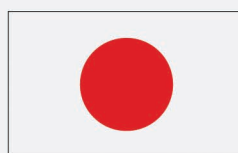


Métodos y herramientas para la **monitorización** de la **calidad** en **servicios de salud**

Pedro J. Saturno Hernández



Instituto Nacional
de Salud Pública



Métodos y herramientas
para la monitorización de la calidad
en servicios de salud

Métodos y herramientas para la **monitorización** de la **calidad** en **servicios de salud**

Pedro J. Saturno Hernández



Métodos y herramientas para la monitorización de la calidad en servicios de salud

Primera edición, 2015

D.R. © Instituto Nacional de Salud Pública
Av. Universidad 655, Col. Santa María Ahuacatlán
62100 Cuernavaca, Morelos, México

ISBN 978-607-511-148-3

Impreso y hecho en México
Printed and made in Mexico

Citación sugerida:

Saturno-Hernández PJ. Métodos y herramientas para la monitorización de la calidad en servicios de salud. Cuernavaca, México: Instituto Nacional de Salud Pública, 2015.

Acerca del autor

Pedro J. Saturno Hernández es médico, doctor en Salud Pública por la Universidad de Harvard, profesor honorífico de Salud Pública de la Facultad de Medicina de la Universidad de Murcia (España) y científico visitante y miembro del Grupo de Salud Global de la Escuela de Salud Pública de la Universidad de Harvard. Dirige una Maestría en línea sobre Gestión de la Calidad en Servicios de Salud, que actualmente se imparte en su 15° edición en la Universidad de Murcia, en su segunda en el Instituto Nacional de Salud Pública (INSP), en México y en su primera en la Universidad de Rio Grande del Norte, en Brasil.

Previamente se ha desempeñado como director general adjunto de Planificación Sanitaria y Educación en el Ministerio de Salud de España. Tiene una extensa experiencia como consultor sobre gestión de la calidad en servicios de salud para agencias internacionales como la Organización Mundial de la Salud (OMS) y la Agencia de Estados Unidos para el Desarrollo Internacional (USAID), entre otras. Ha participado en varios grupos de trabajo y comités nacionales e internacionales en relación con la gestión de la calidad, prevención en salud y educación a lo largo de los últimos 30 años, ha impartido cursos especializados y conferencias en numerosos ámbitos y ha ayudado a la implantación de programas de mejora de la calidad en varios países e instituciones.

Actualmente es AXA Professor on Healthcare Quality y director adjunto del Centro de Investigación en Calidad y Encuestas del INSP.

Contenido

| | |
|---|------------|
| Presentación | 11 |
| Capítulo 1 | 13 |
| Actividades para la monitorización. Construcción, análisis y tipos de indicadores. | |
| Planes de monitorización | 15 |
| Capítulo 2 | 31 |
| Indicadores compuestos | 33 |
| Capítulo 3 | 45 |
| El muestreo para la aceptación de lotes (LQAS) como método de monitorización | 47 |
| Capítulo 4 | 63 |
| Control estadístico de la calidad. Gráficos de desarrollo | 65 |
| Capítulo 5 | 85 |
| Gráficos de control | 87 |
| Capítulo 6 | 101 |
| Gráficos de capacidad de proceso. Sistema Seis Sigma para la mejora continua | 103 |
| Capítulo 7 | 123 |
| Gráficos CUSUM, gráficos EWMA y otros métodos de análisis y monitorización estadística de la calidad | 125 |

Presentación

Hay consenso en aceptar que la calidad no mejora sola. Para conseguir que cambien la forma de hacer las cosas y los resultados que con ella se obtienen, hace falta atención específica en la mejora, y esto implica acercamientos y metodologías diversas.¹ Parafraseando una cita atribuida a Einstein, es de estúpidos pensar que las cosas van a mejorar si seguimos haciéndolas igual. Sin embargo, no todas las actividades que se emprendan para gestionar mejor la calidad son igualmente efectivas. Basarse en la evidencia existente es un imperativo ineludible, pero también lo es utilizar una metodología rigurosa y contrastada para su implementación.² Si no se hace así, las actividades que se emprendan van a contribuir, por defectuosas, al enorme desperdicio que existe en el funcionamiento y provisión de los servicios de salud.³

Este manual, junto con los otros dedicados a la metodología de acuerdo con los tres grandes grupos de actividades que intervienen en la mejora continua de la calidad (ciclos de mejora, monitorización y planificación de la calidad), quiere contribuir a explicitar cómo implementar estas actividades de la forma más práctica posible pero con el rigor científico necesario.

Monitorizar la calidad se presenta como una actividad de utilidad múltiple, variable según las diversas circunstancias y niveles de responsabilidad dentro de los sistemas de salud, pero que en todos los casos presenta el común denominador de servir para controlar los niveles de calidad existentes mediante la detección de situaciones tanto aceptables como problemáticas. Su realización, desde el punto de vista metodológico, básicamente implica el uso de buenos indicadores (válidos, fiables y útiles) en planes de monitorización bien definidos, con métodos apropiados de medición y análisis de la información propios para los fines que se persigan.

En este manual se abordan todos estos componentes haciendo particular énfasis en aspectos comúnmente menos contemplados en otros esquemas de monitorización en servicios de salud, pero de gran relevancia y trascendencia práctica. Como ejemplo, dentro de los tipos de indicadores, se enfatiza la clasificación según el tipo de variable estadística que resume la información, por su trascendencia en el tipo de análisis que corresponde, así como la tipología y estructuración según niveles de responsabilidad (micro, meso y macro) dentro de los sistemas de salud, y una atención particular a los indicadores compuestos, a los que se dedica un capítulo completo. De igual manera, en lo relativo a métodos de medición y análisis, se revisan con especial detalle los acercamientos utilizados en el ámbito industrial

1 Saturno-Hernández PJ, Hernández-Ávila M, Magaña-Valladares L, García-Saisó S, Vértiz-Ramírez J J. Estrategia integral de formación para la mejora continua de la calidad de los servicios de salud. *Salud Pública Mex* 2015;57:275-283.

2 Saturno PJ. Tendencias actuales en los sistemas de gestión de la calidad. En: Secretaría de Salud. *La calidad de la salud en México a través de sus instituciones*. México: SSA 2012;225-238.

3 Organización Mundial de la Salud. *Informe de la salud en el mundo. La financiación de los sistemas de salud: hacia la cobertura universal*. Ginebra: OMS, 2010.

(LQAS, control estadístico, capacidad de proceso, Seis Sigma), cuya aplicación en los servicios de salud está siendo incompresiblemente lenta, a pesar de su innegable utilidad y eficiencia.

En la mayoría de los capítulos, hay una invitación a completar la lectura con una aplicación práctica para evidenciar posibles limitaciones y utilidad en contextos concretos. No hay que olvidar, como rasgo inherente a todas las actividades de monitorización, que ésta por sí sola no va a mejorar la calidad: se necesita de una voluntad explícita y operativa para poner en marcha los ciclos o actividades de mejora sobre los problemas o situaciones que los resultados de la monitorización hayan evidenciado. En este sentido, el presente manual se complementa con los relativos a las otras actividades para la mejora continua con el fin de configurar el necesario acercamiento integral a los sistemas de calidad en los servicios de salud. El enfoque y los contenidos de todos ellos están fielmente basados en los de la Maestría en Gestión de la Calidad en Servicios de Salud, que actualmente se imparte en línea en España, México y Brasil.

Capítulo 1

Actividades para la monitorización. Construcción, análisis y tipos de indicadores. Planes de monitorización

Contenido general y objetivos específicos

1. Evaluar, diseñar y monitorizar la calidad: los tres grupos de actividades que llevan a la práctica el compromiso con la calidad
2. Medición de la calidad. Situaciones y objetivos
3. Monitorización de la calidad: ¿qué incluye?
4. Puntos de partida para monitorizar: origen de los indicadores
5. Construcción y análisis de indicadores. Los puntos clave
6. Tipos de indicadores
7. Componentes de un plan de monitorización. Tipos de plan
8. Métodos de monitorización según las características del plan

Anexo 1.1. Análisis crítico de un indicador

Contenido general

En una visión integral de actividades para la mejora continua de la calidad, consideramos tres grandes grupos: ciclos de mejora, monitorización y diseño de la calidad. En otro manual se presentan los métodos y herramientas para el más conocido de ellos: los ciclos de mejora; en éste, todo lo relativo a la monitorización. Este primer capítulo revisa los principales conceptos y esquemas metodológicos para monitorizar, y los siguientes, las herramientas y técnicas que los llevan a cabo.

Objetivos específicos

1. Distinguir las principales actividades incluidas en la monitorización o control de los niveles de calidad
2. Explicar la relación entre monitorización y ciclos de mejora
3. Delimitar las diferencias prácticas entre criterio o requisito de calidad e indicador
4. Distinguir los diversos tipos de indicadores de utilidad en los servicios de salud
5. Analizar las características y proceso de construcción de un indicador
6. Describir las características de un indicador compuesto
7. Definir los principales componentes de que consta un plan de monitorización
8. Describir los objetivos generales de los planes de monitorización
9. Clasificar los planes de monitorización en función de la periodicidad de las mediciones
10. Enumerar diversas metodologías aplicables en la monitorización de indicadores

Actividades para la monitorización. Construcción, análisis y tipos de indicadores. Planes de monitorización

1. Evaluar, diseñar y monitorizar la calidad: los tres grupos de actividades que llevan a la práctica el compromiso con la calidad

En una visión integral de las actividades para la mejora de la calidad, definimos tres grandes grupos: ciclos de mejora, monitorización y diseño o planificación de la calidad. Hacemos la distinción entre los tres grupos de actividades en función de los diferentes puntos de partida y objetivos inmediatos que persiguen, según se resume en el cuadro 1.1. Por otra parte, los tres tipos de actividades se relacionan entre sí y, si bien los ciclos de mejora tienen sentido (principio y final)

por sí mismos (e incluyen actividades de diseño, al definir la intervención para mejorar), tanto la monitorización como el diseño necesitan complementarse, por lo menos, con otra de las actividades (la monitorización con ciclos de mejora, y el diseño con la monitorización). Esta interrelación, se resume en la figura 1.1.

Los tres grupos de actividades son diferentes puertas de entrada a la dinámica de la mejora de la calidad. La actividad central son probablemente los ciclos de mejora, pero es conveniente aclarar un poco más en detalle lo que implica empezar las actividades del programa de mejorar con la monitorización. Comencemos por resumir y aclarar la relación entre ciclos de mejora y moni-

La monitorización es una de las actividades necesarias para llevar a la práctica el compromiso con la mejora de la calidad.

Cuadro 1.1

Grupos de actividades para la mejora continua de la calidad.
Puntos de partida normales y objetivos inmediatos

| Grupo de actividades | Punto de partida | Objetivo inmediato |
|-------------------------|---|--|
| Ciclos de mejora | — Identificación de un problema de calidad u oportunidad de mejora en algún aspecto de los servicios que se ofrecen. | — Solucionar el problema. — Aprovechar la oportunidad de mejora descubierta. |
| Monitorización | — Identificación de aspectos relevantes de los servicios que se ofrecen y construcción de indicadores sobre su calidad. — Selección de indicadores sobre problemas sometidos a ciclos de mejora. — Selección de indicadores sobre procesos o servicios diseñados. | — Identificar problemas de calidad u oportunidades de mejora. |
| Diseño | — Programación de un nuevo servicio a ofrecer. — Identificación de necesidades y expectativas de los usuarios. — Identificación de parámetros y resultados a conseguir | — Diseñar los procesos de atención para conseguir los resultados deseados predeterminados. |

En este capítulo se da una visión panorámica de los métodos y actividades de monitorización de la calidad.

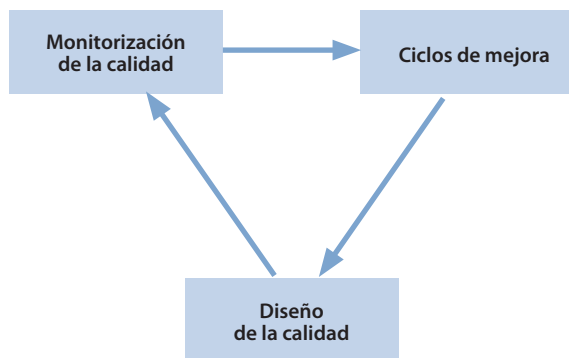


Figura 1.1. Los grupos de actividades para la mejora continua de la calidad

zación, por medio, sobre todo, de la comprensión de las situaciones y objetivos planteados cuando medimos la calidad.

2. Medición de la calidad. Situaciones y objetivos

En los programas de gestión de la calidad, ésta puede medirse en tres situaciones principales, descritas esquemáticamente en el cuadro 1.2, que pueden resumirse en dos objetivos: (i) estimación del nivel de calidad y (ii) comparación con niveles preestablecidos. La distinción entre las diversas situaciones y objetivos es importante porque pueden requerir acercamientos estadísticos diferentes y éstos a su vez pueden ayudar a determinar el grado de utilidad de ciertos métodos.

1. *Estimación del nivel de calidad.* Medición del cumplimiento de requisitos, criterios o indicadores de calidad del servicio o aspecto del servicio que queremos evaluar. Esta estimación puede hacerse con diferentes propósitos; por ejemplo, si hay o no espacio para la mejora; cuál es la situación de partida al inicio de un ciclo de mejora; comparar diversos centros o servicios; etc. En cualquier caso, lo que queremos saber, con una determinada precisión, es en qué nivel concreto está el servicio o el aspecto del servicio cuya calidad se quiere medir.
2. *Comparación de la situación actual con un nivel preestablecido.* Puede hacerse en dos tipos de situaciones: (i) documentar la mejora conseguida tras la implementación de la intervención en un ciclo de mejora, en cuyo caso comparamos con la medición previa, de base, o (ii) asegurarnos que estamos a un nivel deseable predeterminado (estándar), con el cual nos comparamos.

Para tener una estimación del nivel de presencia de los criterios o requisitos (objetivo 1), se utiliza normalmente una muestra, que obtenemos de forma aleatoria con el método más apropiado al caso (aleatoria simple, sistemática, estratificada, etcetera).

Cuando queremos comparar nuestra medición con un nivel preestablecido (objetivo 2), las dos situaciones mencionadas (documentar mejora o controlar que estamos en los niveles predeterminados) pueden utilizar métodos diferentes,

| Cuadro 1.2 Medición de la calidad. Situaciones, objetivos y métodos | | |
|--|---|---|
| Situación | Objetivos | Métodos |
| Evaluación en ciclos de mejora | Estimación del nivel de cumplimiento de los criterios o requisitos de calidad | Estadística descriptiva (proporciones, porcentajes, medias y sus intervalos de confianza) |
| Reevaluación tras intervenir | Comprobar con la situación de partida (¿hemos mejorado?) | Comparación de las proporciones o medias de las dos evaluaciones para ver la significación estadística de la diferencia |
| Monitorizar (mediciones repetidas y planificadas) | Comprobar si estamos o no en los niveles de referencia previamente establecidos | - Estimaciones sucesivas del nivel de cumplimiento - Decisión con base probabilística (aceptación/rechazo) sin estimar el nivel real de cumplimiento |

aunque ambas tienen en común que pueden ser consideradas como una comprobación de hipótesis previas: ¿hemos mejorado? o ¿estamos en niveles determinados preestablecidos?

Documentar la mejora conseguida es en realidad comparar dos estimaciones, antes y después de la intervención para mejorar. En esencia, es comprobar la hipótesis nula de ausencia de mejora frente a la hipótesis alternativa de haber logrado mejorar; es decir, contestar a la pregunta ¿cuál es la probabilidad de encontrar, debido al azar, la diferencia resultante de las dos evaluaciones, si no hay diferencia real? o, lo que es lo mismo, ¿es la mejora estadísticamente significativa? Para ello, podemos utilizar las pruebas estadísticas usuales, viendo la significación de una cola. Ni en este caso ni en ningún otro para el que nos basemos en estimaciones del nivel de calidad, existen métodos especiales desarrollados en el campo del control de la calidad en la industria.

3. Monitorización de la calidad: ¿qué incluye?

Entendemos por monitorización *la medición sistemática y planificada de indicadores de calidad*; es una actividad conducente a controlar que estamos en los niveles preestablecidos y cuyo objetivo es identificar la existencia o inexistencia de situaciones problemáticas que hay que evaluar o sobre las que hay que intervenir. Los indicadores por monitorizar pueden derivarse de ciclos de mejora, de actividades de diseño de los servicios, o ser fruto de una selección de aspectos o servicios relevantes de nuestro centro, cuya calidad nos interesa controlar. En cualquier caso, deben ir acompañados de un

plan de monitorización que incluya periodicidad, mecanismos para la recolección de datos y métodos de análisis o interpretación de los mismos.

4. Puntos de partida para monitorizar: origen de los indicadores

A la monitorización puede llegarse esencialmente por medio de tres situaciones diferentes, resumidas en el cuadro 1.3: (i) tras la elaboración y adopción de un listado de indicadores que representen los aspectos o servicios más relevantes que ofrecemos; (ii) tras completar un ciclo de mejora y seleccionar indicadores sobre el aspecto o servicio mejorado, y (iii) tras implantar un nuevo diseño de la forma de ofrecer un servicio reformado o añadido a nuestra actividad.

La monitorización de un listado de indicadores sobre la actividad del centro o servicio es probablemente la situación más común como punto de partida para monitorizar, y puede servir de entrada a las actividades de gestión de la calidad, tal como contemplan y recomiendan organizaciones como la *Joint Commission* y diversos autores. Con algunas matizaciones, es también la base de programas externos como los conjuntos de indicadores establecidos por los ministerios o entidades supra centro (redes o sistemas), o la evaluación de los diversos ítems para los programas de acreditación y certificación de centros. Su objetivo es asegurarse de que estos indicadores están en niveles aceptables, con o sin comparaciones con los niveles de diversos centros y, a la vez, identificar aspectos-problema que deben evaluarse y mejorarse.

Una situación diferente es la medición de indicadores sobre un aspecto o servicio que ha sido

La monitorización es la medición sistemática, repetida y planificada de indicadores de calidad, con el fin de identificar situaciones problemáticas.

Monitorizar implica identificar los indicadores por medir, definir el plan de monitorización y establecer un método de análisis de los resultados de la medición.

Cuadro 1.3
Monitorización de la calidad. Situaciones y objetivos

| Situación | Objetivos |
|--|--|
| - Medición de listado de indicadores sobre aspectos o servicios relevantes | - Asegurarse de que están en niveles aceptables - Identificar aspectos-problema o priorizar oportunidades de mejora |
| - Medición de indicadores sobre un aspecto o servicio que ha sido sometido a ciclo de mejora | - Asegurarse de que se mantiene la mejora conseguida |
| - Medición de indicadores sobre un aspecto o servicio de nuevo diseño | - Asegurarse de que el diseño funciona como estaba previsto |

A la monitorización se puede llegar tras un ciclo de mejora, tras la elaboración de un listado de indicadores sobre la actividad del centro, o tras actividades de diseño de la calidad.

sometido a un ciclo de mejora. En este caso, si bien la actividad de base es la misma (construcción de indicadores, normalmente seleccionando criterios de los utilizados en la evaluación), el objetivo es asegurarse de que se mantiene la mejora conseguida, por lo menos durante un periodo que nos parezca prudencial en función del aspecto que estemos monitorizando. Sin embargo, el objetivo es, en último término, como en el caso de los listados de indicadores, identificar situaciones (en este caso por no haberse mantenido la mejora conseguida) en las que es necesario investigar (de nuevo) las causas de la calidad deficiente.

Una situación semejante, en cuanto que se parte de la implantación de una intervención diseñada para un determinado aspecto o servicio, es la elaboración y medición de indicadores sobre uno de nuevo diseño. En este caso, lo que nos interesa es fundamentalmente asegurarnos de que el diseño implementado funciona como estaba previsto; pero, en último término, al igual que en las otras situaciones, se monitoriza para identificar la existencia o inexistencia de problemas que merezcan ser investigados.

En las tres situaciones, por lo tanto, el objetivo último de la monitorización es la identificación de problemas, y los métodos de medición pueden ser semejantes en función de cómo establezcamos el plan de mediciones, con independencia de la situación que da origen a los indicadores y a la medición en sí.

En el cuadro 1.4 se resumen las diferencias entre medir para evaluar (para ciclos de mejora) y medir para monitorizar.

Como se ha dicho, sea cuál fuere el punto de partida, las principales actividades necesarias para la monitorización son la construcción o el análisis de los indicadores, el diseño del plan de

monitorización y el método de análisis e interpretación de los datos.

5. Construcción y análisis de indicadores. Los puntos clave

Un indicador es un aspecto relevante que resume, en la medida de lo posible, la calidad de la actividad o problema que se desea monitorizar. Su utilización, más que otra cosa, es una herramienta de *screening*, para asegurarnos de que estamos en los niveles de calidad deseados.

En cualquier caso, sea cual fuere el origen de la monitorización (cuadro 1.3) lo que hay que monitorizar son buenos indicadores, es decir, indicadores válidos, fiables y apropiados (cuadro 1.5).

5.1. Indicadores válidos

Como instrumento de *screening*, un buen indicador será aquel que reúna las características de validez, sensibilidad y especificidad que se resumen en el cuadro 1.5. Asimismo, un indicador es también un criterio de calidad, aunque sea un criterio especialmente relevante al tener vocación de resumen del nivel de calidad; por ejemplo, para el programa de hipertensión, podríamos decidir que sería suficiente la selección de tres indicadores para monitorizar su calidad: uno en relación con la captación; otro, con el diagnóstico correcto, y otro, con el resultado. El que los indicadores sean también criterios de calidad significa que todas las recomendaciones y características de los criterios de evaluación en los ciclos de mejora (realistas, aceptables, medibles, fiables; además de relevantes y válidos) son igualmente aplicables a los indicadores.

Un buen indicador ha de tener las mismas características exigibles a un buen criterio, más las que corresponden a su condición como instrumento de *screening* (sensibilidad y especificidad).

| Cuadro 1.4 | |
|---|---|
| Diferencias entre evaluación y monitorización de la calidad | |
| Evaluación | Monitorización |
| Comprensión amplia de proceso | Menos amplia/resumen |
| Varios aspectos/criterios | Pocos aspectos/un criterio |
| Ad hoc, ocasional | Continua/rutinaria |
| Conduce a intervención y monitorización | Conduce a evaluación (si no se alcanzan los niveles preestablecidos o aparece el criterio centinela) |

Cuadro 1.5

Características de un buen indicador

1. Validez

- 1.1. Validez como indicador de calidad
 - ¿Mide calidad y sirve para monitorizar e identificar situaciones en las que la calidad asistencial puede mejorarse?
- 1.2. Validez *facial*
 - ¿Se entiende su sentido e importancia sin muchas explicaciones?
- 1.3. Sensibilidad
 - ¿Identifica todos los casos en los que hay problemas de calidad?
- 1.4. Especificidad
 - ¿Identifica sólo los casos en los que hay un problema de calidad?

2. Fiabilidad

¿Es medido e interpretado siempre de la misma manera por todos los evaluadores?

3. Utilidad (para la mejora de la calidad)

¿Es apropiado para detectar situaciones a mejorar en el nivel de responsabilidad de quienes valoran sus resultados?

alidad, puede valorarse la sensibilidad y especificidad del indicador.

Para asegurarnos de que construimos o seleccionamos un indicador con las características de validez y fiabilidad, la *Joint Commission on Accreditation of Health Care Organizations* (JCAHO) americana propuso, hace años, un esquema aplicable a los indicadores clínicos que se resume en siete puntos, mismos que deben considerarse uno a uno, a modo de lista de comprobación o examen del indicador en cuestión (cuadro 1.6).

- *Definición de términos.* Equivale en cierto modo a las aclaraciones que también hay que considerar para la construcción de criterios por evaluar en un ciclo de mejora. En esencia, se trata de favorecer la fiabilidad del indicador, explicando sin ambigüedades lo que significa cada uno de los términos que se empleen en su descripción. Cualquier ambigüedad en sus términos provocará falta de fiabilidad en las mediciones. Muchos de los indicadores “oficiales” no pasarían el examen en este primer punto.

- *Tipo de indicador.* Se trata de explicitar si lo que se mide son datos de estructura (existencia de recursos y organizativos), proceso o resultado; además de si está previsto que sea un indicador centinela (un solo caso ya indica un problema por evaluar) o que se mida como tasa, proporción,

La validez y fiabilidad de los indicadores debe comprobarse de forma sistemática.

5.2. Indicadores fiables

La validez asegura que la herramienta mide lo que queremos medir, pero la fiabilidad es una característica que condiciona la validez de la interpretación de los resultados y que se define como el grado de reproducibilidad de los resultados para los mismos casos y situaciones cuando el indicador es utilizado por observadores diferentes. Es decir, que cada uno de los evaluadores haya interpretado y medido el indicador de forma semejante, y por lo tanto obtenga el mismo resultado que el otro o los otros evaluadores al evaluar la misma cosa con el mismo indicador en casos o situaciones semejantes. Sin fiabilidad no hay validez, porque estamos asignando valores a la medición de una forma inconsistente.

La realización de un estudio piloto anterior a la generalización del uso del indicador se considera un paso inexcusable para asegurar la fiabilidad. El piloto permite medir la fiabilidad (concordancia, índice kappa, acuerdo específico) e identificar, discutir y corregir, si las hubiere, diferencias en la interpretación de los indicadores. El proceso del estudio piloto debe repetirse tantas veces como sea necesario hasta alcanzar cotas de fiabilidad aceptables. Una vez asegurada la fiabi-

Cuadro 1.6

Desarrollo / análisis del indicador

Lista de comprobación

- I. Definición de términos
- II. Tipo de indicador
- III. Fundamento (justificación)
- IV. Población a la que se aplica
- V. Fuentes de datos
- VI. Factores subyacentes
 - A. Relacionados con el paciente/usuario
 - B. Ajenos al paciente/usuario
 - 1. Actuación profesional
 - 2. Organización
- VII. Bases de datos (comprobar validez)

Adaptado de: Joint Commission on Accreditation of Health Care Organization (JCAHO): Características de los indicadores clínicos. Control de Calidad Asistencial 1991, 6:65-79.

Al medir la calidad podemos estar interesados en estimar su nivel o en comparar el resultado de la medición con un nivel preestablecido.

El esquema inicial de la JCAHO para indicadores clínicos valora aspectos relativos a la fiabilidad y validez de los indicadores.

Para analizar y construir un indicador conviene utilizar un listado estructurado de aspectos por definir o comprobar, como el propuesto por la Joint Comisión u otros aparecidos con posterioridad.

El principio de utilidad o adecuación de un indicador va unido a la idea de que cada nivel jerárquico debería controlar indicadores relacionados con su responsabilidad en el sistema.

media o como cualquier otro tipo de estimador estadístico. Si es de tipo centinela no se medirá con muestreo, sino por medio de un sistema de vigilancia que controle todos los casos. En el apartado 6 de este capítulo, revisaremos más en detalle los diversos tipos de indicadores.

- *Fundamento.* Describir brevemente por qué lo que vamos a medir es relevante para la calidad del problema, aspecto o servicio por monitorizar. Se relaciona con la validez facial.
- *Población a la que se aplica.* Se refiere a la descripción clara y pormenorizada del universo y de la unidad de estudio en la que se medirá el indicador. La muestra de unidades de estudio será el denominador de los indicadores tipo tasa, proporción y media.
- *Fuentes de datos.* Detallar toda la secuencia de obtención de los datos, desde la identificación de unidades de estudio hasta la medición del indicador en sí.
- *Factores subyacentes.* Se trata de reflexionar sobre la validez del indicador; es decir, reflexionar y hacer explícito qué tipo de factores (problemáticos o no) pueden influir en el resultado de las mediciones. Es esencial distinguir aquellos relacionados con el paciente o usuario, que serán factores de confusión que habrá que controlar, y los que dependen del sistema de salud y de la actuación de los profesionales, que son los que nos interesa identificar como fuente de variación del indicador, de forma que sepamos qué tipo de problema se detectará y qué es necesario evaluar para intervenir, cuando la medición del indicador nos indique calidad defectuosa.
- *Comprobación empírica de la validez.* No siempre es posible, pero sí deseable, fundamentar empíricamente la reflexión sobre los factores subyacentes que pensamos medir con el indicador.

