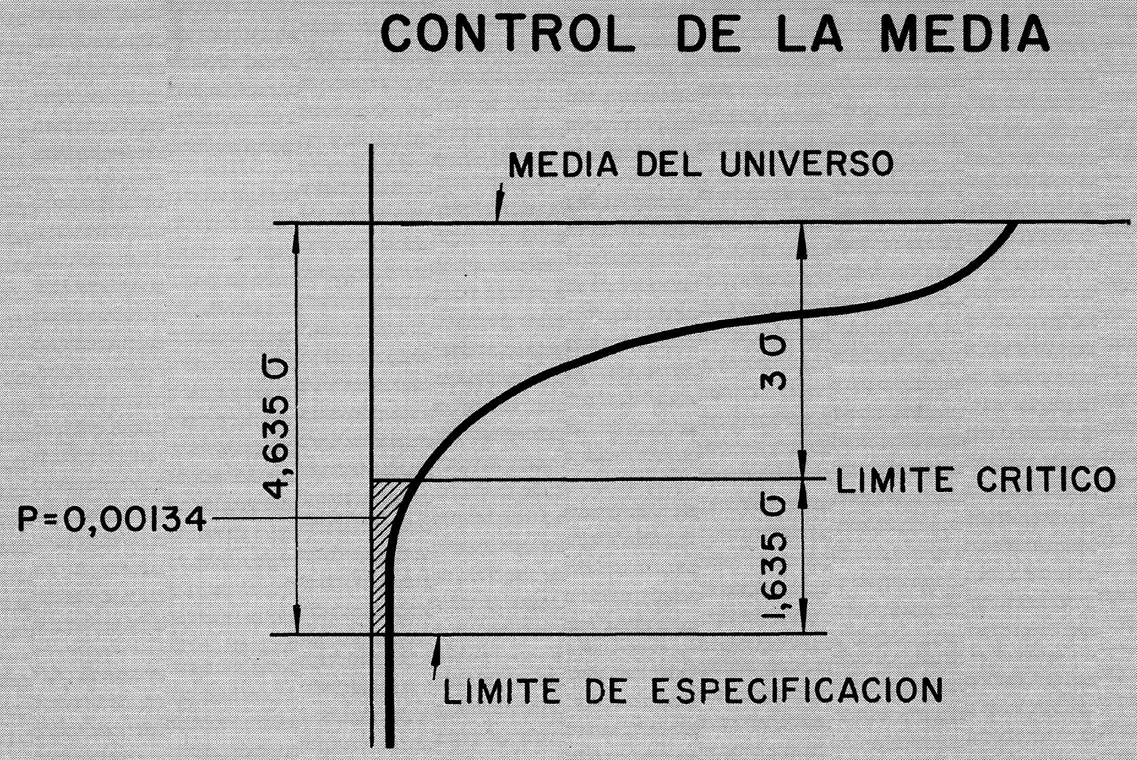


Fig. 11.

unidos a determinaciones analíticas complementarias deben de permitir el establecimiento, por así decir, de MODELO de realización del proceso en una fábrica dada que dará lugar a la consideración de poblaciones aleatorias caracterizadas por las respectivas (μ, σ). A lo largo del tiempo la observación diaria de la información aportada, por los elementos de control aludidos nos informará si es correcta la marcha del proceso el día considerado o si sistemáticamente aparecen desviaciones indicando la existencia de causas asignables, etc.; o sea, que, en suma, la Matemática estadística ofrece, con el estudio de los tipos de distribución, probabilidades y demás consideraciones que he intentado esbozar ante vosotros, un gran instrumento de trabajo para el análisis de la marcha del proceso productor mismo por el Técnico o Director de fábricas de cemento.

Y quiero terminar dándoos las más sinceras gracias por vuestra amable atención a mi pobre exposición y, además, porque estos IV Coloquios me han obligado a estudiar al nivel de como lo hacía cuando iba a examinarme. Algo que ocurrió hace tal número de años que cae ya en la zona de los grandes números. Así, que a vosotros os debo haberme sentido joven... aunque en apuros.

Bibliografía

Das Mathematische Werkzenq.—Rudolf Fueter.—Orell Füssli.—Zurich.
Theory and Problems of Statistics.—Murray R. Spiegel.—Schamm Pub. N. York.
A. S. T. M. Manual on Quality Control of Materials.
Iniciación estadística.—Sixto Ríos.
Control estadístico de la calidad.—E. Blanco Loizeller.
Control estadístico de calidades.—Carlos Paz Shaw.—Instituto de Racionalización del Trabajo.