Una explicación más detallada de este esquema, simple y antiguo pero aún útil, principalmente para los indicadores sobre atención clínica, puede verse en el artículo de la JCAHO en el que se basa, publicado en español por la revista de la Sociedad Española de Calidad Asistencial, que figura en la bibliografía de este capítulo. El Ane-

xo 1.1 contiene un esquema estructurado para aplicarlo en la práctica a los indicadores que creamos oportuno valorar. La aplicación del esquema simplificado de la JCAHO ayuda a seleccionar buenos indicadores, que habrá que someter adicionalmente a pruebas de fiabilidad. Sin embargo, un esquema más reciente, desarrollado en el *Veterans Affairs Center for Practice Management and Outcomes Research*, también en Estados Unidos, contempla un proceso de validación de cuatro pasos y hace más énfasis en la necesidad de pilotaje y comprobación empírica de diversos aspectos de la validez y fiabilidad (cuadro 1.7). En este esquema, también se contempla de alguna forma (en el paso 4) la tercera y más importante característica que ha de tener un indicador para incluirlo en un plan de monitorización: que sea útil o apropiado. Otros esquemas de la JCAHO, posteriores al descrito más arriba, también hacen referencia a esta importante característica.

5.3. Indicadores útiles o apropiados

En los programas de gestión de calidad no tiene sentido la monitorización como actividad aislada.

Cuadro 1.7
Proceso para la construcción de buenos indicadores

| | |
|------|--|
| 1. | Selección de indicadores potenciales |
| 1.1. | Valorar la evidencia científica que justifica la importancia del indicador |
| 1.2. | Valorar la facilidad de medición y la frecuencia de casos problemáticos que identifica el indicador |
| 1.3. | Valorar y clasificar los factores modificables que se asocian con los casos problemáticos que identifica el indicador |
| 2. | Establecimiento de estándares y diseño de herramientas de medición |
| 3. | Pilotaje. Medición del indicador en grupos de casos y controles para valorar sensibilidad, especificidad y valor predictivo. Valoración de la fiabilidad |
| 4. | Simulación de la aplicación del indicador en instituciones y niveles para los que se ha diseñado. Valorar su aplicabilidad y utilidad real |

Adaptado de: Hofer TP, Bernstein SJ, Hayward RA y De Monner S. Validating Quality Indicators for Hospital Care. *Jt Comm J Qual Improv*, 1997;23(9):455-467.

da, es preciso reaccionar a sus resultados; por eso, además de válido y fiable, el indicador debe ser útil para la gestión de la calidad en la institución o nivel dentro del sistema de salud en el que se vaya a utilizar; es decir, tiene que ser apropiado para que, en su caso, sus resultados sugieran acciones de mejora. Cuando se detecta un problema, un indicador no es adecuado para un programa de gestión de calidad, si quienes lo utilizan no pueden tomar acciones para la mejora a partir de los resultados de la monitorización.

En la industria, las ideas más sugerentes sobre este tema son probablemente las que aporta Juran cuando explica el concepto de la "pirámide de control", como representación de las cosas que una determinada empresa debe monitorizar en relación con sus productos y procesos en aras de mantener y mejorar su calidad; es una especie de "plan de delegaciones", cuya base son los con-

troles automatizados, y los sucesivos escalones, cada vez más estrechos, son los controles que han de ejercer los encargados de la mano de obra, los supervisores y mandos intermedios y, finalmente, la cúspide del nivel gestor. La idea es que la función de los altos directivos tiene que ver con el control exclusivo de los grandes objetivos estratégicos y con la seguridad de que existe y funciona un sistema de control en el resto de los segmentos de la pirámide. Cada nivel debería controlar indicadores relacionados con su responsabilidad en el sistema.

Sobre la base de estas ideas, adaptándolas a lo que serían las características de un indicador apropiado, según el nivel/responsabilidad dentro del sistema de salud en el que se quiera utilizar, proponemos el esquema que se detalla en la figura 1.2, como punto de partida para un tema que sabemos que es importante; pero que aún

Validez, fiabilidad y utilidad son las tres características que deben tener los indicadores para monitorizar la calidad. La utilidad implica que de sus resultados pueden tomarse acciones para la mejora, si fuese necesario.

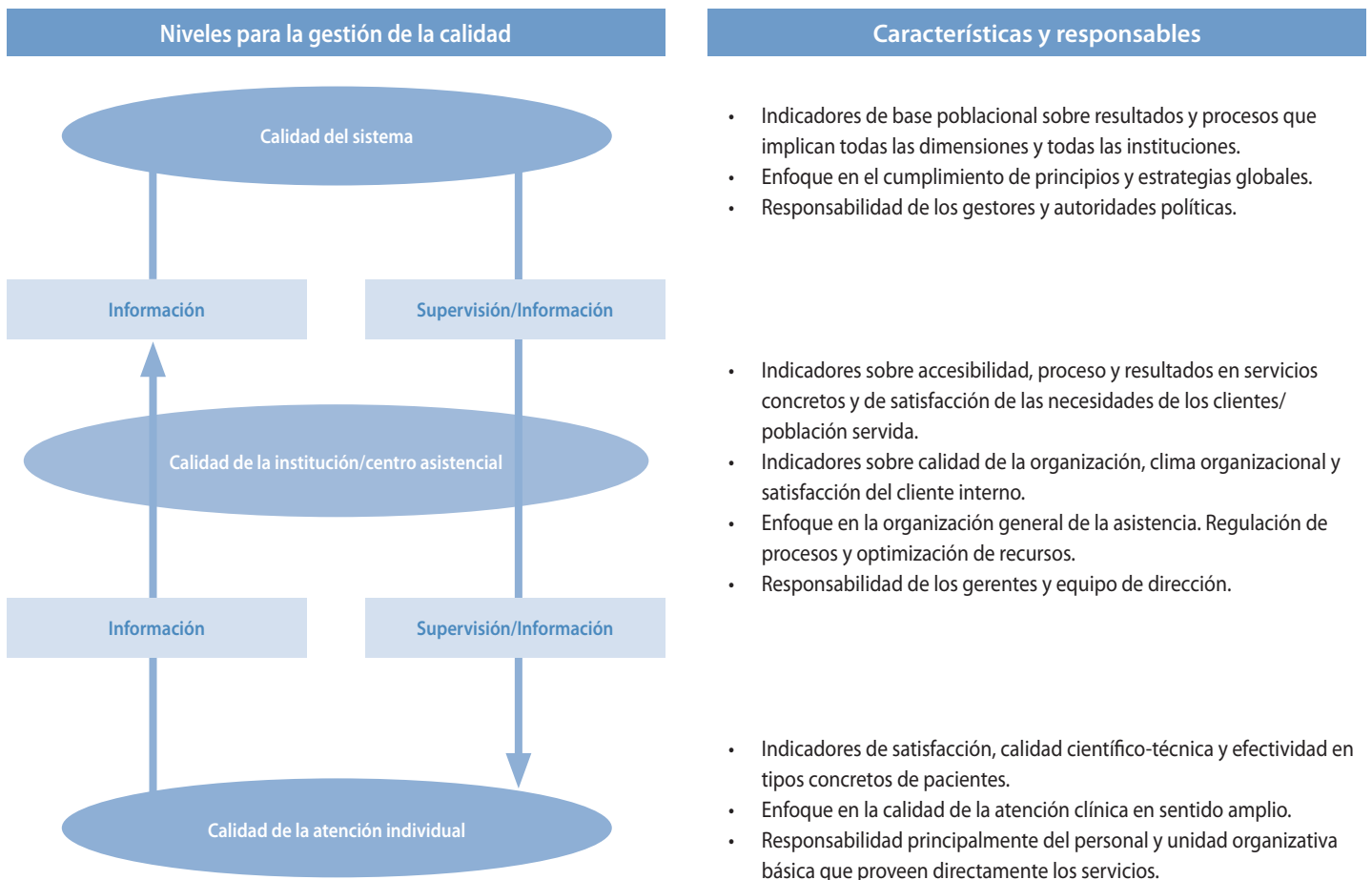


Figura 1.2. Calidad e indicadores en los distintos niveles del sistema

está sin resolver de forma satisfactoria, prácticamente, en ningún sistema de salud. La organización de los programas de gestión de calidad en la red de establecimientos (en la práctica un sistema de salud) del ejército de Estados Unidos (*Veteran Health Administration*), con tres niveles establecidos (nacional, regional y centro hospitalario) y la iniciativa del Servicio Nacional de Salud Británico, que distingue y define indicadores de *alto nivel* (útiles a nivel nacional y de región) y para centros (*trusts*) hospitalarios, son dos claros ejemplos del intento de poner en práctica la delimitación de responsabilidades y niveles para la monitorización.

Uno de los esquemas más referenciados sobre indicadores, en el que se especifican las características que han de tener, es el propuesto en el informe del grupo de trabajo del Instituto de Medicina de Estados Unidos (IOM), *Envisioning the National Health Care Report*, resumido en el cuadro 1.8. Como puede apreciarse, aunque se refiere a los indicadores por utilizar en el ámbito nacional o de sistema, se hace referencia explícita a las tres características esenciales de validez, fiabilidad y utilidad, además de a la preocupación por la factibilidad de su implementación.

En este mismo informe, se propone también un esquema, en realidad una matriz con dos entradas, para identificar los aspectos que deberían tenerse en cuenta para los contenidos de los indicadores. Esta matriz, reproducida en la figura 1.3a, cruza los principales componentes o dimensiones de la calidad que maneja el IOM estadounidense, con la perspectiva o los aspectos que pueden importar al usuario del sistema. Posteriormente (figura 1.3b) se ha ampliado este esquema para incorporar elementos de interés para señalar desigualdades en el sistema. Se supone que a la hora de buscar o construir indicadores sobre un determinado servicio o problema de salud, deberán considerarse indicadores para todas las casillas de la matriz y elegir los que tengan, además, las características que se especifican en el cuadro 1.8.

6. Tipos de indicadores

Además de estar basados en datos de estructura, proceso o resultado, y referidos a determinadas dimensiones de la calidad, pueden distinguirse varios tipos de indicadores en función de cuál sea el tipo de variable que miden (cualitativa o

Cuadro 1.8

Características deseables de los indicadores para el informe nacional sobre la calidad de la atención sanitaria

1. Importancia

- Impacto en la salud: ¿cuál es el impacto en la salud, asociado con el problema que se mide?
- Significado relevante: ¿a los usuarios y al sistema de salud les preocupa lo que se mide?
- Susceptibilidad de ser influido por el sistema de salud: ¿el sistema de salud puede actuar de forma significativa sobre el problema que se mide?

2. Solidez científica

- Validez: ¿el indicador mide realmente lo que se quiere medir?
- Fiabilidad: ¿el indicador proporciona resultados estables en las diversas poblaciones y circunstancias?
- Grado de evidencia explícita: ¿qué evidencia científica existe para respaldar el indicador?

3. Factibilidad de la implementación

- Existencia de prototipos: ¿se está utilizando en la actualidad?
- Existencia de los datos necesarios: ¿la información que se necesita puede recogerse en el nivel y el marco temporal requeridos?
- Coste de la medición: ¿cuánto costará recoger los datos necesarios para medir el indicador?
- Capacidad para realizar análisis comparativo de subgrupos: ¿el indicador puede utilizarse para comparar diferentes grupos de población?

| Perspectiva del usuario | Componentes de la calidad | | | |
|-------------------------------------|---------------------------|-------------|----------------------------------|-------------|
| | Seguridad | Efectividad | Atención centrada en el paciente | Oportunidad |
| Permanecer sano | | | | |
| Mejorar | | | | |
| Vivir con enfermedad o discapacidad | | | | |
| Enfrentar el final de la vida | | | | |

Adaptado de: Hurtado MP, Swift EK, Corrigan JM (eds.). Envisioning the National Health Care Report. National Academy Press, Washington. 2001.

Figura 1.3a. Matriz para clasificar los indicadores para el informe nacional sobre la calidad de la atención sanitaria

| Dimensiones transversales | Componentes de la calidad | Tipo de atención | | |
|---------------------------|--|------------------|-------------------------------------|----------------------------------|
| | | Prevención | Tratamiento de padecimientos agudos | Gestión de enfermedades crónicas |
| Equidad | Efectividad | | | |
| | Eficiencia | | | |
| | Seguridad | | | |
| | Atención centrada en el paciente/familia | | | |
| | Oportunidad | | | |
| | Acceso | | | |
| | Coordinación de cuidados | | | |
| | Infraestructura y capacidad del sistema de salud | | | |

Adaptado de: IOM (Institute of Medicine). Future Directions for the National Healthcare Quality and Disparities Reports. Washington, DC: The National Academies Press. 2010.

Figura 1.3b. Matriz para clasificar los indicadores para el informe nacional sobre la calidad de y disparidades en la atención sanitaria

Un indicador centinela es aquel en que un solo caso significa existencia de problema a evaluar.

cuantitativa) y la forma y parámetro estadístico que resume la medición (número absoluto, proporción, media, ratio, etc.). Estas características (tipo de variable y forma de resumir la medición) pueden repercutir en los métodos de muestreo y medición del indicador, así como en el tipo de distribución estadística que se utiliza para interpretar las mediciones.

En el cuadro 1.9 se describen los principales indicadores de los servicios salud, sus características distintivas y algunos ejemplos de aplicación en la práctica. A continuación revisaremos lo más destacado de todos ellos.

6.1. Indicador centinela

Un indicador se considera centinela cuando un solo caso sirve para disparar la alarma, y ha de ser seguido de la consiguiente evaluación para averiguar las causas de su aparición. Se entiende, pues, que es algo que no debería ocurrir nunca. Con este nivel de exigencia, los indicadores centinela deben reservarse para cuestiones generalmente graves y

transcendentes. En salud pública serían ejemplos de indicador centinela los casos de polio en aquellas comunidades que se suponen con niveles adecuados de vacunación, o los de mortalidad materna a partir de determinados niveles de desarrollo. En los últimos años, se ha identificado un número relativamente elevado de indicadores o situaciones centinela en el medio hospitalario y, la mayoría de las veces, de errores o fallas inaceptables en la asistencia prestada en cuanto a la dimensión de seguridad. Estos eventos centinela (cuadro 1.10) son los que, según los esquemas de la JCAHO, han de conducir al “análisis de las causas raíz” (*Root cause analysis*), aunque se reconoce la dificultad para detectarlos con eficacia y notificarlos. Este listado se ha venido completando y modificando; tanto las actualizaciones como la evolución estadística de los eventos centinela notificados puede verse en la página web de la JCAHO.*

La JCAHO ha propuesto una serie de sucesos centinela que han de conducir al análisis de causas raíz cuando surjan.

* http://www.jointcommission.org/sentinel_event_statistics_quarterly/ (consultado en septiembre 2015)

Cuadro 1.9
Tipos de indicadores

| Tipo | Concepto | Ejemplos |
|----------------------------|--|---|
| Centinela | <ul style="list-style-type: none"> Un solo caso denota problema Medición con sistema de vigilancia | <ul style="list-style-type: none"> Mortalidad materna Casos de polio |
| Número absoluto | <ul style="list-style-type: none"> Nº de eventos que se miden por unidad de estudio No tienen denominador | <ul style="list-style-type: none"> Nº de intervenciones por semana Tiempo puerta-aguja en IAM |
| Proporción y porcentaje | <ul style="list-style-type: none"> El numerador mide la ocurrencia del indicador en un grupo de casos que es el denominador | <ul style="list-style-type: none"> % de pacientes hipertensos controlados % de pacientes que esperan <30' para ser atendidos |
| Tasa | <ul style="list-style-type: none"> Proporción o porcentaje por unidad de tiempo | <ul style="list-style-type: none"> Tasa anual de complicaciones Tasas de cobertura |
| Media | <ul style="list-style-type: none"> Promedio de una serie de mediciones de una variable cuantitativa | <ul style="list-style-type: none"> Media de tiempo de espera antes de ser consultado Estancia media |
| Ratio o razón | <ul style="list-style-type: none"> Cociente entre dos variables o media de medición en cada caso de múltiples variables cualitativas simultáneamente El numerador no está incluido en el denominador | <ul style="list-style-type: none"> Nº de defectos por documento interconsulta Nº de FRCV indagados por paciente |
| Compuesto o índice (index) | <ul style="list-style-type: none"> Resumen o agrupación de varios indicadores específicos | <ul style="list-style-type: none"> Cobertura de screening de cáncer Tasa de problemas de calidad percibida |

Cuadro 1.10

Sucesos centinela en hospitales, según la JCAHO

- Suicidio
- Error médico
- Complicaciones de intervenciones y tratamientos
- Cirugía en sitio equivocado
- Retrasos en el tratamiento
- Muerte tras restricciones indebidas
- Muerte tras fuga
- Asalto, violación u homicidio
- Muerte por transfusión
- Rapto de niños

Adaptado de: 2000 Hospital Accreditation Standards. Standards Intents.

Fuente: The Joint Commission on Accreditation of Healthcare Organization: Oakbrook Terrace, ILL. 2000.

6.2. Indicadores tipo número absoluto

En este tipo especial de indicadores tampoco hay denominador y, por lo tanto, su medición generalmente tampoco se realizará con muestreo. El indicador representa exclusivamente la cantidad, número de eventos o cualquier otra unidad a medir, de aquello que se quiere monitorizar o controlar. La única acotación es la unidad de tiempo en la cual se mide esta cantidad o número absoluto a que se refiere el indicador; se emplean cuando lo relevante es la cantidad y/o porque cualquier resumen en forma de promedio puede enmascarar situaciones relativamente extremas e indeseables. Dos ejemplos serían el número semanal de intervenciones de cirugía menor ambulatoria en un centro de salud (situación en la que lo que interesa es controlar el mantenimiento de la cantidad de intervenciones), y el tiempo puerta-aguja medido en cada individuo que llega al hospital (puerta) con infarto agudo de miocardio, susceptible de tratamiento con trombolíticos (aguja), donde un indicador promedio puede enmascarar una variabilidad indeseable.

6.3. Indicadores tipo proporción, porcentaje y tasa

En todos estos casos, el numerador especifica los casos que cumplen (o incumplen, según se defina

el indicador) una determinada condición o cualidad, de entre todos los que forman el denominador. La diferencia entre los tres tipos de indicador estriba en si lo describimos en tanto por uno (proporción) o por cien unidades del denominador (porcentaje), y entre estos dos y los indicadores tipo tasa (que también pueden medirse como proporción o porcentaje) es que en estos últimos se definen con referencia a una determinada unidad o periodo de tiempo. Esencialmente es la misma diferencia que se establece en epidemiología entre medición de prevalencia (proporción o porcentaje de casos en un momento determinado) e incidencia (casos nuevos, acumulados, en un periodo de tiempo determinado). Todos ellos pueden medirse mediante muestreo. Tres ejemplos serían el porcentaje de pacientes hipertensos controlados; el de pacientes que esperan más de 30 minutos para ser atendidos (indicadores tipo porcentaje), y la tasa anual de complicaciones, reingresos, infecciones, etc. (tipo tasa). Estos son probablemente los más frecuentes en los sistemas de monitorización en servicios de salud.

6.4. Indicadores tipo media

En este tipo de indicadores lo que se mide es una variable cuantitativa (tiempo, días, etc.) en un conjunto de casos, que puede ser una muestra representativa, y se expresa como el valor medio resultante (suma de todas las mediciones dividida entre el total de casos en los que ha hecho medición); por ejemplo, la estancia media en los hospitales y el tiempo de espera expresado también como media. Los indicadores tipo media son de uso relativamente común; pero tienen el inconveniente de no informar por sí solos sobre el grado de dispersión de las mediciones; por ello, son de más utilidad para ser utilizados con técnicas de control estadístico (capítulos 4 y 5) que incorporan también el grado de dispersión para ser interpretados correctamente. Pueden transformarse en indicadores de tipo proporción o porcentaje, al fijar un límite a partir del cual se considera defecto o mala calidad, y medir en cada caso de la muestra si están o no por encima de ese límite. Por ejemplo, en vez de expresar un indicador como la media tiempo de espera, puede expresarse como el porcentaje de pacientes que ha esperado más de un cierto tiempo (por ejemplo, 30 minutos).

En los indicadores tipo número absoluto no hay denominador. Se utilizan cuando lo que interesa es controlar la cantidad, o impedir que valores promedio enmascaren situaciones con variabilidad extrema.

Los indicadores tipo proporción, porcentaje y tasa son probablemente los más utilizados en servicios de salud y pueden medirse con muestreo.

Los indicadores tipo media resumen la medición de una variable cuantitativa en una serie de casos, que puede ser una muestra. Tienen el inconveniente de no informar sobre el grado de dispersión de lo que se mide.

Los indicadores tipo media pueden transformarse en indicadores tipo porcentaje fijando un valor que separe buena y mala calidad.

Los indicadores tipo ratio resumen la medición simultánea de varios criterios, o un mismo criterio varias veces, en cada unidad de estudio. Se expresan generalmente como promedio de errores por caso evaluado y siguen la distribución probabilística de Poisson.

Los indicadores compuestos agrupan varios indicadores en uno solo. Son de especial interés para comparar centros, áreas o sistemas, y se les está prestando un interés creciente en todos los campos.

Los planes de monitorización incluyen la frecuencia de las mediciones, o mecanismos de la recolección de datos, y los métodos para su interpretación. De todo ello, lo que más va a influir en la práctica concreta es la frecuencia de mediciones y el interés que podamos tener en estimar el nivel de calidad.

6.5. Indicadores tipo ratio o razón

En este tipo de indicador el numerador no está incluido estrictamente en el denominador. Ejemplos comunes en epidemiología y demografía son la razón de sexos y la densidad demográfica (habitantes/km²). Otros indicadores tipo razón son aquellos que miden simultáneamente varios criterios o variables cualitativas (por ejemplo, los defectos de un documento interconsulta, o los factores de riesgo cardiovascular que han sido identificados) y el indicador expresa el promedio de cumplimientos (o incumplimientos) por caso analizado. En estos ejemplos puede parecer que fuese en realidad un indicador tipo media, dado que puede mirarse siempre el mismo número de criterios; pero el mismo razonamiento y el mismo tipo de indicador resulta cuando medimos cuántas veces se presenta un mismo defecto en cada caso analizado (como el número de errores tipográficos por página), sin que esté predeterminado cuál es el número máximo de veces que medimos el criterio (en nuestro ejemplo *presencia de error tipográfico*) en cada página (el límite será el número de palabras, pero éste también puede variar de página a página). Estos indicadores siguen, desde el punto de vista estadístico, la distribución probabilística de Poisson y son especialmente interesantes para su uso con métodos de control estadístico, midiéndolos con muestras representativas.

6.6. Indicadores compuestos o índice

Los indicadores compuestos agrupan, resumiéndolos, varios indicadores en uno solo; por lo tanto, por definición, tienen una de las características deseables en un indicador: resumir en un número la calidad del servicio o de un conjunto de aspectos relevantes del mismo. Como ejemplos podemos citar la cobertura de screening de cáncer en la mujer, que incluye cáncer de mama y de cuello de útero, o el conocido como tasa de problemas que resume en las encuestas de calidad percibida todos los problemas encontrados por el entrevistado, expresado como porcentaje de las respuestas que indican problemas sobre el total de preguntas (no encuestas o individuos) respondidas.

Continuamente se están proponiendo y utilizando nuevos esquemas de indicadores com-

puestos que, aunque de inicio son generalmente discutidos, van imponiéndose en la práctica en aras de la tan deseada parsimonia en la cantidad de números que sirvan para monitorizar cualquier servicio, centro o sistema de salud de una forma; sin embargo, lo más amplia e integradora posible. Su característica como número-resumen los hace especialmente útiles como indicadores para comparar centros, áreas o sistemas, aunque también pueden utilizarse para unidades organizativas más simples. Dada su creciente importancia y la existencia de varios y bien definidos subtipos, merece la pena que les prestemos especial atención, por lo que, en el capítulo 2, los tratamos más en detalle.

7. Componentes de un plan de monitorización. Tipos de plan

Una vez definidos y seleccionados los indicadores, sea cual fuere la situación o punto de partida (cuadro 1.3) o el tipo de indicador (cuadro 1.9), las principales características a describir en el plan de monitorización son: (i) la frecuencia de las mediciones; (ii) los mecanismos de recolección de datos, y (iii) los métodos que guían el diseño de la recolección y su interpretación.

La decisión sobre la *frecuencia* de la medición puede determinar en gran medida otras características del plan de monitorización, incluido el método o métodos de recolección e interpretación de los datos que pueden ser de mayor utilidad. Cada situación y tipo de indicador puede requerir de mediciones con mayor o menor frecuencia. Actualmente, en los servicios de salud, lo más común en los programas externos es realizar mediciones anuales o, cuando menos, mensuales; pero puede darse la necesidad u oportunidad de realizar mediciones más frecuentes. En relación con su frecuencia podemos distinguir entre planes con frecuencia elevada (continuos o de intervalos cortos entre mediciones), y planes espaciados (con mediciones a intervalos largos). Vamos a considerar, de una forma un tanto arbitraria, un mes como límite de frecuencia entre ambos tipos de planes.

Una característica común a cualquier plan de medición que se realice con muestras es la necesidad de que éstas sean aleatorias, debiéndose decidir y definir, como en las evaluaciones, el *tipo de muestreo* (aleatorio simple, sistemático, estratificado, etc.) y la *mecánica de obtención* de las unida-

des de estudio en las que mediremos el indicador. Esta mecánica ha de estar en relación con el método de *monitorización elegido*.

En relación con lo anterior, tal como se resume en el cuadro 1.2, siempre está la opción de enfrentarse a las mediciones del indicador como si fuesen estimaciones del nivel de cumplimiento del mismo, cada vez que lo midamos, y comparar nuestros resultados con el estándar o nivel de cumplimiento de referencia; el método y análisis de las mediciones serían entonces semejantes a la estimación del nivel de calidad y comparación de dos evaluaciones. Sin embargo, si renunciamos a conocer el nivel real de cumplimiento del indicador que medimos, y centramos nuestra atención en saber si la situación es o no problemática, podemos utilizar métodos desarrollados en la industria, cuya ventaja es que utilizan muestras más pequeñas, además de que sirven, en el caso de los gráficos de control estadístico, para un conocimiento más profundo del tipo de problema existente (esporádico, sistemático, tendencia). Nos parece especialmente relevante conocer esta metodología, todavía no generalizada en los servicios de salud, a pesar de su utilidad potencial. La medición con el objetivo de estimar con cierta precisión el nivel de cumplimiento de los indicadores es la de mayor tradición en los servicios de salud, y puede consultarse en cualquier libro de estadística básica y, aplicado al campo de la calidad, en el capítulo sobre medición de criterios o requisitos de la calidad para la realización de ciclos de mejora del Manual sobre métodos y herramientas para la realización de ciclos de mejora.

8. Métodos de monitorización según las características del plan

Centrándonos en los métodos desarrollados en la industria, podemos distinguir, siguiendo su propia terminología, dos categorías: para inspección de la calidad y para el control de procesos. La inspección se define como la medición de unos determinados indicadores y su comparación con los requisitos o estándares de calidad para determinar su cumplimiento o su conformidad con ellos; el control de procesos implica la medición y análisis de la variación de los indicadores seleccionados que han de representar aspectos relevantes del proceso sometido a análisis y control. En ambos casos, los planes de medición implican

mediciones sistemáticas y repetidas con una periodicidad dada que es normalmente mayor para el control de procesos.

La inspección se contempla generalmente como un método transversal y estático, orientado a medir la calidad en una población (grupo o lote) del cual se extrae la muestra, para decidir si aceptar o no que su calidad es adecuada; mientras que el control de procesos es más una actividad continua y dinámica para tomar decisiones sobre el proceso controlado. Estas decisiones, esencialmente, se toman en el sentido de aceptar o rechazar si hay que intervenir o investigar una variabilidad que puede resultar atípica o inesperada, lo que le asemeja al tipo de decisión (aceptación/rechazo de situación problemática) que se realiza en la inspección; sin embargo, hoy en día existe una discusión abierta sobre la conveniencia de realizar o no *inspecciones*, dado que el énfasis debe ponerse en el control de procesos y los programas de prevención que consigan los niveles de *conformidad* deseados. Esta discusión no es sobre los métodos en sí, sino sobre la oportunidad y objetivos al aplicar uno u otro enfoque.

El muestreo para la aceptación de lotes (*Lot Quality Acceptance Sampling*, LQAS) y los gráficos de control estadístico de la calidad representan los métodos más comunes de inspección y de control de procesos que pueden utilizarse de forma eficiente en la monitorización de la calidad en los servicios de salud. A pesar de las diferencias conceptuales que podamos encontrarles, ambos tienen en común dos características relevantes:

- (i) No proporcionan como resultado estimaciones del nivel de calidad. En su lugar informan, de distinta manera, sobre diferencias estadísticamente significativas en relación con el nivel estándar o condiciones predefinidas. Esta información se utiliza para clasificar como aceptable o problemático el proceso, producto o servicio bajo escrutinio. El que no sean útiles para dar estimaciones del nivel de calidad puede verse como una desventaja para el personal de los servicios de salud, acostumbrados a trabajar con estimaciones de niveles, incidencias, prevalencias, cifras que miden los diversos parámetros fisiológicos, etc. Sin embargo, estos métodos nos dicen que no es necesario tener estimaciones del nivel de calidad para cumplir con el objetivo de la monitorización, si

Pueden distinguirse dos grandes categorías de métodos de monitorización: inspección de la calidad y control de procesos.

La inspección de la calidad es un método transversal y estático, y consiste en valorar la conformidad del indicador con los niveles (estándar) preestablecidos.

El control de procesos es continuo y valora la variabilidad del indicador a lo largo del tiempo, no sólo su conformidad con el estándar.

Si renunciamos a estimar el nivel de calidad, podemos utilizar métodos desarrollados en la industria como el LQAS y el control gráfico de la calidad, que utilizan en general muestras más pequeñas.

la consideramos como una actividad para el *screening* o identificación de problemas. La renuncia a tener estimaciones es una de las condiciones para la segunda característica de los métodos de control de la calidad en la industria.

- (ii) Permiten generalmente tomar decisiones al utilizar muestras mucho más pequeñas que las que necesitaríamos, si quisiéramos decidir por medio de la comparación de una estimación del nivel de calidad al estándar prefijado. Las decisiones sobre la aceptación (tras *inspecciones*) o la existencia de variación problemática (a través del *control de procesos*) pueden tomarse con muestras relativamente pequeñas, especialmente, con la utilización práctica e inteligente de la estadística y, más en concreto, de la teoría de la probabilidad.

Al margen de las discusiones teóricas, hay una diferencia importante en la aplicación práctica de ambos métodos: el control gráfico sólo es aplicable en planes de medición continua o de intervalos cortos entre mediciones, mientras que el LQAS puede utilizarse en cualquier tipo de plan, sea con mediciones a intervalos cortos o largos, entendiéndose por intervalo corto, tal como apuntábamos más arriba, si es mensual o menor. El cuadro 1.11 resume esta distinción, sobre la que subrayamos que los métodos de control estadístico gráfico son

| Cuadro 1.11 Planes y métodos de monitorización, según la frecuencia de las mediciones | |
|---|--|
| Plan (según frecuencia de mediciones) | Métodos |
| <ul style="list-style-type: none"> ● Espaciado (intervalos largos* o esporádicos) | <ul style="list-style-type: none"> ● Evaluación rápida ● LQAS** |
| <ul style="list-style-type: none"> ● Continuo (intervalos cortos) | <ul style="list-style-type: none"> ● Control estadístico de la calidad (gráficos) ● LQAS |

*: Intervalo largo: mediciones con frecuencia superior a un mes.
 **: LQAS: Lot Quality Acceptance Sampling= Muestreo para la aceptación de la calidad de lotes.

tanto más útiles cuanto más frecuentes (menos espaciadas) sean las mediciones, y que el LQAS es probablemente el método de elección para monitorizar y priorizar problemas de calidad si esta actividad se realiza cada varios meses, anualmente o de forma esporádica. El conocimiento detallado de la práctica de ambos métodos hace evidente esta distinción, como veremos en los capítulos correspondientes en este manual.

BIBLIOGRAFÍA

- Agency for Healthcare Research and Quality. Methods used in the National Healthcare Reports [monografía en internet]. Rockville, MD: AHRQ, 2014 [consultada en febrero de 2015]. Disponible en: <http://nhqrnet.ahrq.gov/inhqrd/resources/methods>
- Halpern J. The measurement of quality of care in the Veterans Health Administration. *Med Care* 1996;34(3):M555-M568.
- Health and Social Care Information Centre [sitio en internet]. Indicator portal. Londres, UK: HSCIC, c2015 [consultado en febrero de 2015]. Disponible en: <http://www.indicators.ic.nhs.uk>
- Hurtado MP, Swift EK y Corrigan JM (eds.). *Envisioning the national health care report*. Washington, DC: National Academy Press, 2001.
- Institute of Medicine of the National Academies. *Future directions for the national healthcare quality and disparities reports*. Washington, DC: The National Academies Press, 2010.
- Joint Commission on Accreditation of Health Care Organization. Características de los indicadores clínicos. *Control de calidad asistencial*. JCAHO 1991;6:65-79.
- Juran, JM, Gryna FM, Bingham RS. *Manual de control de la calidad*. 2º ed. Barcelona: Reverté, 1990.
- Kelley EE, Hurst J. Health care quality indicators project. Conceptual framework paper. DELSA/HEA/WD/HWP(2006)3.
- Organisation for Economic Co-operation and Development. *Geographic variations in health care: what do we know and what can be done to improve health system performance?* [monografía en internet]. OECD Health Policy Studies, 2014 [consultada en febrero de 2015]. Disponible en: <http://dx.doi.org/10.1787/9789264216594-en>
- Organisation for Economic Co-operation and Development. *Patient safety data systems in the OECD: a report of a Joint Irish Department of Health-OECD Conference, Health Care Quality Indicators Project*, Directorate for Employment, Labour and Social Affairs Health Committee, DELSA/HEA/HCQ(2007)1.
- Saturno PJ. Qué, cómo y cuándo monitorizar: marco conceptual y guía metodológica. *Revista de Calidad Asistencial* 1998;13:437-443.

Anexo 1.1

Análisis crítico de un indicador

Paso 1. Selección del indicador

De todos los indicadores conocidos o utilizados en el entorno del participante, seleccionar uno, preferentemente clínico, en relación con un proceso o resultado de la atención a un problema de salud.

Paso 2. Análisis de sus características

- 2.1. Describir sus características de acuerdo con los siete puntos de la *lista de comprobación* del cuadro 1.6.
- 2.2. Indicar en qué aspectos puede mejorar, y si éstos corresponden a la validez y/o a la fiabilidad del indicador.

Paso 3. Reflexión sobre su utilidad

- 3.1. Indicar en qué nivel del sistema de salud podría ser potencialmente útil (apropiado para reaccionar con iniciativas de mejora) como herramienta de monitorización.
- 3.2. Analizar por qué.

Capítulo 2

Indicadores compuestos

Contenido general y objetivos específicos

1. Indicadores compuestos. Concepto, tipos y características
2. Precauciones metodológicas para la construcción de indicadores compuestos
3. Tipos de indicadores compuestos, según el método de construcción
4. Atributos deseables de los indicadores compuestos

Anexo 2.1. Construcción y análisis de un indicador compuesto

Contenido general

En este capítulo se describen los diversos tipos de indicadores compuestos que pueden utilizarse en la monitorización de la calidad de servicios de salud. La importancia que se le concede a este tipo de indicadores va en aumento, sobre todo para su utilización en comparaciones entre centros, áreas y países.

Objetivos específicos

1. Describir las características de un indicador compuesto.
2. Distinguir los diversos tipos de indicador compuesto, según el método de construcción.
3. Identificar las precauciones metodológicas que hay que tener en cuenta en la construcción de los indicadores compuestos.
4. Realizar un análisis crítico de los indicadores compuestos existentes.
5. Describir los atributos deseables en un indicador compuesto.
6. Identificar las situaciones en las que puede ser útil un indicador compuesto.

Indicadores compuestos

1. Indicadores compuestos. Concepto, tipos y características

Los indicadores compuestos resumen en un solo número las mediciones de dos o más indicadores o criterios, que llamaremos subindicadores. Intuitivamente sería deseable que los subindicadores se relacionaran entre sí, incluso estadísticamente; sin embargo, el principal cuerpo teórico que está desarrollándose en este campo considera que los subindicadores no tienen por qué estar necesariamente relacionados, ni siquiera tener una unidad de medida común, ni tener cada uno de ellos una importancia relativa tal, que pueda guiarnos en una eventual ponderación para agruparlos en un solo indicador compuesto.

La atención que se les está prestando a los indicadores compuestos es relativamente reciente; pero en poco tiempo se han situado en el centro del debate teórico-práctico, tanto en el campo social y económico como en el sanitario. En el campo social y económico, las razones para este interés, según se resumen en el documento base de una reunión consultiva de la Unión Europea sobre este tema, celebrada en Bruselas en marzo de 2002, son, entre otras, las siguientes:

- Pueden utilizarse para resumir aspectos o problemas complejos y multidimensionales, en la toma de decisiones.
- Proveen una visión global; pueden ser interpretados más fácilmente que el escrutinio de muchos indicadores por separado, y facilitan los rankings de países y regiones en cuestiones complejas.
- Son generalmente mejor entendidos y atractivos para el gran público, al proporcionar una cifra resumen para comparar sistemas y países.

- Ayudan a reducir el tamaño de los listados de indicadores o a incluir más información dentro de unos límites de espacio establecidos.

Sin embargo, tienen en su contra que podrían ser demasiado simplistas cuando no se les construye bien y se malinterpretan, carecer de transparencia en las valoraciones y decisiones adoptadas en cuanto a la selección de los subindicadores y el método de agrupación y, a pesar de todo, requerir de una gran cantidad de datos para su construcción, puesto que hay que medir todos los subindicadores incluidos.

A pesar de todo, las circunstancias potencialmente negativas se consideran retos metodológicos por solucionar, más que inconvenientes reales que frenen una tendencia y un interés en este tipo de indicadores que van muy visiblemente en aumento. Sirva como ejemplo la iniciativa del *Joint Research Centre* (JRC) de la Comisión Europea y la Organización para la Cooperación y el Desarrollo Económicos (OCDE), cuyo primer taller sobre indicadores compuestos sobre el desempeño económico y social de los países, celebrado en mayo de 2003, se presentó como un verdadero hito en el que se revisó lo realizado hasta ese momento y se propusieron actividades de continuidad sobre el mismo tema.* Es de especial interés, para quien quiera profundizar en el tema, el documento elaborado por el Grupo de Estadística Aplicada del JRC sobre la situación actual de la metodología y práctica para el de-

* Los contenidos, materiales presentados y conclusiones de este taller, pueden consultarse en: http://webfarm.jrc.cec.eu.int/uasa/events/oecd_12may03.

La monitorización es una de las actividades necesarias para llevar a la práctica el compromiso con la mejora de la calidad.

Los indicadores compuestos se han situado en poco tiempo en el centro del debate teórico-práctico en instancias internacionales, tanto en el campo social y económico como en el sanitario.

Los inconvenientes detectados en los indicadores compuestos es que representan retos metodológicos a resolver, más que frenos para su utilización.

La Unión Europea y la OCDE están apoyando explícitamente iniciativas para profundizar en la metodología de construcción y uso de indicadores compuestos para comparar países.

Las organizaciones mundiales también han desarrollado y utilizado indicadores compuestos como el Índice de desarrollo humano, de la ONU, o el Indicador resumen, de la OMS, para comparar los sistemas de salud.

sarrollo de indicadores compuestos (*State-of-the-art Report on Current Methodologies and Practices for Composite Indicator Development*), citado en la bibliografía de este capítulo.

Uno de los indicadores compuestos más referenciado en el campo económico y social es el llamado Índice de desarrollo humano (*Human Development Index*), un indicador utilizado por la Organización de las Naciones Unidas (ONU) que agrega la esperanza de vida, las tasas de escolaridad y analfabetismo, y GDP per cápita. En el campo sanitario, la Organización Mundial de la Salud (OMS) es también un organismo de la ONU, la cual construyó y utilizó, en el Informe de Salud Mundial del año 2000, un indicador compuesto que agrupa cinco componentes para clasificar los sistemas de salud de 191 países y que ha sido discutido desde muchos puntos de vista, pero que, indudablemente, también ha contribuido a centrar la atención en este tipo de indicadores para seguir perfeccionándolos. El cuadro 2.1

contiene ejemplos de utilización de indicadores compuestos en sistemas de salud.

Como consecuencia de estas iniciativas, en la actualidad existe una cierta doctrina sobre la metodología de construcción de los indicadores compuestos, y se han identificado aquellos pasos o decisiones que pueden resultar especialmente problemáticos, los cuales pasaremos a revisar brevemente.

2. Precauciones metodológicas para la construcción de indicadores compuestos

El cuadro 2.2 resume el esquema metodológico general y las principales consideraciones metodológicas para cada paso.

De todas las decisiones mencionadas en el cuadro 2.2, la mayoría de las discusiones se han venido centrando en las correspondientes al paso dos (selección de subindicadores), al cuatro (valoración de la relación existente entre los subin-

Cuadro 2.1
Ejemplos de indicadores compuestos en sistemas de salud

- USA: ranking estatal sobre la calidad de atención a pacientes de Medicare (Smith, 2002):
 - 22 indicadores (sobre atención a infarto, ACV, fallo cardíaco, neumonía, cáncer de mama, diabetes).
 - Cada estado tiene una posición en la lista de cada indicador.
 - El indicador resumen para cada estado es la suma, o el promedio, de las posiciones en los 22 listados.
- España (García-Eroles *et al.*, 2001): Top 20. Clasificación de los hospitales sumando su posición en el ranking de seis indicadores (mortalidad, complicaciones, estancia media, índice de ocupación, coeficiente de ambulatorización, coste por unidad de producción). Un acercamiento parecido al ranking *100 Top Hospitals* de Estados Unidos.
- Canadá: combinación de 15 indicadores (relativos a atención prenatal, salud comunitaria, servicios para los ancianos, eficiencias, recursos) agrupados y ponderados en función de seis categorías, para cada una de las 54 regiones sanitarias (Smith, 2002).
- INHS: indicador compuesto de alto nivel. Tasas medias ponderadas de varios indicadores relacionados (cuadro 2.5) (NHS, 1999).
- OMS: indicador compuesto (que reúne resultados en salud, desigualdades en salud, respuesta del sistema de salud, desigualdades en la respuesta del sistema, financiamiento equilibrado) para clasificar los sistemas de 191 países (World Health Report 2000) (Smith, 2002).
- Indicador del King's Fund para comparar las 120 *Health Authorities* de Inglaterra, Escocia y País de Gales (NHS, 1999), compuesto de seis subindicadores ponderados (muertes por cáncer, muertes por enfermedad cardíaca, número de personas en lista de espera, número de personas en lista de espera más de 12 meses, número de operaciones de cadera, muertes prematuras y evitables).
- USA: construcción de cuatro indicadores compuestos (*atención en la consulta, calidad clínica, accesibilidad y servicio al cliente* (Zaslavsky *et al.*, 2002). y *vacunaciones*) combinando 23 subindicadores (nueve del Consumer Assessment of Health Plan Study, CAHPS, y 14 del Health Plan Employer Data Information System, HEDIS).
- USA: comparación estatal con base en dos indicadores compuestos que representan (a) la distancia a benchmark y (b) comparación con la media nacional, de hasta 300 indicadores (AHRQ, 2014).

*: Referencias en la bibliografía de este capítulo.

Cuadro 2.2
Esquema general para la construcción de indicadores compuestos:
pasos y principales consideraciones metodológicas

| Paso | Considerar |
|--|---|
| 1. Decidir sobre el fenómeno, aspecto o problema por medir | Hasta qué punto se beneficiaría de un indicador compuesto. |
| 2. Selección de subindicadores | Justificación de la selección: relevancia y validez de cada indicador. No hay ninguna fórmula inequívoca para seleccionar qué y cuántos subindicadores. |
| 3. Valorar la existencia y calidad de los datos para la medición de los subindicadores seleccionados | Métodos para tratar la ausencia parcial de datos (<i>missing</i>) para algún subindicador. |
| 4. Valorar la relación existente entre los subindicadores | Pueden ser de utilidad métodos como el análisis de componentes principales, factorial y de correlación en general. |
| 5. Normalización, ponderación y agregación de los subindicadores | Justificar la decisión sobre cuál de los múltiples métodos posibles sería el más adecuado. |
| 6. Comprobar la solidez y estabilidad de los resultados | Realizar análisis de sensibilidad a los cambios en la composición del indicador (ejemplo: subindicadores que se incluyen, ausencia parcial de datos, método de ponderación) y expresar resultados con límites de confianza. |

Fuente: Elaboración propia a partir de: Applied Statistics Group: State-of-the-art Report on Current Methodologies and Practices for Composite Indicator Development. 1st Workshop on Composite Indicators of Country Performance. Joint Research Centre. European Commission, 2003

De todas las decisiones metodológicas que hay que tomar en la construcción de indicadores compuestos, las más discutidas son la valoración de la relación existente entre los subindicadores y, sobre todo, los mecanismos de agregación, normalización de los datos y eventual ponderación de los subindicadores.

dicadores) y, sobre todo, al cinco (normalización, ponderación y agregación de los subindicadores). Sobre la valoración de la relación entre los subindicadores, ya hemos mencionado que existen dos posiciones contrapuestas: considerar que es necesario que se relacionen para que el indicador compuesto sea consistente y, lo contrario, considerar que un indicador compuesto puede resumir varios y diferentes aspectos de un fenómeno, no necesariamente relacionados entre sí, aunque todos tengan el nexo del fenómeno general que contribuyen a medir. La mayoría de los indicadores compuestos utilizados parece asumir este segundo acercamiento, aunque con cierta frecuencia también se utilizan técnicas como el análisis factorial y el análisis de componentes principales para agrupar estadísticamente los subindicadores, en grupos, factores o conglomerados que funcionan entonces como variables para la agregación en un solo indicador compuesto, incluso con la ponderación que pueda resultar de este

análisis estadístico. Todo ello no obvia, sino que complementa, la importantísima consideración inicial sobre la adecuación y representatividad de los subindicadores seleccionados en relación con el fenómeno que se va a medir.

Sin embargo, lo que está produciendo más controversia y discusión metodológica es la relativa necesidad de conversión de las escalas de medida (normalización) de los subindicadores, y de ponderación y sus métodos. En función de estas características pueden identificarse diferentes tipos de indicadores compuestos, con métodos de elaboración más o menos complejos, según se describe de forma resumida en el cuadro 2.3.

Un análisis detallado de cada uno de ellos se escapa del alcance de este manual; no obstante, señalaremos lo más destacado de cada uno de los tres grandes grupos, sin ponderación y con ponderación, y dentro de los primeros sin reconversión y con reconversión de las escalas de medida de los subindicadores.

Cuadro 2.3
Indicadores compuestos: tipos y métodos de agregación

| Tipo | Método |
|---|--|
| I. Sin ponderación de los subindicadores* | |
| I.1. Sin reconversión de escalas de medida de los subindicadores | |
| I.1.1. Suma de rankings | $\sum_{i=1}^N \text{Rank}_i$ |
| I.1.2. Ranking medio | $\frac{\sum_{i=1}^N \text{Rank}_i}{N}$ |
| I.1.3. Porcentaje simple | $\frac{\sum \text{cumplimientos de los subindicadores}}{\sum \text{de ocasiones en los que se evalúa cada subindicador}} \times 100$ |
| I.1.4. Media de cumplimientos | $\frac{\sum \% \text{ de cumplimientos de cada subindicador}}{\text{Núm. de subindicadores evaluados}}$ |
| I.1.5. Porcentaje de 100% de cumplimientos | $\frac{\text{casos en los que se cumplen todos los subindicadores}}{\text{casos evaluados}} \times 100$ |
| I.1.6. Suma de número de subindicadores por encima de la media menos número de subindicadores por debajo de la media. | $\sum_{i=1}^N \text{sgn}[x_i - \bar{x}_i]$ |
| I.2. Con reconversión de escalas de medida de los subindicadores | |
| I.2.1. Suma de ratios o de proporciones de las diferencias con la media. | $\sum_{i=1}^N \frac{x_i}{\bar{x}_i} \quad \text{o} \quad \sum_{i=1}^N \frac{x_i - \bar{x}_i}{\bar{x}_i}$ |
| I.2.2. Suma o media de los valores (absolutos o diferencias con medias) estandarizados. | $\sum_{i=1}^N \frac{x_i - \bar{x}_i}{\sigma_{x_i}} \quad \text{o} \quad \sum_{i=1}^N \frac{x_i}{\sigma_{x_i}}$ solos o promediado a \sum en numerador y N en denominador |
| I.2.3. Suma de valores reconvertidos en función de rango. | $\sum_{i=1}^N \frac{x_i - \min(x_i)}{\text{rango}(x_i)}$ |

Continúa/

3. Tipos de indicadores compuestos, según el método de construcción

Las principales características de los distintos tipos de indicadores compuestos que se reseñan en el cuadro 2.3 son las siguientes:

3.1. Indicadores compuestos sin reconversión de escala ni ponderación de los subindicadores

Son los más sencillos de calcular y fáciles de entender, ya que consisten en la medición de los subindicadores y el establecimiento de una forma concreta de agregar los resultados en un solo nú-

mero que es el indicador compuesto. Lo que los diferencia entre sí es el método de agregación, el cual origina las ventajas e inconvenientes de cada uno de ellos, como veremos a continuación.

La suma o el promedio de los rankings obtenidos en cada subindicador para comparar un número o *ranking-resumen*, método utilizado para comparar los diferentes estados en relación con la calidad de la atención a pacientes de Medicare (cuadro 2.1) y que utiliza la clasificación Top 20 de los hospitales en España, tiene como problema que puede producir valores medios que enmascaran situaciones de subindicadores con niveles muy bajos compensados por otros con niveles altos. Problema semejante al que tiene la media de

Los indicadores compuestos sin reconversión de escala ni ponderación de los subindicadores son los más fáciles de calcular y entender.

/continuación

| II. Con ponderación de los subindicadores | |
|---|---|
| II.1. Análisis de componentes principales y análisis factorial | Combinaciones estadísticas de grupos de subindicadores |
| II.2. Regresión múltiple | Modelo de regresión múltiple con los subindicadores como variables independientes y un indicador de referencia como dependiente |
| II.3. Neutralización del efecto de las correlaciones entre subindicadores | $\sum [2 \cdot (1 + r_{x_i, y_i})]^{1/2} (x_i + y_i)$ |
| II.4. Distancia a frontera de eficiencia | Ponderación en función de los valores <i>benchmark</i> de la combinación de subindicadores |
| II.5. Distancia a valores estándar | Ponderación en función de los valores-objetivo o estándar de cada subindicador |
| II.6. Opinión de expertos (distribución de presupuesto imaginario) | Ponderación según cantidad otorgada a cada subindicador |
| II.7. Opinión de la población | Ponderación según valor otorgado por consultas estructuradas a la población o sus representantes |
| II.8. Proceso analítico jerárquico (AHP) | Utilización de métodos cuantitativos y cualitativos en comparaciones sucesivas por pares de subindicadores |

*: Este grupo de métodos no implica necesariamente ponderación, pero también pueden utilizarse con ponderación previa de los subindicadores, en combinación con alguno de los métodos listados en el apartado II de este cuadro.

N = N° total de subindicadores; i = serie de subindicadores; x = resultado de la medición de un subindicador; sgn = signos;

r_{x_i, y_i} = coeficiente correlación entre dos subindicadores.

cumplimientos de los subindicadores como indicador compuesto, método que utilizaba el extinto Insalud español para resumir la valoración de las llamadas *normas técnicas mínimas* (en realidad una serie de criterios o subindicadores).

El indicador compuesto que indica el porcentaje de casos que cumplen todos los subindicadores (con 100% de cumplimientos) tiene el inconveniente de ser tremendamente restrictivo. Un solo subindicador con niveles bajos de cumplimiento condiciona una impresión general de mala calidad, con independencia de los niveles que puedan tener los otros subindicadores.

Mención aparte dentro de este grupo merecen los indicadores compuestos tipo porcentaje simple, que son probablemente los de más atractivo y utilidad potencial, además de ser, efectivamente, los más utilizados en la práctica, al menos en el ámbito de las encuestas de calidad percibida.

Los indicadores tipo porcentaje simple están contruidos para responder a la pregunta: ¿con qué frecuencia han recibido los pacientes atención de calidad?, que puede referirse a diferentes

ámbitos dentro del sistema (servicios, centros, áreas, regiones, sistemas de salud) y unidades de estudio (global, determinados procesos, pacientes, aspectos, etcétera).

Su principal peculiaridad es que cada caso evaluado proporciona múltiples subcasos, tantos como subindicadores sea pertinente evaluar en cada caso concreto, y podrían necesitarse muestras más pequeñas para contestar con precisión a esa pregunta general (*atención de calidad o sin problemas*), que las que se necesitarían para evaluar correctamente cada uno de los subindicadores. El cuadro 2.4 contiene un ejemplo clínico en este sentido. Como puede verse, aunque sólo se evalúan 15 casos, para el indicador compuesto es como si fueran 40, que son las veces en que ha sido pertinente evaluar los subindicadores, considerando cada una de ellas como una *oportunidad* diferenciada para actuar con calidad o no. El indicador compuesto de este ejemplo nos dice que se ha dado atención de calidad en 70% de las ocasiones en las que se tuvo la oportunidad de decidir una opción de calidad.

El método de agregación de los subindicadores es la base de los diferentes tipos de indicadores compuestos, así como de las principales y peculiares ventajas e inconvenientes de cada tipo.

Dentro del grupo de indicadores compuestos sin reconversión de escala ni ponderación de los subindicadores, tienen especial interés y utilidad los indicadores del tipo porcentaje simple.

Cuadro 2.4
Indicador compuesto: atención de calidad en pacientes en infarto de miocardio: (casos evaluados: 15)

| Subindicadores | Numerador (Cumplimientos) | Denominador (Casos en los que se ha evaluado) | Cumplimiento |
|---|------------------------------|--|--------------|
| Aspirina dentro de las 1 ^{as} 24 h. | 10 | 12 | 0.83 |
| β-bloqueantes dentro de las 1 ^{as} 24 h. | 4 | 6 | 0.66 |
| Aspirina al alta | 4 | 5 | 0.80 |
| β-bloqueantes al alta | 2 | 4 | 0.50 |
| Consejo antitabaco | 1 | 3 | 0.33 |
| Trombolíticos dentro de los primeros 30´ | 7 | 10 | 0.70 |
| Indicador compuesto (calidad de la atención) | 28 | 40 | 0.7 |

Adaptado de: Courtney J, Krumholz H, Wang Y, Turnbull B. Using composite measures for the public reporting of hospital performance data. ISQua Conference. Paris, octubre, 2002.

Los indicadores compuestos tipo porcentaje simple pueden medirse con muestras más pequeñas que los indicadores específicos o subindicadores que los componen.

De forma semejante, los resultados de las encuestas de calidad percibida pueden resumirse con las llamadas *tasas de problemas*, que en realidad son indicadores compuestos tipo porcentaje simple. Las *tasas de problemas* resumen las respuestas de varias preguntas sobre diversos aspectos de una variable compleja o dimensión indagada en la encuesta (por ejemplo problemas percibidos de organización, accesibilidad, competencia profesional, trato, etc.), expresándolo como porcentaje de problemas, según la fórmula:

$$\frac{\sum \text{respuestas que expresan existencia del problema}}{\sum \text{preguntas respondidas}} \times 100$$

En este acercamiento, los ítems individuales pueden servir como indicadores específicos de los diferentes aspectos del concepto, dimensión o variable más compleja que resume el indicador compuesto.

El principal, y quizás único, inconveniente de este tipo de indicadores es que su resultado pueda ser dominado por el subindicador mayormente evaluado (o la pregunta mayormente contestada, en el caso de las encuestas de calidad percibida); sin embargo, este sesgo potencial podría verse como deseable hasta cierto punto, puesto que refleja con mayor fidelidad

la realidad: lo más frecuente está debidamente más representado.

3.2. Indicadores compuestos con reconversión de la escala de medición de los subindicadores

Uno de los problemas al agregar varios subindicadores en uno solo para hacer comparaciones es que se de alguna de las siguientes circunstancias: (i) no tienen la misma unidad de medida; (ii) tienen diferente rango posible de valores máximo y mínimo; (iii) tienen diferentes niveles de precisión y dispersión de los datos, en función de haber sido medidos sobre tamaños de muestra (o características de la misma) también diferentes. Todos estos problemas van a restarle fiabilidad al indicador compuesto hasta el punto de hacerlo inconsistente y poco útil para comparaciones. Cuando se detecta o se piensa que puede ocurrir alguna de estas circunstancias, o simplemente para curarse en salud por si llegara a ocurrir, se realiza una reconversión de las escalas de medida de los subindicadores, para normalizarlos y hacerlos más comparables.

El método más común de reconversión es dividir los valores observados en cada subindicador por alguna medida que represente la dispersión de los datos, como puede ser el rango o, más frecuentemente, la desviación estándar del valor

Los indicadores compuestos tipo porcentaje simple se utilizan también en la forma llamada *tasa de problemas*, para resumir los resultados de las encuestas de calidad percibida.

Los indicadores compuestos tipo porcentaje simple priman a los subindicadores de posible evaluación más frecuente.

promedio (error estándar de la medición). Por ejemplo, para la construcción de los indicadores compuestos que resumen los datos del CAHPS y HEDIS (cuadro 2.1), los ítems del CAHPS tenían escalas de respuesta diferentes (0-10, 1-4, 0-1, 1-3) y los indicadores del HEDIS diferencias también apreciables en su varianza; la solución adoptada fue normalizar todos los subindicadores, antes de su agrupación, dividiéndolos por la desviación estándar de su respectiva media.

En las últimas ediciones del listado Top 20 de los hospitales españoles, también se introduce en cierta forma este concepto al normalizar el ranking de los hospitales en cada uno de los seis indicadores por la distancia del valor observado a los valores de los otros hospitales de la lista; estos valores normalizados son los que se suman para obtener el indicador (ranking) compuesto final.

En ocasiones se realiza un ajuste en relación con un estándar de referencia. Por ejemplo, en el caso de los indicadores de alto nivel del NHS británico, el indicador compuesto es la media de las tasas de los subindicadores estandarizados en relación con la tasa promedio de Inglaterra; así el

indicador *cobertura de detección precoz de cáncer en la mujer* que incluye los subindicadores referentes a las coberturas para cáncer de mama (tasa 1) y de cáncer de cérvix (tasa 2) se calcula de la siguiente manera:

Indicador cobertura de detección precoz de cáncer en la mujer =

$$\frac{100}{\text{Tasa 1 en Inglaterra}} \times \text{Tasa 1 en área} + \frac{100}{\text{Tasa 2 en Inglaterra}} \times \text{Tasa 2 en área} / 2$$

El cuadro 2.5 proporciona ejemplos de indicadores compuestos, utilizados en el NHS británico. En todos ellos se realiza el cálculo utilizando este tipo de estandarización.

3.3. Indicadores compuestos con ponderación de los subindicadores

La cuestión de la normalización de las escalas de los subindicadores puede considerarse, aunque sea simplificando mucho el argumento, una cuestión de razonamiento estadístico; es decir,

En determinadas circunstancias, es necesario realizar una reconversión de las escalas de medida de los subindicadores, con el fin de garantizar la fiabilidad de las comparaciones con base en los indicadores compuestos en que se integran.

El método de normalización de escalas más común es dividir los valores observados por alguna medida de la dispersión de los datos.

Cuadro 2.5

Ejemplos de indicadores compuestos: indicadores de alto nivel del NHS británico

- Manejo de patologías de hospitalización evitable: tasa de admisión estandarizada por causa de
 - Infecciones renales y del tracto urinario.
 - Fallo cardiaco
 - Infección ORL severa
- Manejo de patologías crónicas: tasa de admisión estandarizada por causa de
 - Asma
 - Diabetes
 - Epilepsia
- Atención a la salud mental en atención primaria:
 - Consumo de benzodiazepinas
 - Ratio de antidepresivos/ benzodiazepinas
- Prescripción coste/ efectiva:
 - Coste/prescripción de asociaciones
 - Coste/prescripción de fármacos modificados
 - Coste/prescripción de fármacos de valor clínico limitado
 - Coste/DDD de corticosteroides inhalados
 - Ratio de antidepresivos/benzodiazepinas
- Alta hospitalaria:
 - Tasa de altas <56 días desde admisión por ACV
 - Tasa de altas <.56 días desde admisión por fractura de cuello de fémur

En los indicadores compuestos de alto nivel del NHS británico se normalizan los valores de los subindicadores dividiéndolos por la tasa de Inglaterra, que se considera como estándar.

La ponderación de los subindicadores y el método empleado para ello son siempre objeto de discusión, al implicar generalmente cierta valoración subjetiva.

No existe ningún procedimiento universalmente aceptado como el mejor para ponderar los subindicadores de un indicador compuesto.

La opinión de expertos o de la población ha sido utilizada con cierta frecuencia para ponderar los subindicadores y goza de un cierto predicamento sobre la base de su, al menos aparente, legitimidad.

La utilización de técnicas estadísticas como el análisis factorial es considerada por algunos autores como la forma más adecuada y aséptica para ponderar.

valorar si existe alguna de las circunstancias que lo haría necesario (descritas en el apartado anterior) y encontrar la fórmula de reconversión de los datos que mejor se adecue al indicador compuesto de que se trate. Sin embargo, plantearse la necesidad de ponderación y decidir qué método utilizar para ponderar, raramente va a estar exento de una discusión que va mucho más allá de la fórmula matemática empleada. Para empezar, así como no hay ninguna forma inequívoca de decidir qué y cuántos subindicadores incluir en un indicador compuesto, a la hora de ponderarlos tampoco existe un procedimiento universalmente aceptado. Todos ellos implican algún tipo de juicio de valor. Revisemos brevemente los comúnmente empleados.

La utilización de la opinión de expertos o de la población en general tiene un cierto predicamento sobre la base de su, al menos aparente, legitimidad. La opinión de expertos ha sido utilizada, por ejemplo, para ponderar las seis categorías de subindicadores en el caso de Canadá (cuadro 2.1), y el resultado de encuestas a la población ha sido utilizado en el caso de los subindicadores considerados por el King's Fund, y de los del indicador compuesto de la OMS, aunque en este último caso la encuesta se dirigió a individuos supuestamente informados y conocedores del tema (la mayoría de ellos empleados de la propia OMS). Un proceso semejante de consulta a supuestos conocedores del tema lo realizó el extinto Insalud para establecer la ponderación, que llamaron *valor técnico*, de las coberturas de los diversos servicios de la cartera de servicios de atención primaria, que eran sumadas junto con el cumplimiento de las normas técnicas mínimas, para calcular el *producto* de los centros de salud.

La forma más aséptica de ponderar, no por ello menos discutida, es la utilización de técnicas estadísticas (análisis de componentes principales, análisis factorial, regresión múltiple, etc.). Es el método utilizado, por ejemplo, para decidir qué agregar y qué peso se les dará a los 23 subindicadores CAHPS y HEDIS (cuadro 2.1), a los que se sometió un análisis factorial para identificar los cuatro indicadores compuestos que finalmente se calculan. Conviene señalar que cinco de los nueve subindicadores del CAHPS son, a su vez, indicadores compuestos tipo porcentaje simple (tasa de problemas) procedentes de una encuesta

de calidad percibida, asumidos con toda normalidad como subindicadores simples (aunque son en realidad compuestos) y válidos.

Dos formas de ponderación y agregación de subindicadores que merecen considerarse, pero aun escasamente utilizarse en los servicios de salud, son las llamadas de distancia a la frontera de eficiencia y distancia a valores estándar. Ambos métodos han sido utilizados en el ámbito socioeconómico, pero aún no se les ha prestado suficiente atención en la gestión de la calidad de los sistemas de salud, a pesar de que implican conceptos muy arraigados en este campo, como son la referencia a valores *benchmark* (frontera de eficiencia), y a valores deseables u óptimos de calidad (distancia a valores estándar). Un ejemplo sería el *Índice de desarrollo humano* que utiliza la ONU para comparar países. En este indicador compuesto, los valores de los subindicadores (esperanza de vida, nivel de analfabetismo y escolarización, GDP per cápita) se expresan como decimales de la unidad que representan su distancia al valor 1, que se les adjudica a las metas que deben alcanzarse en estos subindicadores (esperanza de vida de 85 años, 100% de escolarización y ausencia de analfabetismo, y GDP per cápita de 40 000 dólares).

Otro ejemplo, utilizado en la actualidad por la AHRQ (*Methods Used in the National Healthcare Reports*, <http://nhqrnet.ahrq.gov/inhqrdr/resources/methods>), es la comparación estatal de cada indicador, con el benchmark establecido por la media de 10% de los mejores estados (se establece 10% porque no todos reportan los 300 indicadores posibles), que se encuentran resumidos en un gráfico con una sola barra horizontal, que muestra cuántos indicadores alcanzan o superan el benchmark y cuántos están cerca y lejos del benchmark. De forma aun más sintética, se establece un score por estado de 0 a 100, con base en la siguiente fórmula que pondera cada indicador según esté por encima o debajo de la media nacional:

$$((Ax1)+(Bx0.5)+(Cx0))x100 / (A+B+C)$$

Donde:

A: número de indicadores por encima de la media nacional.

B: número de indicadores en la media nacional.

C: número de indicadores por debajo de la media nacional.

Cuadro 2.6
Atributos deseables de los indicadores compuestos

Los métodos de la *distancia a la frontera de eficiencia* y de la *distancia a valores estándar* tienen una gran validez facial para la gestión de la calidad.

- Resumen un conjunto de indicadores/criterios teóricamente relacionados.
- Son fácilmente interpretables (alta validez facial).
- Describen la mayoría de la variabilidad entre las unidades evaluadas (profesional, centro, área...).
- Se han construido teniendo en cuenta de forma explícita y razonada:
 1. Si hay necesidad de normalización de las escalas de medida.
 2. Si hay necesidad de ponderación.
 3. La adecuación de los métodos eventualmente utilizados para los cambios de escala, ponderación y agregación de los subindicadores.

Tanto la *distancia a la frontera de eficiencia* como la *distancia a valores óptimos o benchmark*, tal como resume el método empleado por la AHRQ, pueden ser especialmente útiles para comparar centros, áreas, regiones o sistemas, dentro de una dinámica de mejora continua.

4. Atributos deseables de los indicadores compuestos

Como resumen de lo hasta aquí expuesto en relación con los indicadores compuestos, y atendiendo a lo que se sabe en este momento, podemos subrayar una serie de atributos o características que debemos valorar y exigir en este tipo de herramientas (cuadro 2.6).

Los dos primeros atributos listados en el cuadro 2.6 (correcta selección de subindicadores y facilidad de interpretación) están, inicialmente, sujetos a una valoración relativamente subjetiva. Sin embargo, el resto de los atributos descritos deben valorarse más objetivamente. La forma en que capturan la variabilidad entre las unidades a comparar en relación con lo que pretende medir debe ser sujeto a pilotaje y validación empírica. Asimismo, deben buscarse y explicitarse las razones para la manipulación estadística (normalización, ponderación, etc.) a que se someten los subindicadores, o, en su caso, las razones de su ausencia. El anexo 2.1 resume los pasos para analizarlos críticamente.

BIBLIOGRAFÍA

- Agency for Healthcare Research and Quality (AHRQ). Methods used in the National Healthcare Reports [documento en internet]. 2014 [consultado en febrero de 2015]. Disponible en: <http://nhqrnet.ahrq.gov/inhqrd/r/resources/methods>
- García-Eroles L, Illa C, Arias A, Casas M. Los Top 20, 2000: Objetivos, ventajas y limitaciones del método. Rev Calidad Asistencial 2001;16:107-116.
- National Health Service. Quality and Performance in the NHS; High level performance indicators and clinical indicators. London: NHS Executive, 1999.
- Smith P. Composite indicators of health system performance. En: Smith P (ed.). Measuring up: Improving Health systems performance in OECD Countries [documento en internet]. Paris: OECD, 2002 [consultado en febrero de 2015]. Disponible en: <http://www1.oecd.org/els/health/canconf-fr/smith.pdf>
- State-of-the art Report on Current Methodologies and Practices for Composite Indicator Development. Documento-base preparado por el Applied Statistics Group del JRC de la Comisión Europea y la OCDE para el First workshop on Composite Indicators of Country Performance [document en internet]. Ispra (Italia), mayo de 2003 [consultado en febrero de 2015]. Disponible en: http://webfarm.jrc.cec.eu.int/uasa/events/oecd_12may03/
- Zaslavsky AM, Shaul JA, Zaborski LB, Cioffi MJ, Cleary PD. Combining health plan performance indicators into simpler composite measures. Health Care Finac R, 2002;23(4):101-115.

Anexo 2.1

Construcción y análisis de un indicador compuesto

■ Parte I: Guía para la construcción o análisis de un indicador compuesto

Paso 1: Construcción o selección del indicador

Para este primer paso hay dos opciones alternativas:

- a) Construir un indicador compuesto, por ejemplo, utilizando los criterios empleados en un ciclo de mejora.
- b) Seleccionar un indicador compuesto que se utilice en su entorno de trabajo.

Paso 2: Describir el tipo y método de construcción de un indicador compuesto

Indicar para el indicador compuesto identificado en el paso 1, de qué tipo de indicador se trata y cuál es su procedimiento de construcción, siguiendo el esquema clasificatorio del cuadro 2.2 de este capítulo.

Paso 3: Análisis crítico de un indicador compuesto

Valorar de forma explícita y ordenada la presencia en el indicador compuesto seleccionado de los atributos deseables que se indican en la tabla 16 de este capítulo.

■ Parte II: Guión de puntos por explicitar

Paso 1:

- Indicador construido o
- Indicador seleccionado (ya existente).

Paso 2:

- Descripción del método de construcción.
- Selección de subindicadores.
- Reconversión de escala de medición (en su caso).
- Ponderación de subindicadores (en su caso).
- Forma de agregación de los subindicadores.
- Tipo de indicador compuesto (según listado del cuadro 2.3).

Paso 3:

- Revisión de los atributos deseables listados en el cuadro 2.6.
- Relación entre los subindicadores y de éstos con el problema o aspecto a medir.
- Facilidad de interpretación (valoración de la validez facial).
- Pilotaje empírico (si se ha construido con los criterios del ciclo de mejora pueden compararse los resultados desagregados y agrupados en el indicador compuesto antes y después de la intervención).
- Análisis pormenorizado y razonado de la presencia o ausencia de normalización y ponderación, y los métodos utilizados.

Capítulo 3

El muestreo para la aceptación de lotes (LQAS) como método de monitorización

Contenido general y objetivos específicos

1. Contextualización
2. Planes de mediciones esporádicas o con grandes intervalos de tiempo: la *evaluación rápida* de la calidad
3. Monitorización sistemática, utilizando la distribución binomial
4. Contraste de hipótesis *completo*, utilizando la distribución binomial
5. El muestreo de aceptación de lotes como método de monitorización en los servicios de salud
6. Índices de calidad, riesgos y procedimientos para la aplicación del LQAS
7. LQAS en servicios de salud: dos ejemplos de aplicación en la práctica
8. Esquemas avanzados de muestreo de aceptación de lotes

Anexo 3.1. Procedimiento para la monitorización de indicadores utilizando LQAS

Contenido general

Después de aclarar el concepto, las situaciones en que es útil la monitorización y sus objetivos, así como la importancia y la forma de seleccionar buenos indicadores considerando la conveniencia o inconveniencia de utilizar indicadores compuestos, en este capítulo se abordan las bases teóricas y la aplicación práctica del LQAS. Comenzamos por revisar los fundamentos y aplicación de la distribución binomial en la *evaluación rápida* de la calidad con muestras pequeñas, para después, sobre esta base, entender y practicar el LQAS en planes de monitorización con intervalo largo entre mediciones.

Objetivos específicos

1. Entender el fundamento estadístico de la aplicación de la distribución binomial en la *evaluación rápida* de problemas de calidad.
2. Interpretar, en términos de probabilidad estadística, el resultado de una muestra pequeña (<30 casos) sobre un criterio o indicador dicotómico.
3. Distinguir en la práctica los conceptos de errores α y β en el contraste de hipótesis.
4. Entender el fundamento estadístico del LQAS como método de medición.
5. Diseñar planes de monitorización que incluyen el LQAS como método de medición.
6. Manejar las tablas LQAS de acuerdo con condiciones de referencia (estándar, umbral, errores α y β) prefijadas.
7. Conocer diversos esquemas y variaciones del LQAS, según coste, métodos de muestreo y tipo de probabilidad aplicados.

El muestreo para la aceptación de lotes (LQAS) como método de monitorización

1. Contextualización

Según el concepto abordado en el capítulo 1, entendemos por monitorización la medición sistemática, repetida y planificada de indicadores de calidad; una actividad conducente a controlar que estamos a unos niveles preestablecidos y que tiene como objetivo identificar la existencia o inexistencia de situaciones problemáticas por evaluar o sobre las que haya que intervenir. Los indicadores por monitorizar pueden derivarse de ciclos de mejora, de actividades de diseño de los servicios, o ser fruto de una selección de aspectos o servicios relevantes de nuestro centro, cuya calidad nos interesa controlar. En cualquier caso, deben ir acompañados de un plan de monitorización que incluya periodicidad, mecanismos para la recolección de datos y método de interpretación de los mismos.

El interés que podamos tener en una estimación del nivel de calidad existente y la periodicidad de medición son decisiones de importancia para la elección de los métodos para monitorizar. El muestreo de aceptación de lotes (LQAS) y el control estadístico (gráfico) de la calidad son métodos originados y desarrollados en la industria, que pueden resultar de gran utilidad para la monitorización en servicios de salud, si renunciamos a tener una estimación del nivel de cumplimiento de los indicadores y nos centramos en el conocimiento de la existencia o inexistencia (aceptación/rechazo) de unos niveles de cumplimiento preestablecidos. El control estadístico de la calidad precisa de mediciones frecuentes (mínimo mensuales) para ser útil, mientras que el LQAS y otros métodos de *evaluación rápida* basados en la distribución binomial pueden ser aplicados a planes de mediciones más espaciados. La prácti-

ca de monitorización que utiliza LQAS, aún poco extendida en los servicios de salud, es, sin embargo, muy sencilla; sólo precisa entender y manejar unas tablas ya construidas para fijar el número máximo de incumplimientos del indicador aceptable en una muestra determinada (y normalmente pequeña) para aceptar, a su vez, un determinado nivel de cumplimiento preestablecido, con unos niveles de error estadístico previamente acordados. Este capítulo ofrece los fundamentos estadísticos y el detalle de las implicaciones prácticas del uso del LQAS en servicios de salud.

2. Planes de mediciones esporádicas o con grandes intervalos de tiempo: la *evaluación rápida* de la calidad

Cuando la variable que mide el indicador seleccionado es de tipo dicotómico (presencia/ausencia de una determinada calidad), como ocurre en la mayoría de los indicadores de calidad que se construyen en los servicios de salud, la utilización de las probabilidades de la distribución binomial nos proporciona una forma rápida y eficiente de decidir si estamos o no ante una situación problemática, por medio de un tamaño de muestra relativamente pequeño (máximo 30 casos). Veamos cómo funciona y cuáles son las circunstancias en que es aplicable, partiendo de dos supuestos prácticos.

Supongamos que, refiriéndonos a dos problemas de salud diferentes, nuestro estándar histórico de cumplimiento de toma de tensión arterial para la detección de problemas de hipertensión se sitúa en 85%, y que hemos logrado reducir a 15% la ocurrencia de prescripción de antibióticos en casos de resfriado común. Sin embargo, como nos

En este capítulo se ofrecen los fundamentos teóricos y la aplicación práctica de la distribución binomial y el LQAS, para la *evaluación rápida* de problemas y la monitorización de indicadores en servicios de salud.

Las tablas de distribución binomial pueden utilizarse para decidir de una forma rápida y con una muestra pequeña si existe o no problema de calidad en relación con un estándar preestablecido, y siempre que el indicador se mida con una variable dicotómica (cumple/no cumple).

El tamaño de la muestra ha de ser inferior a 10% del tamaño del marco muestral.

En la que llamamos *evaluación rápida*, interpretamos el resultado de la muestra en términos de la probabilidad de que éste pueda darse, si fuera cierto que se cumple el estándar prefijado en la población o marco muestral.

La probabilidad del resultado encontrado se busca en las tablas correspondientes de la distribución binomial para el tamaño de muestra empleado.

hemos despreocupado de su monitoreo durante varios años, en la actualidad, se percibe un estancamiento en el número de pacientes clasificados como hipertensos y desde hace tiempo no comprobamos si mantenemos el nivel de prescripción inadecuada de antibióticos en el nivel que tanto nos costó alcanzar, por lo que sospechamos que no hemos realizado las actividades de detección como es debido. En ambos casos, es necesario saber si la situación es o no problemática para, en su caso, tomar acciones para solucionarla. Una opción sería realizar un muestreo aleatorio suficientemente grande de las visitas en los últimos seis meses, tanto de pacientes en general (para evaluar la detección de hipertensos), como de los que fueron diagnosticados de resfriado común (para evaluar la prescripción de antibióticos), y estimar en qué nivel nos encontramos.

Sin embargo, ambas situaciones cumplen los requisitos en los que es aplicable la distribución binomial, a saber:

- Lo que se evalúa sólo tiene dos valores posibles (cumple/no cumple)
- Cada una de las unidades de estudio es independiente de la otra.
- Evaluamos en una muestra de “n” casos.
- La probabilidad general de encontrar un valor u otro (cumple/no cumple) es la misma para cada uno de los casos.

Esta última condición se cumple siempre que el tamaño de la muestra, “n”, sea inferior a 10% del universo o marco muestral de donde la hemos extraído (en nuestro caso, el total de visitas en los últimos seis meses y el total de casos de resfriado común, respectivamente). Si “n” es mayor que 10% del marco muestral, habría que hacer un muestreo especial (el llamado muestreo con reposición) o utilizar la distribución hipergeométrica, de la cual no nos vamos a ocupar en este capítulo.

Si es aplicable la distribución binomial, vamos a extraer una muestra aleatoria, pero pequeña, de 15 casos, por ejemplo, e interpretar el resultado que obtengamos en términos de la probabilidad de que sea compatible con la existencia real del estándar prefijado (85% de cumplimiento para la detección de hipertensión, 15% para la prescripción de antibióticos), sin que nos interese conocer el nivel preciso de cumplimiento existente; sólo queremos saber si podemos aceptar o no que estamos en los niveles deseados.

Evaluamos nuestras dos muestras de 15 casos y encontramos que el número de casos de incumplimiento es de cinco para la hipertensión, y que se han prescrito antibióticos en cuatro de los casos de resfriado común. ¿Qué podemos deducir de estos resultados? Obviamente no es apropiado estimar el nivel de cumplimiento con una muestra tan pequeña, pero lo que sí podemos hacer es averiguar cuál es la probabilidad de que esos resultados sean compatibles con la existencia del estándar prefijado en la población de la que hemos extraído la muestra. Si esta probabilidad es baja, concluiremos que este resultado es incompatible con el estándar prefijado y que, por lo tanto, rechazamos que este estándar se cumpla en la población de la que hemos extraído la muestra. Según el consenso más extendido en las pruebas de significación estadística, podemos considerar que una probabilidad es baja si es <5 por ciento.

Para hallar la probabilidad que nos interesa (en nuestro caso, la que tiene el encontrar cinco incumplimientos de toma de tensión en una muestra de 15, si el estándar es de 85% de cumplimiento, y cuatro casos de prescripción de antibióticos en una muestra de 15, si el estándar es de 15%), podemos utilizar la fórmula de probabilidades de la distribución binomial (cuyo cálculo adelantamos que no es preciso efectuar porque existen las tablas oportunas):

$$P(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}$$

en donde:

$P(x)$ es la probabilidad de encontrar x casos de cumplimiento (o incumplimiento, según hayamos definido el estándar) en la muestra.

n es el tamaño de la muestra

x son los casos de cumplimiento (o incumplimiento) que hemos encontrado en la muestra

π es la probabilidad de cumplimiento (o incumplimiento) prefijada (estándar), que supuestamente existe en la población de la que extraemos la muestra.

$\binom{n}{x}$ es una expresión combinatoria que equivale a:

$$\frac{n!}{x!(n-x)!}$$

siendo la anotación “!” indicativa de una multiplicación factorial.

Es importante señalar que los valores de π y x han de ir en el mismo sentido; es decir, si el estándar es de cumplimiento (como en nuestro ejemplo de la hipertensión) x ha de ser número de cumplimientos; pero si el estándar es negativo o de incumplimiento, x debe ser incumplimiento. En nuestros dos ejemplos, n es igual a 15; π es igual a 85% para el caso de la hipertensión y 15% para la prescripción de antibióticos; x sería 10 para la hipertensión (hemos encontrado cinco incumplimientos, pero el estándar lo tenemos en positivo), y cuatro para la prescripción de antibióticos en el resfriado común.

Para averiguar $P_{(x)}$ (la probabilidad de haber encontrado precisamente el número de cumplimientos que hemos encontrado, si el estándar se cumple en la población de donde hemos extraído la muestra) podríamos utilizar la fórmula descrita que, aunque parezca complicada, es fácil de programar en cualquier hoja de cálculo; sin embargo, es más cómodo utilizar los cuadros de la distribución binomial existentes en los libros de estadística. Hay un cuadro de probabilidades para cada tamaño de muestra (normalmente hasta $n=20$, o como máximo $n=30$).

Estos cuadros dan los cálculos de probabilidad que nos interesan para cada uno de los posibles resultados en un determinado tamaño de muestra. En consecuencia, lo que tenemos que hacer en primer lugar es localizar el cuadro que corresponde al tamaño de muestra que hemos utilizado.

En nuestro caso ha sido 15, y el cuadro correspondiente a $n=15$ es el que reproducimos como cuadro 3.1. En él podemos ver que las columnas están encabezadas por unas cifras que corresponden a los diversos valores (de 0.05 a 0.50) que podemos hacer corresponder con el estándar prefijado. Para localizar las probabilidades que nos interesan en nuestros ejemplos, nos situaremos en la columna de π (estándar) igual a 0.15 (15%), tanto para el caso de la prescripción de antibióticos (que ya está expresado así: el indicador busca la no prescripción de antibióticos), como en el caso de la hipertensión (un estándar de cumplimiento de 85% equivale a un estándar de incumplimiento de 15 por ciento).

A continuación, buscamos en la columna de “ x ” el número de incumplimientos encontrados: cinco para la hipertensión y cuatro para la pres-

Normalmente se rechaza que sea cierto el cumplimiento del estándar de referencia, si la probabilidad del resultado obtenido es <0.05 .

Para buscar en los cuadros, hay que localizar primero la correspondiente al tamaño de la muestra, y después mirar en la columna del estándar correspondiente.

Cuadro 3.1.
Probabilidades binomiales (para $n=15$)

| x | Estándar de incumplimientos (π) | | | | | | | | | |
|-----|---------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | .05 | .10 | .15 | .20 | .25 | .30 | .35 | .40 | .45 | .50 |
| 0 | .4633 | .2059 | .0874 | .0352 | .0134 | .0047 | .0016 | .0005 | .0001 | .0000 |
| 1 | .3658 | .3432 | .2312 | .1319 | .0668 | .0305 | .0126 | .0047 | .0016 | .0005 |
| 2 | .1348 | .2669 | .2856 | .2309 | .1559 | .0916 | .0476 | .0219 | .0090 | .0032 |
| 3 | .0307 | .1285 | .2184 | .2501 | .2252 | .1700 | .1110 | .0634 | .0318 | .0139 |
| 4 | .0049 | .0428 | .1156 | .1876 | .2252 | .2186 | .1792 | .1268 | .0780 | .0417 |
| 5 | .0006 | .0105 | .0449 | .1032 | .1651 | .2061 | .2123 | .1859 | .1404 | .0916 |
| 6 | .0000 | .0019 | .0132 | .0430 | .0917 | .1472 | .1906 | .2066 | .1914 | .1527 |
| 7 | .0000 | .0003 | .0030 | .0138 | .0393 | .0811 | .1319 | .1771 | .2013 | .1964 |
| 8 | .0000 | .0000 | .0005 | .0035 | .0131 | .0348 | .0710 | .1181 | .1647 | .1964 |
| 9 | .0000 | .0000 | .0001 | .0007 | .0034 | .0116 | .0298 | .0612 | .1048 | .1527 |
| 10 | .0000 | .0000 | .0000 | .0001 | .0007 | .0030 | .0096 | .0245 | .0515 | .0916 |
| 11 | .0000 | .0000 | .0000 | .0000 | .0001 | .0006 | .0024 | .0074 | .0191 | .0417 |
| 12 | .0000 | .0000 | .0000 | .0000 | .0000 | .0001 | .0004 | .0016 | .0052 | .0139 |
| 13 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0001 | .0003 | .0010 | .0032 |
| 14 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0001 | .0005 |
| 15 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 |

x = nº de incumplimientos encontrados en la muestra de 15 casos.
Fuente: Adaptado de Rosner, 2011.

La distribución binomial generalmente no tiene utilidad práctica para valorar muestras de más de 30 casos.

cripción de antibióticos. El cruce de la fila de este número con la columna de 0.15 expresa la probabilidad de encontrar precisamente ese número de incumplimientos si el estándar se cumple. En el cuadro 3.1 podemos ver que estas probabilidades son 0.0449 (4.49%), para el caso de la hipertensión, y 0.1156 (11.56%) para el caso de los antibióticos. ¿Cómo interpretarlo? Ateniéndonos a la norma para decidir establecida (rechazar que sean ciertas las situaciones de cumplimiento del estándar, si el resultado de la muestra tiene una probabilidad <5%), identificamos como problemática la situación de la detección de hipertensos, mientras que aceptamos como no problemática (en relación con nuestro estándar) la prescripción de antibióticos en el resfriado común.

Éste es un proceder sencillo, que sólo precisa de:

1. Establecimiento de unos parámetros de referencia:
 - ▶ Estándar del cual queremos verificar si se cumple o no (p) para el indicador por comprobar.
 - ▶ Un tamaño de muestra (n) que nos resulte manejable para tomar una decisión rápida.
2. Extracción de la muestra de forma aleatoria.
3. Buscar en los cuadros de la distribución binomial la probabilidad de encontrar el resultado que hayamos obtenido en la muestra, en la columna correspondiente al valor del estándar cuyo cumplimiento queremos comprobar. Si esta probabilidad es baja, clasificamos la situación como problemática y la investigamos más a fondo.

Es muy importante asegurarse de entender bien los diversos parámetros que utilizamos porque los cuadros de la distribución binomial que contienen los diversos libros de estadística no utilizan siempre los mismos símbolos. Por ejemplo, “x” puede aparecer como “a”, “c”, “k” o “d”. Por otra parte, las probabilidades a veces son exactas (como en el cuadro 3.1) o acumuladas; p puede estar en positivo (cumplimientos) o negativo (incumplimientos). La estructura del cuadro es, sin embargo, siempre la misma.

Este proceder de *evaluación rápida* es muy sencillo y práctico para interpretar muestras pequeñas para decidir la aceptación o rechazo sobre la existencia de un determinado estándar. Adviértase que el rechazo puede indicar identificación de situación problemática o que es mejor que el estándar. Véase, por ejemplo, en el cuadro 3.1, que para un estándar de 30%, es prácticamente igual de improbable encontrar uno u ocho incumplimientos en una muestra de 15 casos, y con ambos resultados rechazaremos que exista un incumplimiento de 30%; sin embargo, en el primer caso (un incumplimiento), lo más probable es que en realidad la tasa sea menor, y en el segundo (ocho), que sea mayor que el estándar de 30% de incumplimiento utilizado como referencia.

Sin embargo, a pesar de su sencillez, es necesario conocer tanto las utilidades como las limitaciones de la *evaluación rápida*. Estas últimas son su falta de adecuación para monitorizaciones sistemáticas y su parcialidad como método de contraste de hipótesis (cuadro 3.2).

La que hemos llamado *evaluación rápida* sólo puede utilizarse para monitorizaciones esporádicas (no sistemáticas) y sólo tiene en cuenta uno de los dos posibles errores que pueden darse en el contraste de hipótesis.

Cuadro 3.2

Utilidad y limitaciones de la *evaluación rápida* binomial como método de identificación de problemas con indicadores tipo tasa (proporción o porcentaje)

| Utilidad | Limitaciones |
|---|--|
| Monitorización esporádica (no sistemática) de indicadores | Inadecuada para monitorización sistemática, aunque puede ser adaptada para ello. |
| Interpretación de resultados de muestras pequeñas (<30 casos), en relación con un estándar de referencia. | Es incompleta como método de contraste de hipótesis: la decisión sólo tiene en cuenta la probabilidad puntual de falsos positivos (identificar como problemática una situación que no lo es) o de falsos negativos (aceptar como buena una situación problemática), pero no ambos. |

3. Monitorización sistemática, utilizando la distribución binomial

Los resultados de una evaluación puntual, esporádica, pueden ser interpretados en la forma que lo hemos hecho con las tablas de distribución binomial, pero no sería correcto mantener el mismo proceder como método de evaluación sistemática, repetida o rutinaria. En este caso, para tomar una decisión, lo correcto sería considerar, no la probabilidad del resultado concreto que hemos encontrado en la muestra puntual (tal como hemos hecho), sino la probabilidad de todos los resultados posibles simultáneamente (probabilidad acumulada) para el tamaño de muestra que hemos decidido utilizar. Asimismo, establecemos una probabilidad límite para aceptar o rechazar (que puede seguir siendo <0.05), pero su cálculo implica la suma de las probabilidades de que encontremos un determinado número de incumplimientos y cualquier número de incumplimientos inferior. La decisión no se tomaría entonces en función de la probabilidad para un resultado concreto (como hemos hecho en nuestros ejemplos), sino que buscaríamos un número límite de incumplimientos de tal manera que la probabilidad de encontrar ese número de incumplimientos o

más sea <0.05 . Sólo aceptaríamos como buenos los resultados con número de incumplimiento por debajo de ese límite. Si en vez de incumplimientos quisiéramos valorar cumplimientos, el razonamiento sería semejante aunque, naturalmente, en sentido opuesto: buscaríamos un número mínimo de cumplimientos de forma que la probabilidad de encontrar ese número o más fuese >0.95 ; así identificaríamos como problemático todo resultado que tuviera un número de cumplimientos inferior al que hemos señalado como límite para la decisión. La decisión se toma, como se ve, en función de la probabilidad acumulada de los diversos resultados posibles. Para ello debemos utilizar los cuadros de la distribución binomial con probabilidades acumuladas, tal como el que se reproduce también para una muestra de 15 casos, en el cuadro 3.3.

En el cuadro 3.3, las cifras de probabilidades que aparecen son la suma de la probabilidad de cada una de las posibles x , y de todas las anteriores. Por ejemplo, si nos colocamos en la columna que corresponde al estándar de incumplimiento de 15% ($\pi=0.15$), la probabilidad que encontramos para $x=4$ es 0.9383 (o 93.83%) y corresponde a la suma de la probabilidad de que obtengamos uno, dos, tres, cuatro o ningún incumplimientos

Para monitorizar de forma sistemática utilizando la distribución binomial, deben utilizarse los cuadros de probabilidad acumulada.

Cuadro 3.3
Probabilidades binomiales acumuladas (para $n=15$)

| x | Estándar de incumplimientos (π) | | | | | | | | | |
|----|---------------------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--|
| | .05 | .10 | .15 | .20 | .25 | .30 | .35 | .40 | .50 | |
| 0 | .4633 | .2059 | .0874 | .0352 | .0134 | .0047 | .0016 | .0005 | .0001 | |
| 1 | .8290 | .5490 | .3186 | .1671 | .0802 | .0353 | .0142 | .0052 | .0017 | |
| 2 | .9638 | .8159 | .6042 | .3980 | .2361 | .1268 | .0617 | .0271 | .0107 | |
| 3 | .9945 | .9444 | .8227 | .6482 | .4613 | .2969 | .1727 | .0905 | .0424 | |
| 4 | .9994 | .9873 | .9383 | .8358 | .6865 | .5155 | .3519 | .2173 | .1204 | |
| 5 | .9999 | .9978 | .9832 | .9389 | .8516 | .7216 | .5643 | .4032 | .2608 | |
| 6 | 1.0000 | .9997 | .9964 | .9819 | .9434 | .8689 | .7548 | .6098 | .4522 | |
| 7 | 1.0000 | 1.0000 | .9994 | .9959 | .9827 | .9500 | .8868 | .7869 | .6535 | |
| 8 | 1.0000 | 1.0000 | .9999 | .9992 | .9958 | .9848 | .9578 | .9050 | .8182 | |
| 9 | 1.0000 | 1.0000 | 1.0000 | .9999 | .9992 | .9963 | .9876 | .9662 | .9231 | |
| 10 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | .9999 | .9993 | .9972 | .9907 | .9745 | |
| 11 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | .9999 | .9995 | .9981 | .9937 | |
| 12 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | .9999 | .9997 | .9989 | |
| 13 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | .9999 | |
| 14 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | |
| 15 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | |

La decisión para aceptar/rechazar que el estándar se cumpla en la población o universo de donde se ha extraído la muestra, se basa en que la probabilidad acumulada de obtener el resultado sea <0.05 .

en la muestra; es decir, cuatro incumplimientos o menos, mientras que, en el cuadro 3.1, lo que veíamos era la probabilidad de obtener exactamente cuatro.

Para utilizar el procedimiento de *evaluación rápida* en la monitorización sistemática, utilizaríamos los cuadros de probabilidad acumulada, como el cuadro 3.1, buscando en el cuadro correspondiente al tamaño de muestra que hayamos decidido utilizar, el número de incumplimientos que represente una probabilidad acumulada >0.95 y sea lo más cercano posible a este valor. Este número sería el que utilizaríamos para decidir, sabiendo que tendríamos una posibilidad de equivocarnos al clasificar como problemática la situación <0.05 (probabilidad de que encontremos un número de incumplimientos mayor que el que utilizamos para clasificar la situación como problemática).

Para el caso de la prescripción de antibióticos ($p=0.15$), y con una muestra de 15 casos (cuadro 3.3), el número límite para decidir sería cinco (probabilidad acumulada=0.9832). Clasificaremos como aceptables los resultados de cinco incumplimientos o menos, e inaceptables (problemáticos) si encontramos seis o más; con ello

la probabilidad de falsos positivos (decir que hay problema cuando no lo hay) que estamos aceptando es de 0.0168 ($1-0.9832$), que es la probabilidad de obtener más si el estándar es de 15%. Si fijamos “cuatro” como número de decisión, la tasa de falsos positivos que arriesgamos con nuestro proceder sería $1-0.9383=0.062$ o 6.2% (0.9383 es la probabilidad de encontrar cuatro incumplimientos o menos en una muestra de 15 casos, si el real es 15 por ciento).

En la figura 3.1 se ha representado la curva de probabilidades que contiene el cuadro 3.1 para el estándar de 15%. En esta figura puede observarse la probabilidad de todos los valores de incumplimiento posibles en una muestra de 15 casos, y las consecuencias en cuanto a falsos positivos (estándar real de 15%, pero que clasificamos como problemático al situar en cinco el límite aceptable de incumplimientos). Aunque la representación como curva no es la más adecuada para las probabilidades binomiales, al no ser realmente continuas, sino “a saltos” por no haber probabilidad de valores intermedios entre un número de incumplimientos y el siguiente, resulta más fácil de entender gráficamente y por eso la utilizamos.

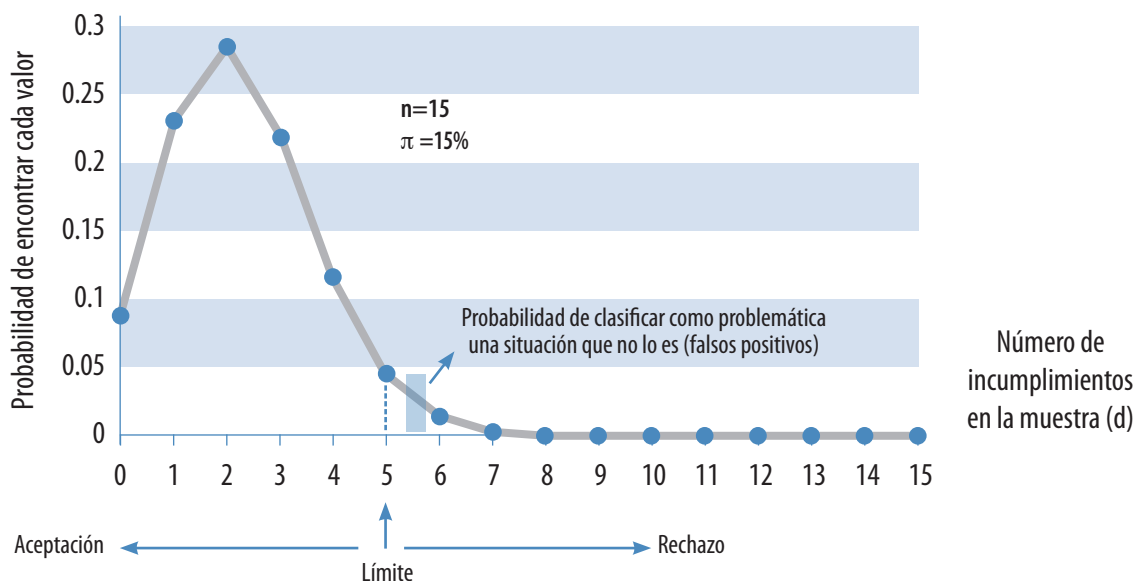


Figura 3.1. Curva de probabilidades de los diversos valores de incumplimiento que pueden encontrarse en una muestra de 15 casos, extraída de una población con un nivel de incumplimiento =15%

El cuadro 3.4 es semejante al 3.3 (probabilidades acumuladas), pero p y x son cumplimientos, en vez de incumplimientos.

En el ejemplo del indicador sobre captación de hipertensos, para el que hemos fijado un estándar de cumplimiento de 85% ($p=0.85$), y para su monitorización sistemática con una muestra de 15 casos, en el cuadro 3.4 buscamos el número límite de cumplimientos, cuya probabilidad acumulada sea $<5\%$, es decir, que la suma de las probabilidades de todos los valores de cumplimientos por debajo del número límite, considerados en su conjunto, sea >0.05 . Observando el cuadro 3.4, ese número límite para aceptar o rechazar que el estándar se cumple es 10 (probabilidad acumulada de encontrar nueve o menos = 0.0168), de manera que identificamos como problema siempre que el número de cumplimientos sea inferior a 10. Obsérvese que esta norma para monitorizar es la misma que hemos determinado para el caso de la prescripción de antibióticos, sólo que viendo cumplimientos, en vez de incumplimientos. No es de extrañar, porque un estándar de cumplimientos (en "positivo") de 85% es equivalente a uno de incumplimientos (en "negativo") de 15%, sólo que en un caso decidimos con base en un número mínimo de cumplimientos, y en el otro con base en un número máximo.

4. Contraste de hipótesis "completo", utilizando la distribución binomial

En los ejemplos expuestos, únicamente hemos tenido en cuenta uno de los dos posibles errores que podemos cometer con ella para tomar una decisión: controlamos y aceptamos una cierta probabilidad de clasificar como problemático algo que no lo es; pero con nuestra decisión, no sabemos hasta qué punto estamos aceptando como bueno algo que en realidad no cumple el estándar. Es como si en una prueba de screening sólo conociéramos su especificidad, pero no su sensibilidad o viceversa. En términos estadísticos, solamente manejamos uno de los dos errores posibles, α o β , tipo I o tipo II. En la industria, en términos de control de calidad, sería como decidir únicamente sobre la base del riesgo del productor: la probabilidad de que un buen producto se rechace al clasificarlo como malo; pero sin tener en cuenta el riesgo del consumidor: probabilidad de que un mal producto se acepte como bueno. El LQAS, basado generalmente en la distribución binomial, contempla explícitamente ambos riesgos o tipos de error, una vez que se define qué se entiende por "buen" producto (estándar de calidad del indicador) y "mal" producto (límite inaceptable o umbral del indicador que lo hace definitivamente problemático).

Para contrastar adecuadamente la hipótesis de si se cumple o no el estándar prefijado, deberíamos considerar no sólo la probabilidad de que rechacemos que se cumple cuando en realidad sí se cumple, sino también la de aceptar que se cumple cuando en realidad no.

Cuadro 3.4
Probabilidades binomiales acumuladas (para $n=15$)

| x | Estándar de cumplimientos (π) | | | | | | | | | |
|----|-------------------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | .50 | .55 | .60 | .65 | .70 | .75 | .80 | .85 | .90 | .95 |
| 0 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 |
| 1 | .0005 | .0001 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 |
| 2 | .0037 | .0011 | .0003 | .0001 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 |
| 3 | .0176 | .0063 | .0019 | .0005 | .0001 | .0000 | .0000 | .0000 | .0000 | .0000 |
| 4 | .0592 | .0255 | .0093 | .0028 | .0007 | .0001 | .0000 | .0000 | .0000 | .0000 |
| 5 | .1509 | .0769 | .0338 | .0124 | .0037 | .0008 | .0001 | .0000 | .0000 | .0000 |
| 6 | .3036 | .1818 | .0950 | .0422 | .0152 | .0042 | .0008 | .0001 | .0000 | .0000 |
| 7 | .5000 | .3465 | .2131 | .1132 | .0500 | .0173 | .0042 | .0006 | .0000 | .0000 |
| 8 | .6964 | .5478 | .3902 | .2452 | .1311 | .0566 | .0181 | .0036 | .0003 | .0000 |
| 9 | .8491 | .7392 | .5968 | .4357 | .2784 | .1484 | .0611 | .0168 | .0022 | .0001 |
| 10 | .9408 | .8796 | .7827 | .6481 | .4845 | .3135 | .1642 | .0617 | .0127 | .0006 |
| 11 | .9824 | .9576 | .9095 | .8273 | .7031 | .5387 | .3518 | .1773 | .0556 | .0055 |
| 12 | .9963 | .9893 | .9729 | .9383 | .8732 | .7639 | .6020 | .3958 | .1841 | .0362 |
| 13 | .9995 | .9983 | .9948 | .9858 | .9647 | .9198 | .8329 | .6814 | .4510 | .1710 |
| 14 | 1.0000 | .9999 | .9995 | .9984 | .9953 | .9866 | .9648 | .9126 | .7941 | .5367 |
| 15 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

El LQAS tiene en cuenta ambos tipos de posibles errores en la decisión que tomamos al contrastar la hipótesis de cumplimiento del estándar.

Para monitorizar con LQAS, además del nivel de calidad que se quiere comprobar (estándar), hay que especificar los riesgos α y β que estemos dispuestos a correr: lo que equivale a especificar la probabilidad de falsos positivos y falsos negativos que queremos que tenga nuestro proceder para identificar situaciones problemáticas.

Para aplicar el LQAS, más que el tamaño del "lote" (marco muestral) es importante considerar que éste sea homogéneo (mismo tipo de paciente o casos por evaluar).

El LQAS junto con sus cuadros, tal como se ha desarrollado en la industria, puede parecer complejo y difícil de manejar; sin embargo su simplificación y adaptación para la monitorización de indicadores de salud es relativamente sencilla y puede limitarse a un manejo adecuado de los cuadros de probabilidades acumuladas de la distribución binomial o, más simple aún, de los elaborados directamente para su uso con LQAS.

5. El muestreo de aceptación de lotes como método de monitorización en los servicios de salud

De acuerdo con la definición de control de calidad en la industria, el muestreo para la aceptación de lotes es el proceso de evaluar una porción (muestra) de un lote de un determinado producto, con el propósito de aceptar o rechazar el lote en su totalidad. Las decisiones se toman sobre la base de la probabilidad de encontrar un número determinado de casos defectuosos (número de decisión) en muestras tomadas de cada lote (que sería el marco muestral), asumiendo la existencia de un determinado nivel de calidad o porcentaje de cumplimiento de los requisitos inspeccionados. El número de decisión se escoge de manera que los riesgos de rechazar lotes *buenos* (es decir, rechazar lotes que en realidad tengan el nivel de calidad deseado: riesgo α o *riesgo del productor*), y de aceptar lotes *malos* (riesgo β o *riesgo del consumidor*) estén a un nivel conveniente y predeterminado. Estos riesgos y el plan de muestreo en sí mismo se basan en la distribución binomial en la mayoría de los esquemas o en la de Poisson (si la proporción de casos defectuosos es inferior a uno por ciento). Existen también esquemas de LQAS para variables cuantitativas continuas, tomando entonces la decisión sobre la base de medias y desviaciones estándar, pero aquí sólo tratamos las del LQAS para variables cualitativas dicotómicas (cumple/no cumple).

Los parámetros que definen el plan de muestreo son: (i) el tamaño del lote (o marco muestral), (ii) el tamaño de la muestra por evaluar, y (iii) el número para la aceptación o número decisional, una vez que hemos definido los estándares en relación con el nivel de calidad asumido (llamados índices de calidad en la literatura de la industria), y los riesgos (α y β) del muestreo.

El tamaño del lote tiene poco efecto sobre la probabilidad de aceptación: en general, el tamaño del lote no importa, a menos que el tamaño de la muestra sea mayor que 10% del tamaño del lote. Sin embargo se aconseja que cuanto más grandes sean los lotes, más pequeños deben ser los riesgos o errores admitidos para la toma de decisiones, para minimizar las consecuencias de decisiones incorrectas. En los servicios de salud, conviene recordar que, más que el tamaño del lote, que puede definirse en términos de los casos ocurridos en un determinado periodo (visitas en un año, diagnósticos en los últimos seis meses, etc.), hay que prestarle atención a la homogeneidad del lote, el cual debe estar constituido por el mismo tipo de casos; es decir, lotes del mismo producto y no mezclas de ellos.

El tamaño de la muestra y el número decisional dependen en gran medida de la definición de los índices de calidad y los riesgos que se acepta asumir con la toma de decisiones. Las combinaciones de tamaño de muestra y número decisional más eficientes para las características deseadas para el plan de monitorización, pueden encontrarse en los cuadros y figuras de las curvas de características operativas (operating characteristics curves, OC) en la literatura de la industria. Estos cuadros y figuras no están exentas de una cierta dificultad de comprensión y manejo; pero podemos simplificarlas para nuestros propósitos al entender sus raíces y mecánica de construcción. La figura 3.2 contiene el esquema básico de una curva OC, las cuales representan la probabilidad de aceptación de diversos niveles de calidad para un determinado tamaño de muestra y número de decisión. Hay, pues, una curva OC para cada número de decisión, una vez fijado el tamaño de la muestra. Son, esencialmente, la representación gráfica de las probabilidades binomiales acumuladas que corresponden a un determinado número decisional, en un determinado tamaño de muestra, dependiendo del valor real de incumplimientos que pueda tener el lote de donde se extrae la muestra.

Pero, en definitiva, lo que vamos a manejar en la monitorización con LQAS no son las curvas de OC, sino las tablas de donde se derivan y que combinan todo lo necesario para diseñar la medición e interpretar los resultados de la monitorización.

6. Índices de calidad, riesgos y procedimientos para la aplicación del LQAS

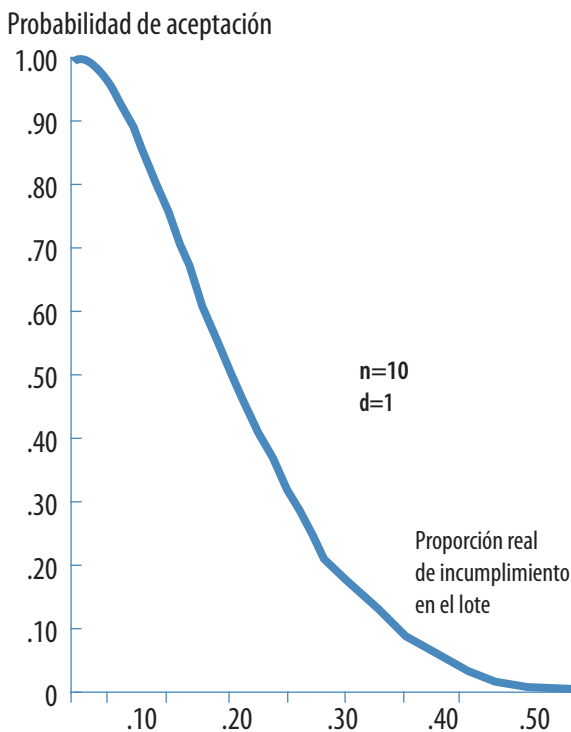
En los diversos esquemas de cuadros para LQAS en la industria, se manejan diversos índices de calidad, pero los dos más comunes, importantes y fácilmente traducibles a un LQAS para servicios de salud son los del nivel de calidad aceptable (acceptable quality level, AQL) y nivel de calidad limitante (limiting quality level, LQL). AQL es el porcentaje máximo de defectos o incumplimientos que puede considerarse satisfactorio como media en el proceso o producto evaluado. LQL es un nivel inaceptable, el límite que define una calidad claramente insatisfactoria. Los riesgos o errores de las decisiones al monitorizar se establecen de manera que la probabilidad de aceptar lotes con nivel de cumplimiento igual o mayor que el que corresponde al AQL es alto (y la probabilidad

de rechazarlos, el error α , o riesgo del productor, bajo), mientras que la probabilidad de aceptar un lote con cumplimiento de los requisitos de la calidad al nivel de LQL debe ser convenientemente baja (error β , o riesgo del consumidor). En la figura 3.3 están representados visualmente, en un esquema de curva OC, la relación entre AQL, LQL y errores α y β .

Una traducción obvia del concepto de AQL es la del estándar que queremos para el indicador a monitorizar, y para el LQL el de umbral para ese indicador que consideramos definitivamente problemático y que debe llevarnos necesariamente a evaluar e intervenir.

Uno de los cuadros más fáciles de manejar de los originados en la industria es el ANSI/ASQL Z1.4 (American National Standards Institute, ANSI), adaptación del MIL-STD-105E (Military Standard 105E) y equivalente a la ISO 2859 (International Organization for Standardization, ISO).

El tamaño de la muestra para LQAS depende, sobre todo, de los índices de calidad prefijados para el indicador (estándar y umbral) y los riesgos de equivocarse (falsos positivos y falsos negativos) que se quieren asumir.



*: Hay una curva de características operativas para cada valor de "d", dado un determinado tamaño de muestra.

Fuente: adaptado de Schilling, 1982.

Figura 3.2. Ejemplo de curva de características operativas (n=10, d=1)

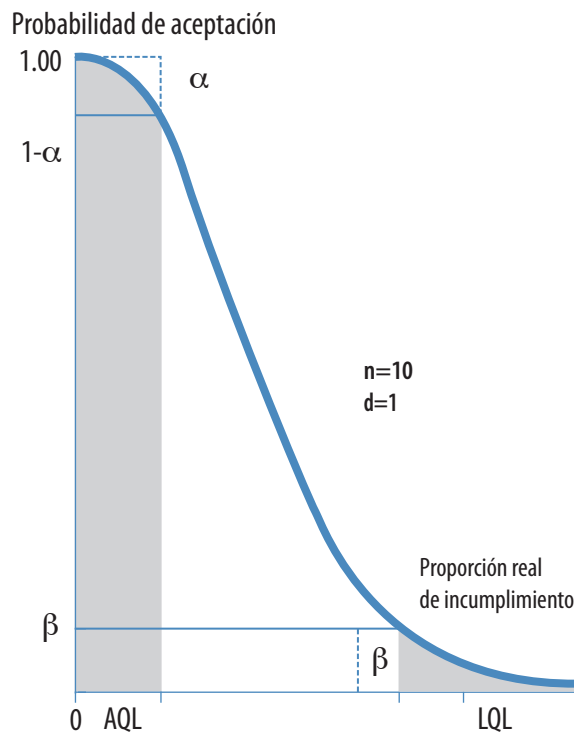


Figura 3.3. Curva de características operativas: Representación gráfica de índices de calidad (AGL, LGL) y errores alfa y beta

Las curvas OC representan gráficamente la relación entre los diversos resultados posibles en una muestra y los riesgos de equivocarse al aceptar/rechazar un determinado nivel de calidad. En la práctica, sin embargo, se utilizan los cuadros de LQAS, más que las curvas OC.

En el ámbito de la industria las tablas de LQAS van referidas a diversos niveles de AQL y LQL (nivel de calidad aceptable y nivel de calidad limitante) que son equiparables a los conceptos de estándar y umbral para los indicadores de calidad en los servicios de salud.

Los cuadros de LQAS utilizados en la industria son complejos de entender, y proponemos aprender a manejar los que se han adaptado, simplificándolos, para su utilización en los servicios de salud.

Las tablas de LQAS indican el tamaño de muestra más eficiente y el número decisional (de cumplimientos o incumplimientos) para aceptar o rechazar si hay problema de calidad (incumplimiento o no del estándar prefijado).

Aunque las normas para decidir pueden derivarse de los cuadros de la distribución binomial, proponemos para monitorización con LQAS utilizar los ya elaborados que se incluyen como cuadros 3.5 y 3.6 en este capítulo, adaptados de los realizados por Lemeshow et al. para uso en servicios de salud.

Se trata de un conjunto de cuadros basados en diversos valores de AQL y errores α , y diversos esquemas de inspección más o menos rígidos. Su aspecto a primera vista es complejo; sin embargo, existen algunos cuadros que adoptan el LQAS a los servicios de salud, cuya comprensión y manejo son más sencillos que los originales de la industria. Dos de las pocas propuestas para uso en servicios de salud son las de Lemeshow y otros, y la de J. J. Valadez, para valorar esencialmente prevalencias y coberturas vacunales, cuyo enfoque y forma de utilización son ligeramente diferentes, aunque su mecánica sea finalmente idéntica a la que debemos seguir para utilizar los cuadros adaptados por nosotros para monitorización de indicadores de calidad, que incluimos en este capítulo. Consiste en lo siguiente:

- Se define el *lote* o marco muestral en el que queremos comprobar el cumplimiento del indicador; por ejemplo, las visitas de un determinado tipo de pacientes en los últimos seis meses o un año, los pacientes incluidos en un determinado diagnóstico, programa o tratamiento, etcétera.
- Se establecen el estándar (nivel de cumplimiento deseado, AQL) y el umbral (nivel mínimo de cumplimiento aceptable, LQL) para el indicador.
- Se deciden los riesgos (α y β , del *productor* y del *consumidor*, o falsos positivos y falsos negativos) que estamos dispuestos a tener al clasificar el *lote* como aceptable.

En los cuadros, se busca la combinación más eficiente (muestra más pequeña) de tamaño de muestra (n) y número decisional (d) con la que podemos efectuar la clasificación, dado el estándar, el umbral y los riesgos prefijados.

Una vez identificados " n " y " d ", la práctica de LQAS es tremendamente simple: se reduce a la extracción de una muestra aleatoria de " n " casos, tomando el lote como marco muestral, y se cuentan los casos en los que el requisito de calidad evaluado no se cumple: si este número es mayor que " d ", rechazamos el *lote*: concluimos que AQL (el estándar) no se cumple porque la probabilidad del resultado obtenido es muy baja si AQL fuese cierto; hay, por lo tanto, un problema de calidad. Si, por el contrario, el número de incumplimientos es igual o menor que " d ", aceptamos la calidad del *lote*, equivalente a aceptar que el

estándar se cumple, además de que sabemos cuál es la probabilidad de que estemos aceptando un "lote" con cumplimiento a nivel de LQL o peor. En definitiva, lo único que hacemos es contar el número de incumplimientos y ver si es mayor o menor que " d ". No hace falta calcular nada. Una vez identificados " n " y " d ", cualquier persona puede monitorizar, sin necesidad de saber cómo se ha llegado a ellos. Recordemos, sin embargo, que " d " va a ser el número de cumplimientos o incumplimientos (en realidad se complementan), en función de que el estándar lo definamos a su vez como cumplimiento o incumplimiento y, que en los cuadros de diferentes manuales puede aparecer como " c ", " a ", " k ", " d " o " x ".

7. LQAS en servicios de salud: dos ejemplos de aplicación en la práctica

Vamos a suponer que queremos monitorizar con LQAS los dos ejemplos de indicadores ejemplificados en el principio de este capítulo: la proporción de pacientes con revisión de tensión arterial y la de los diagnosticados de resfriado común con prescripción inadecuada de antibióticos, medidos ambos cada seis meses. Como lo hemos definido hasta ahora, el AQL (el estándar) es de 85% para la primera y 15% para la segunda (85% también, si lo definimos como no prescripción de antibióticos).

Necesitamos definir ahora LQL (el umbral) o nivel mínimo aceptable. Para la detección de hipertensos podemos establecerlo en 60%; sin embargo, para la no prescripción de antibióticos, nos parece que deberíamos ser más exigentes y situarlo en 65%; aunque nos parece bajo y consideramos que es deseable situarlo en 90 por ciento.

Asimismo, es necesario tomar en cuenta las situaciones problemáticas. Creemos que sería aceptable considerar un error α (falsos positivos) de 5% y una especificidad de 80% (20% de falsos negativos). Para buscar " n " y " d " tenemos tres opciones: utilizar los cuadros de probabilidad acumuladas de la distribución binomial, o los de LQAS que hemos adaptado de los de Lemeshow y otros, como los reproducidos en los cuadros 3.5 y 3.6, o los de Valadez.

El cuadro 3.5 contiene los tamaños de muestra y número decisional para aquellas decisiones que arriesgan 5% de falsos positivos y 20% de falsos negativos en la identificación de situaciones

Cuadro 3.5
 Tamaño de muestra (n) y número de decisión (c)
 Para nivel de significación: 5%; poder: 80% (error $\alpha=5\%$, error $\beta=20\%$)
 Estándar (po%)

| Umbral (P _a %) | 50 | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 |
|------------------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|-------|
| | n c | n c | n c | n c | n c | n c | n c | n c | n c | n c |
| 10 | 8 1 | 6 1 | 5 1 | * | * | * | * | * | * | * |
| 15 | 11 2 | 8 2 | 7 2 | 5 1 | * | * | * | * | * | * |
| 20 | 15 4 | 11 3 | 9 2 | 7 2 | 5 1 | * | * | * | * | * |
| 25 | 23 7 | 16 5 | 12 4 | 9 3 | 7 2 | 5 2 | * | * | * | * |
| 30 | 37 13 | 24 9 | 16 6 | 12 5 | 9 4 | 6 2 | 5 2 | * | * | * |
| 35 | 67 26 | 38 15 | 24 10 | 16 7 | 11 5 | 8 3 | 6 3 | * | * | * |
| 40 | 153 66 | 68 30 | 38 17 | 23 11 | 16 8 | 11 5 | 8 4 | 5 2 | * | * |
| 45 | 617 288 | 154 74 | 67 33 | 37 19 | 22 11 | 15 8 | 10 5 | 7 4 | 5 3 | * |
| 50 | | 615 317 | 151 80 | 65 35 | 35 20 | 21 12 | 13 8 | 9 5 | 6 4 | * |
| 55 | | | 600 340 | 145 84 | 62 37 | 32 19 | 19 12 | 12 8 | 7 4 | * |
| 60 | | | | 573 353 | 136 86 | 57 37 | 29 19 | 16 11 | 10 7 | 5 3 |
| 65 | | | | | 534 356 | 125 85 | 50 35 | 25 18 | 13 9 | 7 5 |
| 70 | | | | | | 483 346 | 109 80 | 43 32 | 20 15 | 9 7 |
| 75 | | | | | | | 419 321 | 91 71 | 33 26 | 14 11 |
| 80 | | | | | | | | 342 279 | 69 58 | 22 19 |
| 85 | | | | | | | | | 253 219 | 44 39 |

Cuadro 3.6
 Tamaño de muestra (n) y número de decisión (c)
 Para nivel de significación: 5%; poder: 90% (error $\alpha=5\%$, error $\beta=10\%$)
 Estándar (po%)

| Umbral (P _a %) | 50 | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 |
|------------------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|-------|
| | n c | n c | n c | n c | n c | n c | n c | n c | n c | n c |
| 10 | 10 2 | 8 2 | 6 1 | 5 1 | * | * | * | * | * | * |
| 15 | 14 3 | 11 3 | 8 2 | 7 2 | 5 1 | * | * | * | * | * |
| 20 | 20 6 | 15 5 | 11 3 | 9 3 | 7 2 | 5 2 | * | * | * | * |
| 25 | 31 10 | 21 7 | 16 6 | 12 5 | 9 4 | 7 3 | 5 2 | * | * | * |
| 30 | 50 19 | 32 12 | 22 9 | 16 7 | 12 5 | 9 4 | 7 3 | 5 2 | * | * |
| 35 | 92 38 | 52 22 | 33 15 | 22 10 | 16 8 | 11 5 | 8 4 | 6 3 | 5 3 | * |
| 40 | 211 93 | 93 43 | 52 25 | 32 16 | 22 11 | 15 8 | 11 6 | 8 5 | 6 4 | * |
| 45 | 853 402 | 212 104 | 93 48 | 51 27 | 31 17 | 21 12 | 14 8 | 10 6 | 7 4 | * |
| 50 | | 852 444 | 210 114 | 91 51 | 49 29 | 30 18 | 19 12 | 13 8 | 9 6 | 5 3 |
| 55 | | | 834 477 | 203 120 | 87 53 | 46 29 | 27 18 | 17 12 | 11 8 | 7 5 |
| 60 | | | | 798 496 | 191 123 | 80 53 | 42 29 | 24 17 | 14 10 | 8 6 |
| 65 | | | | | 746 501 | 176 122 | 72 52 | 36 27 | 20 15 | 11 9 |
| 70 | | | | | | 676 488 | 156 116 | 62 48 | 30 24 | 15 12 |
| 75 | | | | | | | 589 455 | 131 104 | 49 40 | 21 18 |
| 80 | | | | | | | | 484 398 | 102 86 | 34 30 |
| 85 | | | | | | | | | 362 316 | 67 60 |

problemáticas. La manera como está construido el cuadro incluye el tamaño de muestra más eficiente y el número decisional en el mismo sentido que el estándar; es decir, la decisión se toma sobre la existencia de un número mínimo de cumplimientos, si el estándar es de cumplimiento, en vez de un número máximo, como vimos en el apartado anterior. Esto se debe a que Lemeshow y otros orientan su utilización para clasificar las poblaciones en función de niveles de prevalencia (enfermedades, vacunación, etc.), de modo que lo que nosotros hemos llamado estándar de cumplimiento, para ellos es nivel de prevalencia (que figura en el cuadro como P_o); el umbral o LQL, en el cuadro es la prevalencia alternativa, P_a ; lo que figura como “ c ” es el mínimo de casos de la enfermedad (o, en general, cumplimiento de lo que mida P_o) que hay que encontrar en la muestra “ n ” para aceptar P_o . Cualquier valor de cumplimientos menor a “ c ” nos ha de conducir a rechazar P_o , es decir, rechazar la *calidad* del lote (que exista el nivel de prevalencia P_o según el enfoque de Lemeshow y otros).

Volviendo a nuestros ejemplos, la monitorización con LQAS del indicador sobre hipertensión (estándar o $P_o=85\%$, umbral o $P_a=60\%$) podría hacerse, según el cuadro, con una muestra de 16 casos, en la cual debemos encontrar un mínimo de 11 cumplimientos. Para el indicador de la no prescripción de antibióticos (estándar: 90% , umbral= 65%), podemos hacer la monitorización con una muestra de 13 casos, en la cual debemos encontrar un mínimo de nueve cumplimientos para no clasificar la situación como problemática. Estos valores de “ n ” y “ c ” se obtienen del cuadro, viendo donde se cruzan los valores asignados por nosotros a P_o (estándar) y P_a (umbral). Así procederíamos para buscar los “ n ” y “ c ” más eficientes para cualquier estándar y umbral prefijados.

Puede observarse que cuanto más cerca esté P_o de P_a , el tamaño de muestra que necesitamos es mayor. Por ejemplo, para un estándar de 85% , si ponemos el umbral en 65% (en vez de 60% como hemos hecho) necesitaríamos muestras de 25 casos. De igual manera, si el umbral para la no prescripción de antibióticos lo elevamos a 70% , necesitaremos muestras de 20 casos para la monitorización, y 33, si lo elevamos a 75% (cuadro 3.5).

Por otra parte, si decidimos, disminuir los riesgos de equivocarnos al tomar la decisión (errores α y β), también necesitaremos muestras

más grandes. Por ejemplo, el cuadro 3.6 contiene los valores de “ n ” y “ c ” para que las decisiones de aceptar P_o (estándar) tengan un error β (aceptación de lotes con cumplimiento P_a), reducido a 10% (en el cuadro anterior es 20%). Buscando en este cuadro las mismas respuestas que hemos buscado en el 3.5, vemos que para monitorizar el indicador de hipertensión ($P_o=85\%$, $P_a=60\%$), necesitaríamos una muestra de 24 casos (en vez de 16); para la no prescripción de antibióticos ($P_o=90\%$, $P_a=65\%$), una de 20 (en vez de 13). Lemeshow y otros han construido cuadros para diversos errores α y β , con valores de error α de 1, 5 y 10% ; en combinación con errores β de 10, 20 y 50% . Con ello tenemos diversas opciones, entre las que debemos escoger en función del tipo de problema a monitorizar y la importancia diferencial que podemos darle al riesgo de tener falsos positivos o falsos negativos al decidir sobre la identificación de la situación como problemática.

Los cuadros construidos por Valadez, cuyo objetivo es identificar el tamaño de muestra y el número decisional para LQAS, están elaborados con un enfoque diferente. En estos cuadros, el punto de entrada es el tamaño de la muestra y el estándar de cumplimiento, valorando posteriormente qué nivel de LQL corresponde al error β (falsos negativos) que estamos dispuestos a admitir. Para ello, Valadez ha construido 45 cuadros de probabilidad binomial, uno para cada tamaño de muestra desde $n=5$ a $n=50$. En ellos se contemplan diversos niveles de cumplimiento del indicador por analizar (desde 5 hasta 95%), en los que el número decisional se da, sin embargo, como número de incumplimientos (recuérdese que tanto en los cuadros de probabilidad binomial como en los de LQAS de Lemeshow y otros, el estándar y el número decisional va en el mismo sentido). Con ello, una vez fijado el estándar, lo que se busca en el cuadro es el tamaño de muestra más conveniente y el número máximo de incumplimientos que pueden permitirse para aceptar que la situación no es problemática (es decir, que el estándar se cumple), y los riesgos de equivocarse que conlleva según los diversos valores del umbral. Los cuadros de Valadez son más numerosos y más engorrosos para manejar que los de Lemeshow y otros, por lo que proponemos utilizar los que hemos adaptado, incluidos en este capítulo; estos dos cuadros (3.5 y 3.6) pueden ser suficientes para los planes de monitorización con LQAS

La muestra necesaria aumenta si al determinar las condiciones para monitorizar acercamos el umbral al estándar o disminuimos los riesgos de equivocarnos al aceptar/rechazar si hay problema de calidad.

La magnitud de los errores α y β que elijamos dependerá del efecto que puedan tener los falsos positivos y falsos negativos como resultado de la monitorización. Normalmente α lo fijamos en 5% y β en 20 o 10 por ciento.

en servicios de salud. Para una visión más amplia de todos ellos, puede consultarse la bibliografía original que se indica al final de este capítulo.

Una vez localizados “*n*” y “*c*” en los cuadros, según las condiciones (estándar, umbral, errores α y β) que hemos decidido para la monitorización, ya no hacen falta ni los cuadros ni hacer ningún cálculo adicional para interpretar los resultados, hasta que decidamos cambiar los parámetros (por ejemplo el estándar) para la monitorización. El anexo 3.1 resume los pasos para la monitorización de indicadores utilizando LQAS.

8. Esquemas avanzados de muestreo de aceptación de lotes

Es evidente que la muestra para monitorizar con LQAS puede dejar de ser pequeña si queremos acercar los valores de estándar y umbral o reducir la probabilidad de equivocarnos en la decisión de aceptación/rechazo. Por tal motivo, se han desarrollado en la industria diversos esquemas de LQAS que introducen modificaciones tendientes a minimizar este inconveniente y poder monitorizar con AQL y LQL relativamente cercanos y errores α y β pequeños. Estas modificaciones incluyen: (i) cuadros con diversos niveles de rigor para utilizar de forma alternativa, (ii) esquemas de muestreos dobles y secuenciales y (iii) esquemas de muestreo bayesianos.

8.1. Diferentes niveles de exigencia según la evolución y el coste del muestreo

Los métodos expuestos en los apartados anteriores son normas para aplicar a la inspección de lotes de forma independiente y teniendo en cuenta sólo los parámetros (AQL, LQL, y riesgos α y β) que hemos establecido. Sin embargo, si la monitorización se hace realmente rutinaria y con una cierta frecuencia, podemos replantear esquemas menos exigentes, si lo normal viene siendo la no existencia de problemas de calidad, o, por el contrario, más estrictos, si nuestro esquema inicial viene detectando problemas con una cierta frecuencia.

Cambios en los niveles de exigencia conllevan normalmente cambios en el tamaño de la muestra necesaria para decidir. Estos cambios están contemplados en la mayoría de las tablas para LQAS de la industria, que distinguen, por lo menos, tres niveles: inspección convencional, reducida y es-

tricta. La reducida es la que emplea tamaños de muestra menores y se aplica cuando los niveles de calidad vienen siendo consistentemente elevados, a la vez que se establecen señales de alarma para pasar a esquemas más estrictos en función de la problemática que vayamos encontrando.

Otro aspecto que se considera para establecer el nivel de inspecciones es su coste en sí, en relación con lo que puede evitar en términos de coste de mala calidad que pase inadvertida si no se inspecciona. Esta es, por ejemplo, la base de los tres niveles de inspección de los cuadros ANSI/ASQC Z1.4, pensados para sistemas en los que la ratio entre el coste de inspección por caso (I_c) y el coste por caso de la mala calidad evitable por la inspección (C_n) son de 0.4, 1.0 y 1.6. En sus extremos, esta relación $I_c/C_n=P_c$ puede llevar a aconsejar no inspeccionar o, por el contrario, revisar 100% de los casos. En la decisión interviene la ratio de costes (P_c) y el nivel de calidad esperado (Q_e) en términos de proporción de casos defectuosos (incumplimientos); en general la inspección será más estricta (hasta llegar a 100%, como en el caso de los indicadores centinela) cuanto más pequeña sea la ratio (es decir, mayor el coste de la mala calidad en relación con el coste de la inspección) y mayor sea la diferencia entre el nivel de calidad esperada y el valor de la ratio de costes (Q_e-P_c). Sin embargo, este tipo de cálculos, aunque interesantes, son relativamente más fáciles de hacer en la industria, con más tradición y delimitación clara de sus productos y costos, y pueden resultar demasiado complejos para los servicios de salud.

8.2. LQAS con muestreo doble y secuencial

Sobre la base de que las muestras pequeñas pueden detectar las situaciones claramente problemáticas, pero conllevan errores mayores si la situación es intermedia, en la industria se han desarrollado cuadros de LQAS con esquemas de muestreo doble. En estos cuadros se comienza con una muestra pequeña y un número de decisión que permite detectar situaciones muy problemáticas (LQL muy alejado de AQL) con pequeño riesgo de error, y se realiza una muestra adicional en el caso de que la primera más pequeña no permita tomar decisiones.

Dadas las ventajas de este proceder al evitar muestreos ineficientes (muestras más grandes de

El manejo de las tablas de monitorización con LQAS en la práctica es muy simple y no precisa de conocimientos estadísticos especiales ni de cálculo adicional alguno una vez fijados “*n*” y “*c*”.

Los esquemas de LQAS que combinan diversos niveles de exigencia consideran explícitamente el coste de la inspección y de la mala calidad detectados; establecen muestreos secuenciales, y se basan en probabilidad bayesiana. Son modificaciones de gran interés; pero aún no adaptadas para su uso en servicios de salud.

lo necesario para decidir), con la misma filosofía se han desarrollado esquemas de muestreo para LQAS múltiples o secuenciales. En ellos se empieza con una muestra pequeña y el lote se acepta o se rechaza si el número de incumplimientos es muy pequeño o relativamente grande. Hay, pues, dos números decisionales, de manera que los resultados del muestreo con incumplimientos entre esos dos números aplazan la toma de decisiones para realizar una segunda muestra que se suma a la anterior, en donde se procede de igual manera; y así sucesivamente con una tercera, cuarta, etc. Siguiendo este proceder, en el peor de los casos, podríamos llegar a tener una muestra total equivalente a la que necesitaríamos para una estimación del nivel de calidad con una determinada precisión, en función de lo alejado que queramos que esté LQL del nivel AQL.

En el sector salud, el muestreo secuencial se está empleando con timidez en algunos ensayos clínicos que investigan la eficacia de un determinado tratamiento y puede resultar poco ético esperar a tener los resultados de un muestreo tradicional, si con anterioridad puede decidirse con cierta seguridad sobre la bondad o maldad del tratamiento. Sin embargo, el muestreo secuencial, a pesar de sus ventajas, al igual que el LQAS, en general sigue sin formar parte de los métodos de muestreo habituales en los servicios de salud.

8.3. Métodos de LQAS con base bayesiana

En los procedimientos de LQAS revisados, se define el esquema de muestreo sobre unos paráme-

tros predeterminados (AQL, LQL, errores α y β), pero no se tiene en cuenta el nivel real de calidad que puede existir en los lotes que se evalúan. Los planes de LQAS tradicionales están pensados para protegerse de un nivel de mala calidad determinado, que muchas veces no existe en la realidad que valoramos o tiene muy pocas probabilidades de existir; si esto es así, el resultado es que se utilizan muestreos demasiado grandes para protegernos de niveles de mala calidad inexistentes o identificarlos como problema. Con estas premisas, se han desarrollado cuadros y planes de LQAS llamados planes de muestreo empíricos o bayesianos que incorporan los datos de cada evaluación efectuada a la evaluación del lote siguiente. Las comparaciones que se han hecho entre los planes habituales y los bayesianos revelan que estos últimos requieren muestras mucho más pequeñas, en especial cuando la variación en el tiempo de la proporción de incumplimientos es reducida. El hecho de que haya que conocer (o asumir) un determinado nivel de calidad previo y que funcionen, sobre todo, cuando este nivel es estable, al parecer ha dado lugar a que no se utilicen más.

En su conjunto, sin embargo, el conocimiento y aplicación de los métodos de muestreo desarrollados en la industria para el control de la calidad es un campo muy fecundo que puede hacer avanzar considerablemente la práctica de los programas de gestión de la calidad en los servicios de salud. Comencemos por aprender a aplicar los esquemas y cuadros de LQAS más simples, como las explicadas a propósito de los cuadros 3.5 y 3.6 de este capítulo.

Los muestreos secuenciales tratan de minimizar la realización de muestras ineficientes. Con este proceder se decide siempre con el mínimo tamaño de muestra posible, y, en el peor de los casos, la muestra total es equivalente a la que se realizaría para hacer una estimación del indicador.

BIBLIOGRAFÍA

- Richards LE, LaCava JJ. Business statistics. Why and when. 2ª ed. New York: McGraw Hill, 1993. Cap. 6.
- Schilling EG. Probability Functions. En: Acceptance sampling in quality control. New York: Marcel Dekker Inc., 1982. Cap. 3.
- Schilling EG. Concepts and terminology. En: Acceptance sampling in quality control. New York: Marcel Dekker Inc., 1982. Cap. 4.
- Lemeshow S, Hosmer DW, Klar J, Lwanga SK. Lot quality assurance sampling. En: Adequacy of sample size in health studies. Willshire: WHO/John Wiley & Sons, 1992.
- Valadez JJ. Assessing child survival programs in developing countries. Testing Lot Quality Assurance Sampling. Boston: Harvard University Press, 1991.
- Saturno PJ. La distribución binomial y el muestreo para la aceptación de lotes (LQAS) como métodos de monitorización en servicios de salud. Revista de Calidad Asistencial 2000;15(2):99-107.
- Rosner AB. Fundamentals of Biostatistics. 7th ed. Pacific Grove CA: Duxbury Press, 2011.

Anexo 3.1

Procedimiento para la monitorización de indicadores utilizando LQAS

1. Elegir dos indicadores medibles como variable dicotómica (cumple/no cumple). Pueden extraerse de listados existentes en el entorno del participante, seleccionados de entre los criterios que se hayan utilizado en un ciclo de mejora o diseñados para efectos de este trabajo práctico.
2. Definir el estándar (AQL, P0) y el umbral (LQL, Pa), cuya existencia queremos comprobar (aceptar/rechazar).
3. Decidir qué riesgos de falsos positivos y falsos negativos (α y β) estamos dispuestos a correr (elegir entre las dos combinaciones de los cuadros 3.5 y 3.6 de este capítulo).
4. Describir qué tamaño de muestra (n) y qué número decisional (especificando si es de cumplimientos o de incumplimientos) necesitamos para la monitorización.
5. En el caso de que el tamaño de muestra ("n") sea >30 para alguno de los dos indicadores, indicar qué proceder alternativo podríamos utilizar o qué parámetros de referencia podríamos modificar para utilizar una muestra más pequeña.
6. Describir cómo se interpretarían los resultados de la monitorización con las condiciones que se han prefijado.

Capítulo 4

Control estadístico de la calidad. Gráficos de desarrollo (*run charts*)

Contenido general y objetivos específicos

1. Contextualización
2. Control estadístico de la calidad. Conceptos básicos
3. La base estadística de los gráficos de control
4. Construcción de la plantilla de referencia
5. Análisis gráfico de la variabilidad: los cuatro tipos principales de gráficos
6. El control estadístico de procesos en la práctica: construcción e interpretación de los gráficos de control estadístico más sencillos
7. Gráficos de desarrollo
8. Interpretación de los gráficos de desarrollo

Anexo 4.1. Procedimiento para el análisis de una serie de mediciones con un gráfico de desarrollo

Contenido general

Se describe la base teórica del control estadístico de la calidad así como su aplicación en la monitorización y otras actividades de gestión de la calidad. Se presta especial atención a los gráficos de desarrollo y control, como herramientas de elección para el análisis y monitorización de indicadores con planes de mediciones frecuentes. Finalmente, se ejemplifica su uso en los cuatro tipos de indicadores más habituales en servicios de salud.

Objetivos específicos

1. Entender la base teórica del control estadístico de la calidad.
2. Describir las características básicas de las herramientas del análisis estadístico gráfico.
3. Describir los objetivos esperables del análisis de datos con gráficos de control estadístico.
4. Distinguir los conceptos de variabilidad de causa común y variabilidad de causa especial.
5. Distinguir las diferentes situaciones en las que son de utilidad para el control estadístico de la calidad.
6. Describir los requisitos de utilización del control estadístico para la monitorización continua de indicadores.
7. Distinguir la diferencia entre los gráficos x , \bar{x} , p y u .
8. Construir una plantilla de gráficos de desarrollo.
9. Interpretar los gráficos de desarrollo, identificando los patrones propios de variabilidad especial.

Control estadístico de la calidad. Gráficos de desarrollo

1. Contextualización

El control estadístico de la calidad y sus herramientas más genuinas, los gráficos de control, son probablemente la aportación más importante del ámbito de la industria para los programas de gestión de la calidad en cualquier actividad productiva o de servicios. En el capítulo 1, se menciona el control estadístico (gráfico) de la calidad como método de elección para monitorizar indicadores que se miden de forma continua o muy frecuente (como mínimo cada mes); a ello hay que añadirle su utilidad como herramienta de análisis de los procesos, para detectar, no sólo si hay problema (algo común a todos los métodos de monitorización), sino qué *tipo de problema* (esporádico o sistemático). Al mismo tiempo, los gráficos de control estadístico son sensibles a cualquier cambio introducido en el proceso que monitorizan, por lo que también pueden utilizarse para poner en evidencia los efectos de las intervenciones que se implementan para mejorar las situaciones que se han identificado como mejorables o problemáticas. Los gráficos de control estadístico son, pues, instrumentos muy versátiles, con múltiples aplicaciones.

La utilización de los gráficos de control estadístico es simple, muy clara visualmente, y consiste esencialmente en un muestreo comparativo frecuente con muestras pequeñas. Para comprender cómo funcionan, sólo hace falta un entendimiento básico sobre probabilidades y distribuciones más frecuentes, sobre todo, la normal; pero pueden ser utilizados aún sin estos conocimientos.

Tanto el análisis de la variabilidad como la monitorización del nivel estable de calidad se realizan por medio del uso masivo e inteligente de la

teoría de la probabilidad, utilizando como referencia distintas funciones de probabilidad (normal, binomial, Poisson, etc.), según las características del indicador que se mide. Sin embargo, su utilización rutinaria está normalizada y no precisa de ningún conocimiento estadístico. Adicionalmente, el tamaño de muestra necesario para la monitorización puede ser pequeño (5 a 10 casos para variables cuantitativas, 25 a 30 para cualitativas).

El análisis básico realizado con el control estadístico consiste en determinar si los valores obtenidos del indicador están dentro de los esperados, dada la variabilidad estadística normal, o si son estadísticamente incompatibles con una situación estable de partida. Sin embargo, este análisis es visual, gráfico y no entraña, una vez construida la plantilla gráfica, ningún cálculo adicional a la propia medición del indicador.

La gran utilidad potencial del control gráfico de la calidad contrasta con la lentitud de su utilización rutinaria en los servicios de salud. Probablemente se deba a que se precisa de series temporales con mediciones muy frecuentes para que puedan aplicarse. Sin embargo, además de que pueden utilizarse con éxito dentro de los ciclos de mejora y en el control de aspectos relevantes específicamente diseñados para su monitorización, en los servicios de salud se realizan muchas mediciones rutinarias con periodicidad mensual, incluso, diaria, cuyo análisis y presentación gráfica se beneficiarían enormemente si se hicieran por medio de gráficos de control estadístico.

En este capítulo vamos a describir las bases teóricas del control estadístico y ejemplificar los principales tipos de gráficos utilizables, centrándonos después en el más simple de todos: los gráficos de desarrollo o gráfico de rachas (*run chart*).

El control estadístico de la calidad es el método de elección para analizar y monitorizar indicadores que se miden con una frecuencia mensual o menor. Son de utilidad en todas las actividades de gestión de la calidad.

El control estadístico (gráfico) de la calidad se basa en la teoría de la probabilidad y el uso de distintas funciones de probabilidad, según el tipo de variable que mide el indicador.

El análisis del control estadístico se realiza de forma visual, una vez construida la plantilla de referencia.

2. Control estadístico de la calidad. Conceptos básicos

2.1. ¿Qué es un gráfico para el control estadístico de la calidad?

Un gráfico para el control estadístico de la calidad es esencialmente una plantilla en la que se colocan las mediciones del indicador que se está analizando o monitorizando, sólo que está construida de tal forma que sirva como referencia en la interpretación de los valores del indicador a analizar o controlar.

La propuesta original y utilización como herramienta para la identificación de fuentes de dificultades y defectos durante la producción la realizó Walter Shewart a principios de los años 20; pero su desarrollo y aplicaciones posteriores se han ido ampliando hasta constituir una de las ramas más fructíferas en cuanto a adaptaciones y utilización con ventaja, sobre otros enfoques o

métodos, prácticamente, en relación con todas las actividades de los programas de gestión de la calidad (ciclos de mejora, diseño, monitorización).

En la figura 4.1 se representan los esquemas básicos de los dos grupos de gráficos de control estadístico que vamos a describir en detalle en este capítulo y en el siguiente: gráficos de desarrollo (*run charts*) y gráficos de control. En ambos casos, la plantilla consta de un eje de ordenadas con la escala de valores que puede ir tomando el indicador, y un eje de abscisas con la secuencia temporal de las mediciones; hay también en ambos una línea central que representa el promedio del indicador por analizar, y, en el caso del gráfico de control, dos líneas adicionales llamadas *límites de control*, que delimitan la zona de variabilidad permisible para el indicador cuando el proceso que mide es estable. En estos ejes de coordenadas y con estas líneas de referencia se van colocando los puntos correspondientes a las mediciones que vayamos realizando del indicador para analizar

Un gráfico para el control estadístico de la calidad es una plantilla que sirve de referencia para interpretar la variabilidad del indicador analizado.

La plantilla de los gráficos más comunes consiste en unos ejes de coordenadas y una línea que representa el promedio estadístico del indicador (gráficos de desarrollo) y dos líneas adicionales (límites de control) que acotan la variabilidad esperable.

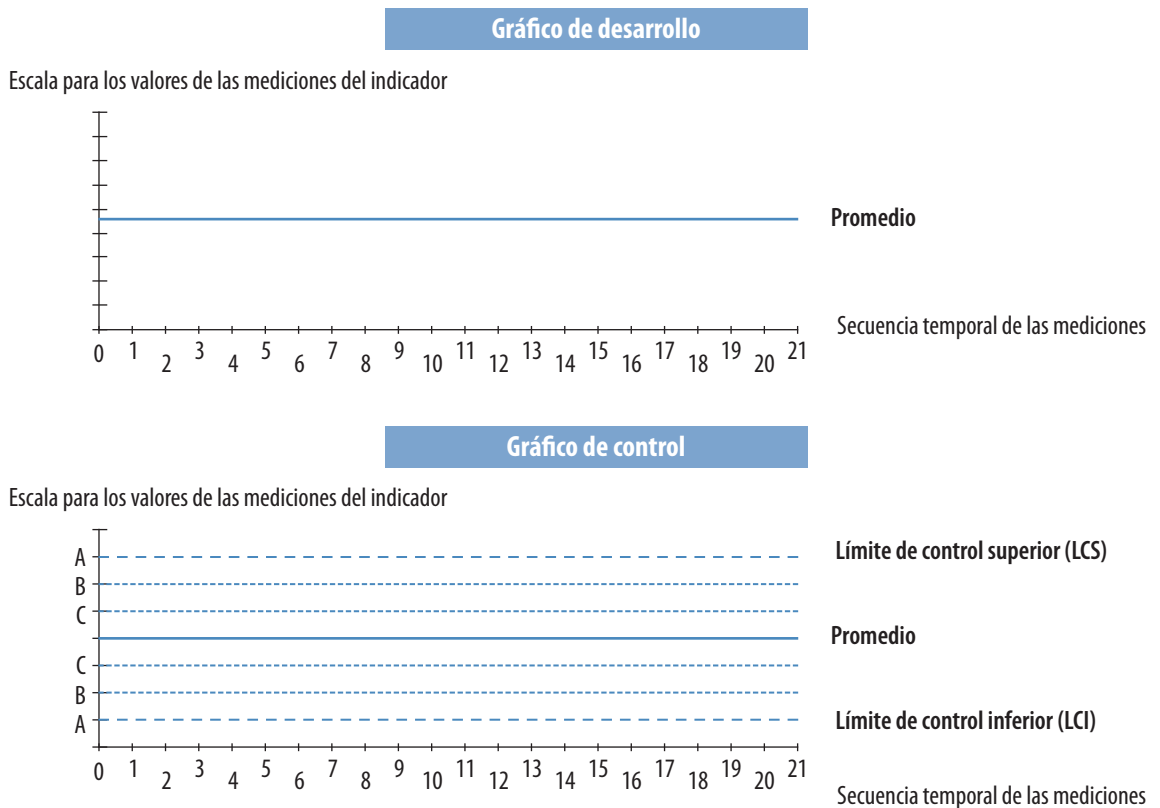


Figura 4.1. Elementos de los gráficos para el control estadístico de la calidad

la variabilidad del proceso que medimos y concluir si es estable o no, en función de las posiciones que vayan teniendo los puntos sucesivos en el gráfico.

2.2. Objetivo principal: analizar la variabilidad

El objetivo principal del control estadístico de la calidad es monitorizar continuamente la estabilidad de los procesos de producción, sea de bienes o de servicios, de manera que se detecten las situaciones problemáticas (*fuera de control*) y se puedan tomar acciones. Para ello se realiza un análisis gráfico de la variabilidad de las mediciones del indicador para identificar lo que se conoce como causas especiales, no aleatorias o asignables, y causas comunes, aleatorias o no asignables. Las causas *especiales* son aquellas que producen *inestabilidad*, un *exceso* de variación no esperable por puro azar; las *comunes* son aquellas inherentes al proceso y que producen una variabilidad estable, sistemática y predecible, propia de la aleatoriedad de los elementos que intervienen en el proceso analizado. Esta distinción es importante porque el tipo de investigación e intervención consiguiente también lo es. Las causas especiales son, en cierto modo, excepcionales y añadidas a lo que sería el funcionamiento normal del sistema, mientras que las comunes son inherentes al sistema y una actuación sobre ellas significa una remodelación en profundidad, innovadora, del proceso analizado.

Las decisiones sobre el tipo de variabilidad, especial o común, se basan en la teoría de la probabilidad y, en concreto, en las distribuciones de probabilidad que corresponden a cada tipo de indicador analizado. Estas decisiones son similares en su estructura a las pruebas de hipótesis; la hipótesis nula (de base) indica que el proceso monitorizado es estable, se mantiene en niveles constantes y la variabilidad de las mediciones es propia del azar dentro de lo esperado para la variabilidad inherente al proceso en sí; la hipótesis alternativa indicaría que la variabilidad observada no se debe al azar, sino que es fruto de cambios en el proceso por causas que habría que averiguar.

Cuando se monitoriza con gráficos de control estadístico, este contraste de hipótesis se realiza implícitamente con cada una de las mediciones del indicador, de manera que puedan identificarse visual y rápidamente aquellas situaciones

significativamente diferentes de la normalidad estadística (es decir, de un proceso estable con una variación en las mediciones esperable por azar). El riesgo de identificar como problemática una situación que no lo es (error α , tipo I o falsos positivos) se ha establecido en la industria tradicionalmente en <0.01 . La razón es puramente empírica, y busca minimizar investigaciones innecesarias de causas especiales. El error β , tipo II o falsos negativos, no se considera explícitamente en las decisiones que se toman en los gráficos de control estadístico; si bien parece, según la experiencia en la industria, que el proceder secuencial y centrado lo minimiza.

2.3. Situaciones en las que es útil el control estadístico. Condiciones para su utilización como herramienta de monitorización

Conviene señalar desde el principio que la monitorización con gráficos de control estadístico tiene sentido cuando se cumplen dos condiciones previas:

1. Cuando el proceso es estable o *estadísticamente controlado*. En este caso, en una serie de mediciones, se observa que los valores del indicador se mantienen dentro de la normalidad estadística (mientras siga habiendo variabilidad excesiva, indicativa de la existencia de causas *especiales*, lo que hay que hacer es averiguarlas e intervenir sobre ellas).
2. Cuando la estabilidad o normalidad estadística está dentro de los límites de calidad que queremos para ese indicador (llamados especificaciones de calidad en el lenguaje de la industria).

En la consecución de estas dos condiciones también son útiles los gráficos de control estadístico; pero, en este caso, se utiliza como método para las diversas fases de un ciclo de mejora. La monitorización en sí como actividad de control (tal como la hemos definido en el capítulo 1) se haría una vez estabilizada la variabilidad y obtenidos los niveles de calidad que nos parecen aceptables.

Hay que señalar también que la consecución de estas dos condiciones debe ser secuencial, de modo que primero hay que conseguir la estabilidad del proceso y, a continuación, asegurarnos del cumplimiento (*conformidad*) con los límites de calidad. En la figura 4.2 se representan en forma de

El control estadístico analiza la estabilidad de los procesos, distinguiendo variabilidad esperable por el azar, dentro del funcionamiento normal del proceso, de variabilidad no esperable. A las causas de la variabilidad esperable se les conoce como causas "comunes" y a las de variabilidad no esperable o excesiva, como "especiales".

Normalmente, la actuación sobre causas comunes, para disminuir la variabilidad o cambiar el promedio, implica un rediseño de los procesos.

La identificación de uno y otro tipo de variabilidad se realiza con base en un contraste continuo de la hipótesis de estabilidad, según la distribución de probabilidades que corresponde al tipo de variable que mide el indicador.

La utilización del control estadístico como herramienta para la monitorización continua requiere que el indicador esté previamente estabilizado (variabilidad estadística predecible) y acorde con el nivel de calidad deseado.

El algoritmo la interrelación entre estas situaciones y su correspondencia con las actividades dentro del esquema de actividades de gestión de la calidad.

El hecho de que los gráficos de control estadístico puedan utilizarse prácticamente en todas las actividades de los programas de gestión de la calidad ha dado como resultado que algunos au-

tores los describan como el paradigma de la optimización sucesiva de los procesos de producción, un sistema en sí mismo para saber cómo lo estamos haciendo, para que sepamos cómo hacerlo mejor.

La versatilidad de usos ha llevado también a una cierta confusión entre los diversos objetivos y

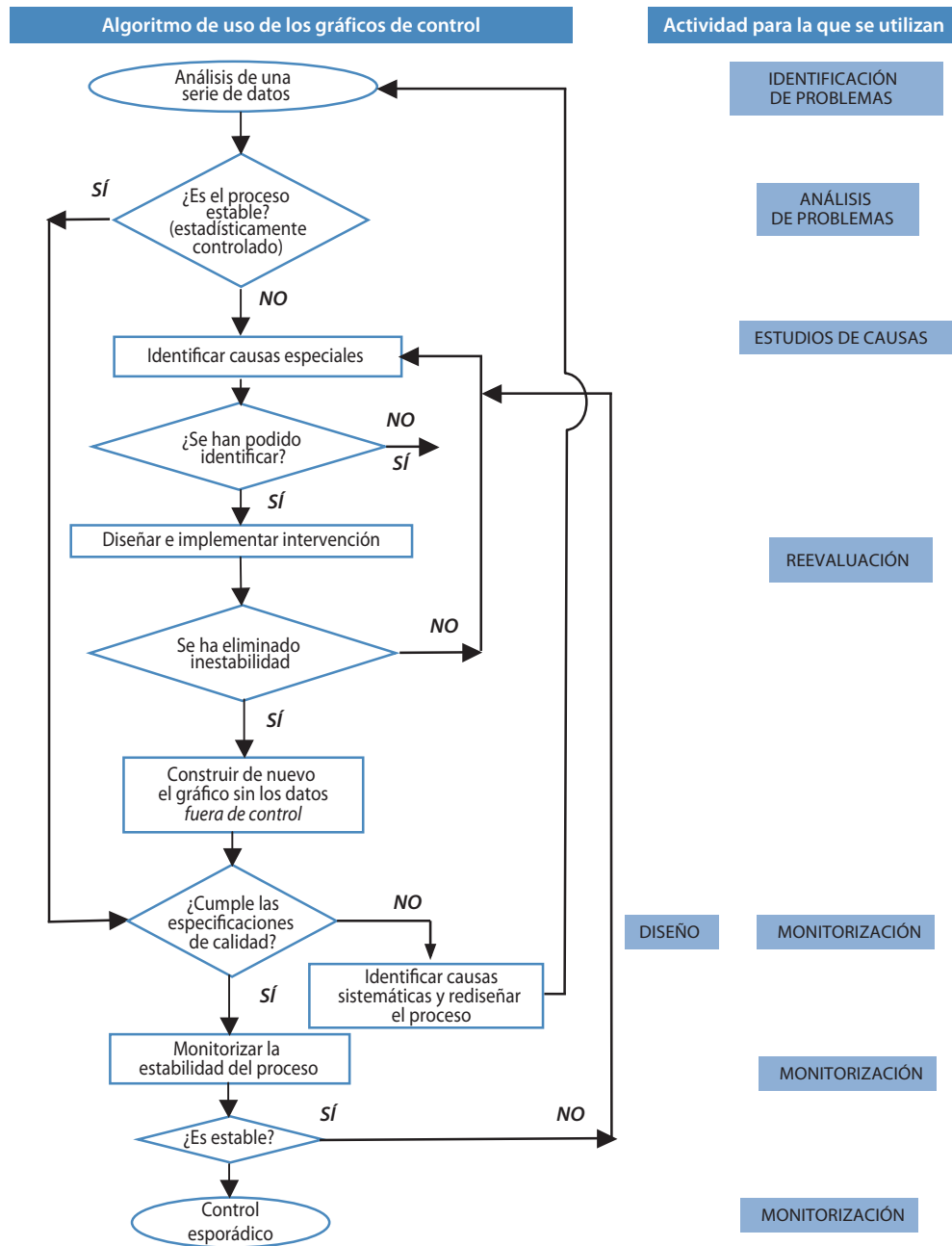


Figura 4.1. Algoritmo de utilización de los gráficos de control y su relación con las actividades de gestión de la calidad

condiciones de aplicación para cada uno de ellos; sin embargo, es muy importante no confundir objetivos y situaciones para los que se utiliza el control gráfico estadístico de los indicadores. Esta confusión aparece, incluso, en algunos manuales de control de la calidad y puede tener consecuencias prácticas indeseables. El hecho es que, aunque el método y las herramientas son similares, hay matices importantes dependiendo de su utilización. En resumen, podemos decir que los tres principales usos de los gráficos de control, cada uno con sus premisas y condiciones, son:

1. Análisis y control de la variabilidad de un proceso (representado por medio de sus indicadores correspondientes). Aplicable a cualquier situación. Puede ser equivalente a la identificación y análisis de un problema de calidad u oportunidad de mejora. Su objetivo final es conseguir la estabilidad (control estadístico) del proceso, por medio de la identificación y posterior eliminación de las causas *especiales*. Los gráficos de control sirven también, en este caso, para poner en evidencia el efecto de las intervenciones implementadas.
2. Análisis y control de la conformidad *de un proceso estable* con las especificaciones de calidad.

3. Monitorización de un proceso *estable y conforme* con las especificaciones, requisitos o estándares de calidad.

Los objetivos 1 y 2 comprenden a actividades propias de ciclos de mejora y/o diseño de la calidad. Más adelante veremos con ejemplos los diferentes usos de los gráficos de control para estas tres situaciones (cuadro 4.1).

3. La base estadística de los gráficos de control estadístico de la calidad

El razonamiento estadístico básico para el control gráfico de la calidad no puede ser más simple ni su utilización práctica más ingeniosa. Se trata en esencia de lo siguiente:

En todo proceso de producción existe variabilidad, que puede deberse a la variación en la forma de proceder (métodos de trabajo); al material empleado; a la forma de funcionar de la tecnología empleada e, incluso, a la forma de medir el indicador para valorar la variabilidad. Ahora bien, si toda esta variación se mantiene dentro de unos límites normales y constantes, las mediciones que hagamos girarán en torno a un valor medio (el propio del proceso), siguiendo un patrón de probabili-

Además de para la monitorización continua, las herramientas del control estadístico sirven para analizar la situación de partida de un indicador (identificación y análisis de problemas) comprobar el efecto de una intervención, y analizar la conformidad con los niveles de calidad requeridos.

Es conveniente tener claro en qué situación (ciclo de mejora, diseño, monitorización) y para qué objetivos vamos a utilizar las herramientas de control estadístico.

Cuadro 4.1

Control estadístico (gráfico) de la calidad:
utilidad, objetivos y lugar dentro de las actividades de gestión de la calidad

| Utilidad | Objetivos | Tipo de actividad |
|--|--|--|
| Análisis de series de datos (análisis de procesos) | <ul style="list-style-type: none"> ● Analizar la variabilidad ● Investigar las causas ● Comprobar el efecto de intervención | Ciclos de mejora (identificación y análisis de problemas, estudio de causas, reevaluación) |
| Análisis de la "capacidad de proceso" (en procesos estadísticamente estables) | Comprobar si las características de la variabilidad inherente al proceso analizado afectan los requisitos de calidad (especificaciones) que deben tener | <ul style="list-style-type: none"> ● Diseño de la calidad ● Monitorización |
| Monitorización (control de procesos estadísticamente estables) | <ul style="list-style-type: none"> ● Controlar continuamente la estabilidad de un determinado proceso (mediante un indicador o indicadores oportunos) ● Identificar las situaciones problemáticas ("fuera de control") que es necesario investigar | Monitorización |

La base estadística del control gráfico de los indicadores es la probabilidad de que los valores en cada medición se alejen más o menos de un valor promedio. Si el proceso es estable, es más probable que los valores de las mediciones resulten cercanos a la media. Por otra parte, la probabilidad de cualquier valor viene determinada por una distribución estadística conocida de antemano.

Si el proceso es estable, su valor medio es constante y 99.7% de las mediciones ha de situarse entre el promedio ± 3 desviaciones estándar.

dades conocido y cuantificable. Un patrón de probabilidades (llamado en estadística función de probabilidad o distribución de probabilidad) que generalmente nos indica que es más probable que los valores que vamos a encontrar en nuestras mediciones resulten cercanos a la media propia del proceso que lejos de ella; de hecho, cuánto más alejados estén del valor medio que caracteriza al proceso, menos probable será que nos aparezcan en nuestras mediciones, a menos que el proceso haya cambiado por alguna causa.

Ese valor medio del indicador que medimos se representa como línea central en la figura 4.1, con o sin otras líneas (límites de control) que delimitan los valores extremos esperables por azar. Lo que hacemos con cada medición del indicador es averiguar si el valor que encontramos es o no compatible con la variación estadística normal en torno al valor medio. Esta averiguación se hace visualmente (sin hacer cálculos) y se basa (como veremos después), no sólo en la medición que acabamos de realizar, sino en cuáles han sido los resultados de mediciones anteriores (que también estarán en la figura), pero siempre considerando la probabilidad de encontrar el valor que hemos obtenido si el valor medio permanece constante. En consecuencia, si es muy baja la probabilidad de que el valor encontrado en la medición sea compatible con la situación de partida (probabilidad que no hay que calcular, sino que se deduce por la situación del punto que representa nuestra medición en el gráfico de control), concluiremos que algo *especial* ha ocurrido; que el proceso está *fuera de control*, y es conveniente averiguar cuál o cuáles son las causas de esta variación extraordinaria.

Supongamos, por ejemplo, que la distribución de probabilidades que deben seguir las diversas mediciones del indicador, si el proceso es estable, es la distribución normal (la más conocida y la primera que explica cualquier libro de estadística elemental). Esta distribución se caracteriza por ser perfectamente simétrica en torno a la media, de manera que ocupa el valor central y hay exactamente 50% de probabilidades para valores por encima y 50% para valores por debajo, con menos probabilidades para cada valor individual más pequeñas, cuando estén más alejados del valor de la media. Estas probabilidades son perfectamente conocidas en función del número

de desviaciones estándar por encima o por debajo de la media (figura 4.3).

Según las probabilidades de la distribución normal (figura 4.3), si extraemos una muestra de un universo (marco muestral) de media m tenemos 68.27% de probabilidades de que el valor que encontremos en la muestra esté comprendido en el intervalo determinado por la media ± 1 desviación estándar ($m \pm 1s$); de igual manera, hay 95.45% de probabilidades de que esté comprendido en el intervalo $m \pm 2s$ (es decir, menos de 0.05% de que encontremos en la muestra valores por encima o por debajo de $m \pm 2s$), y 99.73% de que esté entre $m - 3s$ y $m + 3s$. Esto significa que, si encontramos en nuestra medición valores fuera del intervalo $m \pm 3s$, es muy poco probable (¡menos de uno por ciento!) que el universo de donde hemos extraído la muestra tenga la media m que suponíamos debe tener; concluiremos, por lo tanto, que ese nivel medio no se cumple,

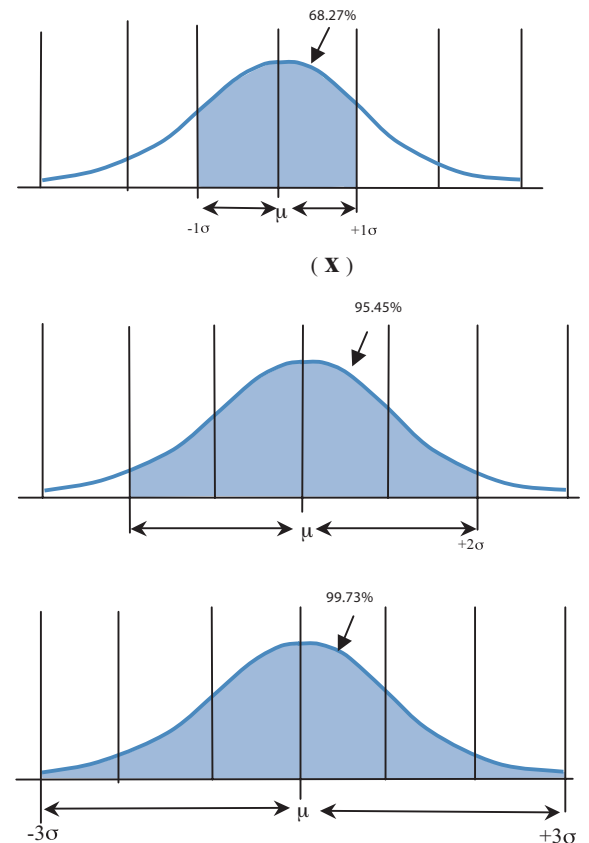


Figura 4.3. Distribución normal. Probabilidad de valores en torno a la media

que algo ha cambiado y la media del proceso se ha *desestabilizado*.

Este razonamiento es aplicable al valor de una medición aislada del indicador; pero ya hemos apuntado que también se detectan situaciones de inestabilidad con base en la secuencia de valores que obtenemos en mediciones sucesivas del indicador; por ejemplo, es signo de inestabilidad si encontramos un determinado número de mediciones sucesivas a un mismo lado (por encima o por debajo) de la media, o constantemente hacia arriba o hacia abajo. ¿Por qué son estos patrones indicativos de cambios o alteración en el proceso analizado? El razonamiento es igualmente probabilístico y tiene como base las probabilidades que seguirían las mediciones si se cumple la distribución normal (lo cual debe ocurrir en situaciones de estabilidad o control estadístico). En el caso de secuencias de valores, nos interesaremos por la probabilidad, ya no de un valor concreto (punto aislado de la figura), como hemos hecho en el razonamiento anterior, sino de que tengamos una secuencia concreta de varios puntos (cada uno representa una medición del indicador) considerados conjuntamente. Esta probabilidad conjunta de toda una secuencia de valores se calcula, según nos indica la estadística, multiplicando las probabilidades de cada uno de los valores considerados.

Como ejemplo, hagamos los cálculos para el caso más simple, con el fin de ver cómo funciona, pero con la advertencia de que, en la práctica, no es preciso hacerlos porque están perfectamente definidos los patrones o secuencias de puntos en el gráfico, ya que, estadísticamente, es imposible que mantengan la misma media; en la práctica sólo hay que ver si aparece alguno de ellos en el gráfico.

El caso más simple, como decíamos, es considerar la probabilidad de una secuencia de valores del indicador a un mismo lado (por encima o por debajo) de la media. Sabemos que en la distribución normal hay 50% de probabilidades de que una medición del indicador caiga a un lado determinado, y, por supuesto, un 100% de que caiga a uno u otro lado indistintamente. Con estos supuestos, la probabilidad de que los valores del indicador que vamos obteniendo en mediciones sucesivas caigan siempre a un mismo lado de la media serían las siguientes:

- Para la primera medición (primer punto a colocar en el gráfico): 0.5 (o 50%).
- Para la segunda medición: $0.5 \times 0.5 = 0.25$ (multiplicación de las dos probabilidades sucesivas).
- Para la tercera medición: $0.25 \times 0.5 = 0.125$ (probabilidad anterior multiplicada por la probabilidad de un nuevo punto *en el mismo lado*) y así sucesivamente:
- Para la cuarta medición: $0.125 \times 0.5 = 0.0625$.
- Para la quinta medición: $0.0625 \times 0.5 = 0.03125$.
- Para la sexta medición: $0.03125 \times 0.5 = 0.016$.
- Para la séptima medición: $0.016 \times 0.5 = 0.008$.

¿Qué significa esto? Que la probabilidad de obtener cinco mediciones sucesivas por encima o por debajo de la media es baja (<0.05 ; exactamente 0.03125) y debemos rechazar la hipótesis nula de que estamos ante la misma situación que teníamos antes de que ocurriese esta secuencia. Si queremos estar aún más seguros (disminuir el error α), esperaríamos a tener seis o siete mediciones sucesivas en el mismo lado de la media. Este último patrón tiene menos de 1% de probabilidades de presentarse, si la situación sigue siendo la que correspondería al promedio del gráfico.

Existen otros patrones o secuencias concretos de distribución de puntos en el gráfico que son indicativos, con una significación estadística determinada, de que la situación está *fuera de control*, es decir, que no corresponde a lo que sería esperable por azar, si se mantiene en la realidad el mismo valor medio. Revisaremos estos patrones más adelante; insistimos de nuevo en que en ninguno de los casos hay que hacer cálculo alguno en la práctica, ya que se han hecho previamente y se sabe que son patrones anormales, en virtud de la distribución de probabilidades que deben seguir los valores del indicador, cuando la situación es estable.

No todos los gráficos de control estadístico tienen como base la distribución normal; pero la forma de proceder y el razonamiento es idéntico en todos los casos: buscar el valor que represente la media general del indicador que medimos y ver la probabilidad de obtener los diversos valores, individualmente y en secuencia, que se obtengan con las mediciones sucesivas del indicador. En todos los casos, una vez construida la plantilla, la decisión sobre la presencia o ausencia

Una determinada secuencia de valores en las mediciones de un indicador puede ser también un signo de inestabilidad estadística del proceso que se analiza o monitoriza.

Aunque los patrones de *descontrol o inestabilidad* problemáticos pueden calcularse con base en la función de probabilidades correspondiente, en la práctica no es necesario realizar cálculo probabilístico alguno, el análisis es simplemente visual en busca de estos patrones que se conocen de antemano.

Sea cual fuere la distribución de probabilidad que corresponda, el aspecto y la interpretación de los gráficos de control estadístico son idénticos. La única diferencia es la forma de calcular el promedio y los límites para la plantilla de referencia.

El valor medio (línea central) y los límites de referencia se calculan y construyen, por lo menos, a partir de 20 mediciones sucesivas del indicador por analizar o monitorizar.

de causas *especiales* (proceso *fuera de control*, inestable, con media diferente a la que estamos utilizando en la monitorización) se toma por simple inspección visual. En realidad, sea cual fuere la distribución o función de probabilidad sobre la que se base la figura (normal, binomial, Poisson, no paramétrica), Shewart comprobó, y en ello se basó para tomar las decisiones sobre la inestabilidad de los procesos, que se cumple la llamada desigualdad de Tchebyshev. Según ésta, la probabilidad de que la medición de una variable se desvíe de su media más de k veces, su desviación estándar es igual o menor de $1/k^2$, sea cual fuere la distribución probabilística de base. Lo que cambia al aplicar diversas distribuciones de probabilidad en el control gráfico es la forma de calcular las desviaciones estándar y los valores promedio, pero la plantilla y la interpretación visual de la figura son idénticas en todos los casos.

4. Construcción de la plantilla de referencia

¿Cómo calculamos el valor medio y los límites de referencia para construir la plantilla sobre la que se grafican las mediciones y saber si la situación está o no *bajo control*? El conocimiento de estas características del proceso por monitorizar se realiza con una serie de mediciones sucesivas que utilizamos para estimar la media y desviación estándar, que nos sirven de base para construir la plantilla del gráfico y realizar los análisis o monitorizaciones posteriores. En estadística, las estimaciones de la media de una variable cuantitativa mediante muestra se anotan como \bar{x} , y la desviación estándar como s . Más adelante veremos que, en el control gráfico, para obtener el promedio de la plantilla (y, por lo tanto, para tomar las decisiones sobre la estabilidad del indicador que se mide), se utiliza la media de medias ($\bar{\bar{x}}$) o su equivalente, según el tipo de variable que mide el indicador, y la desviación estándar de la media ($S_{\bar{x}}$) para los límites de control.

¿Cuántas mediciones hacen falta para poder estimar \bar{x} y $S_{\bar{x}}$? La mayoría de los manuales aconsejan realizar, por lo menos, entre 20 y 30 muestras o mediciones sucesivas del indicador para construir la plantilla del gráfico. Ésta es una de las razones por las que no sería posible utilizar los gráficos de control estadístico para indicadores que no se midan frecuentemente. Adicional-

mente, para detectar situaciones fuera de control en la monitorización, si éstas aparecen, no como valores extremos, sino como secuencia atípica, nos haría falta en determinadas figuras un mínimo de cinco mediciones seguidas. ¿Sería posible esperar cinco años o cinco semestres para detectar una situación problemática? Cuando Shewart propuso las figuras de control estadístico como método para detectar problemas de la forma más inmediata posible en los procesos de producción, llegó a proponer, realizar y analizar determinados indicadores con una periodicidad de una hora entre mediciones. Sin llegar a estos extremos, el control estadístico de la calidad puede aplicarse con provecho en los servicios de salud en mediciones que pueden ser hasta mensuales, como veremos a continuación.

5. Análisis gráfico de la variabilidad: los cuatro tipos principales de gráficos

Aunque todos responden a un mismo esquema metodológico básico, existe gran cantidad de tipos diferentes de gráficos para el control estadístico de la calidad, y continuamente siguen apareciendo nuevas modalidades. De hecho, el control estadístico de procesos (Statistical Process Control, SPC, definido como la aplicación de métodos estadísticos en la medición y análisis de la variación) en el cual se encuadra, y es principal protagonista, el análisis gráfico, es prácticamente una disciplina en sí mismo dentro del amplio campo de la gestión de la calidad. En algunos manuales de SPC se propone una serie de clasificaciones para los diferentes tipos de gráficos, como la que se recoge en el cuadro 4.2; pero quizá lo más útil en términos prácticos sea la consideración del tipo de gráficos, no en función del tipo de variable, sino de una complejidad creciente en cuanto a su construcción, sea cual fuere el tipo de variable que se mida, además de unas posibilidades de utilización también diferentes. Distinguiremos cuatro grupos de gráficos: de desarrollo, de control, de capacidad de proceso y de cálculos secuenciales.

1. *Gráficos de desarrollo*, también llamados de *rachas* o, en algunos países latinoamericanos, de *corridos*, con una traducción literal pero confusa del original en inglés (*run chart*). Estos gráficos únicamente tienen como referencia el valor medio del indicador; por lo tanto, son los más sencillos de

Existe una gran cantidad de gráficos para el llamado control estadístico de procesos, habitualmente clasificados según el tipo de variable que se analiza.

Cuadro 4.2
Gráficos de control de uso más frecuente en la industria.
Clasificación según el tipo de dato que se mide

| Gráficos sobre indicadores que miden variable cuantitativa | Gráficos sobre indicadores que miden variable cualitativa |
|---|--|
| Gráficos de valores medios \bar{x} , gráfico de medias x , gráfico de valores individuales \tilde{x} , gráfico de medianas NLG, gráfico de límites estrechos (precontrol) CUSUM, gráfico de suma acumulada de medias EWMA, gráfico de media móvil, exponencialmente ponderada | Gráficos sobre mediciones de una variable dicotómica p , gráfico de proporciones np , gráfico de número de defectos CUSUM, gráfico de suma acumulada de desviaciones |
| Gráficos de medidas de dispersión R , gráfico de rangos S , gráfico de dimensiones MR , gráfico de rangos móviles NLG , gráfico de límites estrechos (precontrol) | Gráficos sobre mediciones de más de una variable dicotómica u , gráfico de medias de defectos por unidad de estudio c , gráfico de número de defectos CUSUM, gráfico de suma acumulada de desviaciones. |

construir e interpretar, pero su utilidad se reduce al análisis de tendencias y detecta cambios por medio de una determinada secuencia de valores (en un mismo lado de la media, continuamente ascendentes o descendentes, etc.). Nos ocuparemos de ellos más adelante.

2. *Gráficos de control*. En éstos, además del promedio, como en los gráficos de desarrollo, se representan los llamados *límites de control superior e inferior* (LCS y LCI), que normalmente señalan el promedio más tres desviaciones estándar (LCS, límite de control superior) y el promedio menos tres desviaciones estándar (LCI, límite de control inferior). Suelen establecerse habitualmente otras líneas de referencia marcando una y dos desviaciones estándar por encima y por debajo del promedio; esta subdivisión facilita la aplicación de patrones adicionales de distribución de las mediciones en el gráfico que son significativas de *descontrol* (problemas), aumentando la sensibilidad del gráfico para la identificación de situaciones problemáticas. Los gráficos de control son tema del capítulo 5.

3. *Gráficos de capacidad de proceso*. Con ellos se analiza la relación entre la variabilidad estadística normal (controlada o estable) del indicador y las especificaciones que debe tener para estar en el nivel de calidad deseado. Para ello se hacen inter-

venir otros límites, los de especificación superior e inferior (LES y LEI), contra los que se compara la distribución de valores del indicador en el gráfico de control en situación estable, controlada (una vez eliminadas todas las causas especiales, no aleatorias o asignables). Esta comparación entre los límites de especificación y los valores del indicador, dentro de los límites de control estadístico, se realiza con base en una serie de índices, cuyos valores indican cuál sería la mejor decisión para lograr que el proceso analizado, aparte de ser estadísticamente estable, produzca el nivel de calidad óptimo. Los gráficos de capacidad de proceso son esenciales para las actividades de diseño de procesos y se revisan en el capítulo 6.

4. *Gráficos de control secuenciales*, estadísticamente más elaborados. Incluimos aquí un serie de gráficos, como los CUSUM (*Cummulative sum*, suma acumulada) y EWMA (*Exponentially weighted moving average*, media móvil exponencialmente ponderada), que incluyen para su aplicación una recalculación continua de los límites y el promedio con base en los valores previos del indicador. Lo que se considera en el gráfico es la diferencia con valores anteriores, en vez de la medición del indicador en sí. Con ello, al parecer, se consigue detectar desviaciones de patrones de estabilidad con mayor prontitud y exactitud. Este tipo de gráficos se revisan en el capítulo 7.

En nuestro caso, vamos a distinguir cuatro tipos principales de gráficos, en función de su complejidad y utilidad: de desarrollo (o rachas), de control, de capacidad de proceso y de cálculos secuenciales.

Tanto los gráficos de desarrollo como los de control, reciben diversos nombres (x , \bar{x} , p , u) según el tipo de variable que miden y resumen. No hay sin embargo diferencias ni en su aspecto ni en las normas de interpretación.

6. El control estadístico de procesos en la práctica: construcción e interpretación de los gráficos de control estadístico más sencillos

Todo aspecto cuantificable y medido con frecuencia es susceptible de ser analizado y monitorizado con un gráfico de control estadístico. Adicionalmente, aunque se utilicen distintas distribuciones de probabilidad para establecer los parámetros de referencia, la construcción, visualización e interpretación del análisis gráfico es semejante en todos ellos. Por ello los distintos gráficos que vamos a revisar parecerían uno solo si no indicamos cuál es la variable que se mide y cómo se mide. En este capítulo, detallaremos la construcción y uso de los *gráficos de desarrollo*, y en el siguiente, los *gráficos de control* que responden a los cuatro principales tipos de mediciones o indicadores (números absolutos, medias, proporciones y ratios) que pueden realizarse en nuestro medio.

Según se resume en el cuadro 4.3, los diferentes tipos de gráfico (diferentes por el tipo de medición que se representa, no por el aspecto del gráfico) pueden basarse en la distribución nor-

mal, binomial o, incluso, en ninguna en concreto (no paramétrica), en función de que midamos medias, proporciones o números absolutos. De la misma manera, los cuatro tipos de gráficos de control seleccionados, que vemos en detalle en el capítulo 5, se basan en diferentes distribuciones de probabilidad, en función de las características del indicador al que se aplican. No hay, en cambio, ninguna diferencia en el aspecto, ni en las normas para su interpretación, ni, como veremos, en la forma de construirlos.

Ilustremos los procedimientos de construcción e interpretación de los gráficos de control estadístico, aplicándolos en cuatro situaciones diferentes.

En la primera de ellas, queremos controlar el tiempo que dedicamos a las consultas a demanda, no programadas. Hace tiempo que nos propusimos como objetivo de calidad que nos dé tiempo, por lo menos, de saludar correctamente al paciente; dar lugar a que nos explique por qué viene a la consulta, y conducir una mínima entrevista clínica. Nuestro objetivo es una media de 10 minutos; pero sabemos que son varios los factores que pueden justificar dedicarle más tiempo o menos, de

Cuadro 4.3
Tipos de gráficos de control estadístico más comunes, indicadores para los que se utilizan y distribución de probabilidad en que se basan

| Tipo de gráfico | Tipo de indicador | Distribución de probabilidades |
|--|---|-----------------------------------|
| Gráfico de desarrollo | Media, proporción o número absoluto del aspecto por monitorizar | Normal, binomial o no paramétrica |
| Gráfico "x" de valores individuales | Número de eventos por unidad de tiempo (ejemplo: visitas/día) | Normal |
| Gráfico " \bar{x} " o de medias (acompañado o no de rangos o de desviaciones estándar) | Media de una variable cuantitativa por unidad de estudio (ejemplo: media de minutos/visita) | Normal |
| Gráfico "p" o de proporciones | Proporción de una cualidad (cumple/no cumple) en una muestra de casos (ejemplo: proporción de visitas que esperaron más de 30') | Binomial |
| Gráfico "u" o de ratios | Ratio de una (o varias) cualidades por unidad de estudio (ejemplo: núm. de factores de riesgo indagados/caso evaluado) | Poisson |

manera que podamos controlar hasta qué punto la variabilidad es estable y se mantiene en torno a los valores esperados. En el cuadro 4.4 están los resultados de una monitorización realizada en 25 días sucesivos. En cada día se ha medido el tiempo de consulta en cinco pacientes elegidos al azar, sin que se supiese a quién le tocaría. El indicador mide una variable *cuantitativa* (tiempo en minutos) y se expresa como **media** de minutos/visita. ¿Cuál es nuestro promedio?, ¿se mantiene estable?, ¿debemos hacer algo para mejorar?

En la segunda situación, el aspecto que nos preocupa es el control de la espera excesiva, des-

pués de que hemos puesto en marcha sistemas organizativos de cita previa más ágiles y flexibles, pensados para que las visitas de quienes vienen a consultarnos sean fluidas, sin aglomeraciones y con poco tiempo de espera. Establecemos como indicador del funcionamiento de nuestro sistema la **proporción** de pacientes que tienen que esperar más de 30 minutos para ser atendidos. Hemos realizado la medida de este indicador durante un mes (un total de 22 mediciones) en una muestra diaria de 50 pacientes; considerando que toda persona que haya esperado más de 30 minutos incumple nuestro indicador de calidad (cuadro 4.5).

Cuadro 4.4

Indicador: media de minutos por consulta a demanda.
Resultado de 25 mediciones en días sucesivos

| Día | Tamaño de la muestra | Media |
|-------|----------------------|-------|
| 1 | 5 | 10.2 |
| 2 | 5 | 8.4 |
| 3 | 5 | 7.6 |
| 4 | 5 | 7.0 |
| 5 | 5 | 8.0 |
| 6 | 5 | 10.2 |
| 7 | 5 | 7.4 |
| 8 | 5 | 8.0 |
| 9 | 5 | 10.8 |
| 10 | 5 | 6.0 |
| 11 | 5 | 9.2 |
| 12 | 5 | 8.2 |
| 13 | 5 | 6.0 |
| 14 | 5 | 10.8 |
| 15 | 5 | 11.4 |
| 16 | 5 | 5.2 |
| 17 | 5 | 8.6 |
| 18 | 5 | 10.0 |
| 19 | 5 | 6.6 |
| 20 | 5 | 9.3 |
| 21 | 5 | 9.1 |
| 22 | 5 | 8.2 |
| 23 | 5 | 8.1 |
| 24 | 5 | 8.0 |
| 25 | 5 | 7.9 |
| Total | | 210.2 |

Cálculos para la línea promedio según las fórmulas del cuadro 4.8 para gráfico de medias:

$$\bar{X} = \frac{210.2}{25} = 8.4$$

Cuadro 4.5

Indicador: proporción de pacientes que tienen que esperar más de 30' antes de ser atendidos. Resultados de las mediciones efectuadas en todos los días de un mes

| Muestra número | Tamaño de muestra (n) | Número de incumplimientos | Proporción de incumplimientos |
|----------------|-----------------------|---------------------------|-------------------------------|
| 1 | 50 | 12 | 0.24 |
| 2 | 50 | 15 | 0.30 |
| 3 | 50 | 8 | 0.16 |
| 4 | 50 | 10 | 0.20 |
| 5 | 50 | 4 | 0.08 |
| 6 | 50 | 7 | 0.14 |
| 7 | 50 | 16 | 0.32 |
| 8 | 50 | 9 | 0.18 |
| 9 | 50 | 14 | 0.28 |
| 10 | 50 | 10 | 0.20 |
| 11 | 50 | 5 | 0.10 |
| 12 | 50 | 6 | 0.12 |
| 13 | 50 | 17 | 0.34 |
| 14 | 50 | 12 | 0.24 |
| 15 | 50 | 22 | 0.44 |
| 16 | 50 | 8 | 0.16 |
| 17 | 50 | 10 | 0.20 |
| 18 | 50 | 5 | 0.10 |
| 19 | 50 | 13 | 0.26 |
| 20 | 50 | 11 | 0.22 |
| 21 | 50 | 20 | 0.40 |
| 22 | 50 | 18 | 0.36 |
| Total | 1.100 | 252 | |

Cálculos para la línea promedio según las fórmulas del cuadro 4.8 para gráfico de proporciones:

$$\bar{p} = \frac{252}{1\ 100} = 0.23$$

En este capítulo y en el siguiente, utilizaremos como ejemplo cuatro tipos de indicadores diferentes que se corresponden con cuatro tipos de gráficos y distribuciones de probabilidad: media (gráfico \bar{x} , distribución normal), proporción (gráfico p , distribución binomial), valores absolutos (gráfico x , no paramétrico o normal) y ratio (gráfico u , distribución de Poisson).

Nuestro tercer ejemplo es la monitorización del funcionamiento de un programa nuevo que no tienen todos los centros, pero que nosotros, que tenemos una relación estupenda con nuestro hospital de referencia, hemos logrado poner en marcha y estamos orgullosos de ello. Se trata de la realización en nuestro centro de primer nivel de intervenciones de cirugía menor, que normalmente irían a engrosar la lista de espera de los cirujanos del hospital. No son muchas las intervenciones que podemos realizar; pero queremos mostrar que estamos utilizando de forma regular la formación recibida y el equipamiento adquirido expresamente para este programa. Medimos el **número** de intervenciones realizadas por semana (no parece posible un indicador tipo media o proporción). El cuadro 4.6 contiene los resultados de las últimas 20 semanas. ¿Cómo podemos analizarlos?, ¿estamos actuando de forma regular y constante?, ¿cómo podemos monitorizar el funcionamiento del programa?

El cuarto ejemplo es más complicado. Tiene que ver con el control de una mejora importante, conseguida por medio de la relación con los compañeros de otro centro, servicio o nivel asistencial. Nos hemos puesto de acuerdo en qué información necesitan de nosotros cuando les remitimos un paciente, y qué esperamos nosotros de ellos cuando nos lo devuelven una vez cumplido su cometido. Tenemos claro el concepto de *cliente interno* y sabemos que ambos tenemos que estar atentos a las necesidades y expectativas del otro para ofrecer un servicio de calidad, y que éste redunde en una mejor atención a nuestros pacientes. Nos hemos comprometido a informarles de cuatro puntos concretos cada vez que remitimos un paciente: datos de filiación, concisa descripción de antecedentes y exploraciones realizadas, sospecha diagnóstica y tratamientos prescritos. Consideramos la falta de cualquiera de los puntos mencionados un defecto de calidad igualmente importante, de forma que monitorizamos el **número de defectos por unidad de estudio**, siendo esta última cada uno de los documentos interconsulta en los que hemos evaluado el cumplimiento de los cuatro criterios. Realizamos la medición de una muestra diaria de 10 documentos. Si los resultados son satisfactorios, pensamos monitorizar con mediciones semanales o quincenales hasta que estemos seguros de que esta forma de relacionarse se ha convertido en rutina. Sin embargo, no sabemos muy bien cuál sería el mejor

| Semanas | Número de intervenciones |
|--------------|--------------------------|
| 1 | 13 |
| 2 | 7 |
| 3 | 10 |
| 4 | 8 |
| 5 | 11 |
| 6 | 12 |
| 7 | 13 |
| 8 | 12 |
| 9 | 12 |
| 10 | 8 |
| 11 | 9 |
| 12 | 5 |
| 13 | 7 |
| 14 | 9 |
| 15 | 9 |
| 16 | 5 |
| 17 | 7 |
| 18 | 8 |
| 19 | 10 |
| 20 | 11 |
| Total | 186 |

Cálculos para la línea promedio según las fórmulas del cuadro 4.8 para gráfico de valores absolutos:

$$\text{Mediana: } 9 \text{ o } \bar{x} = \frac{186}{20} = 9.3$$

mecanismo para realizar esta monitorización. ¿Es aplicable algún gráfico de control estadístico a este tipo de mediciones?, ¿hay que hacer un gráfico para cada criterio? Los datos de 20 muestras sucesivas de 10 interconsultas figuran en el cuadro 4.7.

Así pues, nuestras cuatro situaciones utilizan cuatro tipos de mediciones diferentes como indicador: una **media** (minutos por consulta), una **proporción** (pacientes que esperan más de 30 minutos), **números absolutos** por unidad de tiempo (intervenciones quirúrgicas por semana), y una **ratio** de variables categóricas por unidad de estudio (número de defectos por documento interconsulta). Veamos cómo se construyen y utilizan para cada uno de ellos los correspondientes gráficos de control estadístico, empezando por

Cuadro 4.7

Indicador: número de defectos (criterios incumplidos) por documento de interconsulta. Resultados de 20 mediciones sucesivas en muestras de 10 documentos

| Medición | Tamaño de muestra (n) | Núm. total de defectos | Ratio |
|----------|-----------------------|------------------------|-------|
| 1 | 10 | 4 | 0.4 |
| 2 | 10 | 1 | 0.1 |
| 3 | 10 | 4 | 0.4 |
| 4 | 10 | 2 | 0.2 |
| 5 | 10 | 3 | 0.3 |
| 6 | 10 | 10 | 1.0 |
| 7 | 10 | 0 | 0.0 |
| 8 | 10 | 4 | 0.4 |
| 9 | 10 | 11 | 1.1 |
| 10 | 10 | 5 | 0.5 |
| 11 | 10 | 2 | 0.2 |
| 12 | 10 | 5 | 0.5 |
| 13 | 10 | 1 | 0.1 |
| 14 | 10 | 12 | 1.2 |
| 15 | 10 | 3 | 0.3 |
| 16 | 10 | 3 | 0.3 |
| 17 | 10 | 11 | 1.1 |
| 18 | 10 | 4 | 0.4 |
| 19 | 10 | 0 | 0.0 |
| 20 | 10 | 3 | 0.3 |
| Total | 200 | 88 | |

Cálculos para la línea según las fórmulas del cuadro 4.8 para gráfico de ratios:

$$u = \frac{88}{200} = 0.44$$

los gráficos de desarrollo en este capítulo, y empleando gráficos de control en el siguiente.

7. Gráficos de desarrollo

Para la construcción de un gráfico de desarrollo lo único que se necesita es calcular la línea central o *promedio*, que es la que nos servirá de referencia para interpretar las mediciones del indicador.

Para calcular el *promedio* necesitamos de 20 a 30 mediciones del indicador, las cuales se promedian de la siguiente manera, siguiendo las fórmulas indicadas en el cuadro 4.8, según midamos medias, proporciones, valores absolutos o ratios:

7.1. Gráfico de desarrollo para medias

Si lo que medimos son *medias* (por ejemplo: la media de minutos por visita a demanda, el indicador

de nuestro primer problema) se calcula la *media de medias* ($\bar{\bar{x}}$) de las 25 mediciones que hemos hecho del indicador; $\bar{\bar{x}}$ es la suma de la media obtenida (valor del indicador) en cada una de las mediciones, dividida entre el número de mediciones (en nuestro caso 25), si todas las mediciones las hemos hecho con un mismo tamaño de muestra. En el cuadro 4.4 se encuentra este cálculo. El promedio (la media de medias) de nuestra serie de mediciones es 8.4, por lo tanto, debemos situar la línea central del gráfico en este nivel para, a continuación, colocar la secuencia de mediciones efectuadas (figura 4.4).

En caso de que el tamaño de la muestra no fuera constante, habría que calcular el promedio de forma ponderada (multiplicando cada media por el tamaño de su muestra; sumar los resultados de estas multiplicaciones, y dividirlo por el total general de casos: $\frac{\sum(\bar{x} \cdot n_i)}{\sum n_i}$). Éste sería el caso, por ejemplo, de haber medido diariamente el

Para el gráfico de desarrollo o de rachas lo único que hay que calcular es el promedio para ubicar la línea central de la plantilla.

Cuadro 4.8
Fórmulas para el cálculo de la línea central

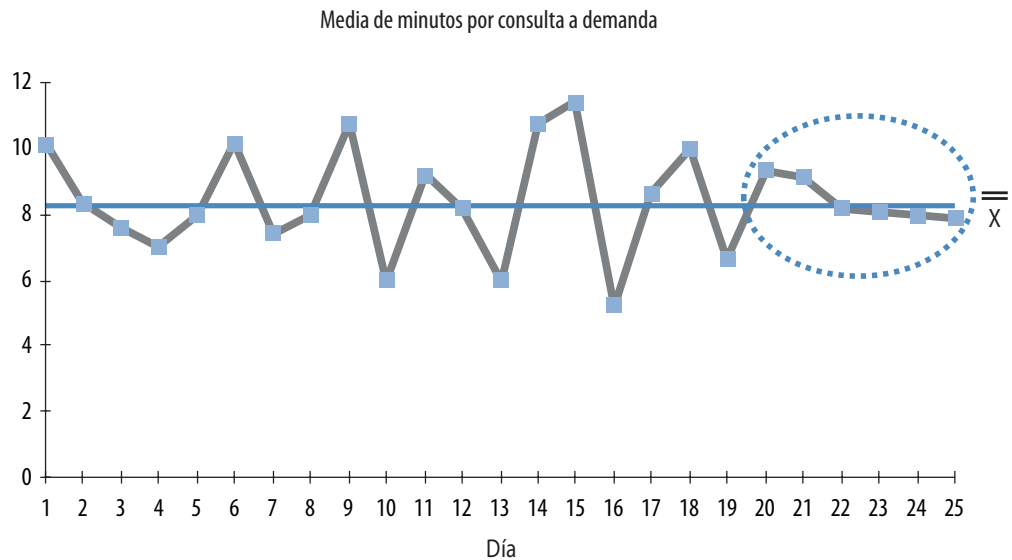
| Gráfico | Línea central (promedio) |
|------------------------------|--|
| X valores individuales | $\bar{x} = \frac{\text{núm. total de eventos en el periodo estudiado}}{\text{núm. mediciones realizadas}}$ (Ejemplo: para el indicador intervenciones/día, medido 20 días: $\frac{\text{núm. total de eventos en el periodo estudiado}}{\text{núm. mediciones realizadas}}$) |
| $\bar{\bar{x}}$ (medias) | $\bar{\bar{X}} = \text{Media de las medias de cada medición del indicador en la serie que analizamos o utilizamos para construir el gráfico} = \frac{\sum \bar{X}_i}{k}$ |
| p (proporciones) | $\bar{P} = \text{Proporción media de la serie de mediciones analizada o utilizada para construir el gráfico} = \frac{\sum p_i \cdot n_i}{\sum n_i}$ |
| u (ratios) | $\bar{U} = \frac{\text{Total ocurrencias de la cualidades que se miden}}{\text{Total de casos evaluados}} = \frac{\sum c_i}{\sum n_i}$ |

ni : tamaño de la muestra en cada medición.

k : núm. de mediciones (muestras) realizadas para analizar el indicador o construir el gráfico.

c : cumplimientos (o incumplimientos) de diversos criterios, evaluados todos ellos simultáneamente en cada unidad de estudio.

i : secuencia ordinal de las mediciones efectuadas (1, 2, 3, ...i).



*: datos del cuadro 4.4

Figura 4.4. Gráfico de rachas (run chart) para el indicador "media de minutos por consulta a demanda"^{ix}

tiempo de consulta en todos los pacientes, en vez de hacerlo en una muestra, porque lo lógico es que el número total diario (equivalente en concepto a la muestra diaria) sea diferente cada día.

7.2. Gráfico de desarrollo para proporciones

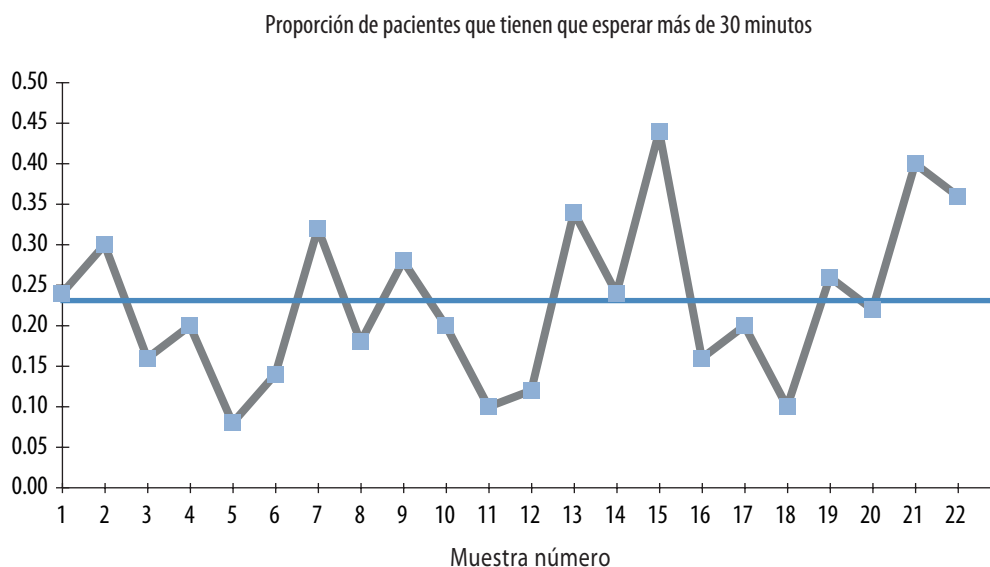
Si lo que se mide son proporciones, calculamos igualmente la proporción media (\bar{p}) de todas las mediciones efectuadas. Éste es el caso de nuestro segundo ejemplo en el que monitorizamos la proporción de pacientes que esperan más de 30 minutos para ser atendidos, para lo cual hemos medido esta característica (esperar más de 30 minutos o no) en una muestra de 50 pacientes durante un mes (22 días). Como el tamaño de la muestra es constante, la proporción media sería la suma de todas las proporciones encontradas, dividido por el número de muestras o, lo que resulta incluso más cómodo de calcular, el promedio *conjunto* (total general de incumplimientos dividido por total de casos evaluados). Si el tamaño de la muestra no fuera siempre el mismo, habría que calcular el promedio *conjunto* (considerar todas las muestras como si fuese una sola y contar el total de incumplimientos para dividirlo entre el total de casos evaluados). En nuestro ejemplo (cuadro 4.5) tenemos un promedio de 23%. A este nivel ponemos

la línea central, y sobre la plantilla así construida colocamos las 22 mediciones (figura 4.5).

7.3. Gráfico de desarrollo para valores absolutos

Si nuestro indicador es el total de casos por cada unidad de tiempo en la que hacemos la medición, como en el ejemplo del programa de cirugía menor (cuadro 4.6), calcularemos el promedio de todas las mediciones, o bien la mediana, lo cual sería más adecuado como referencia para interpretar los datos con prueba no paramétrica. La mediana (\tilde{X}) es el valor central de la serie, que deja igual número de valores por encima y por debajo. Para ello, hay que ordenar las mediciones desde el valor más pequeño (en nuestro ejemplo, cinco intervenciones) hasta el más grande (13 intervenciones) y buscar el valor central. Si el número de mediciones fuera impar, la mediana sería el valor que ocupa el valor central directamente; si la serie de mediciones es par (como en nuestro caso) la mediana es la media aritmética de los dos valores centrales. En nuestro caso, los dos valores centrales de la serie, ya ordenada de menor a mayor, son los dos números nueve, de forma que la mediana es nueve, la cual sería la línea central del gráfico (figura 4.6).

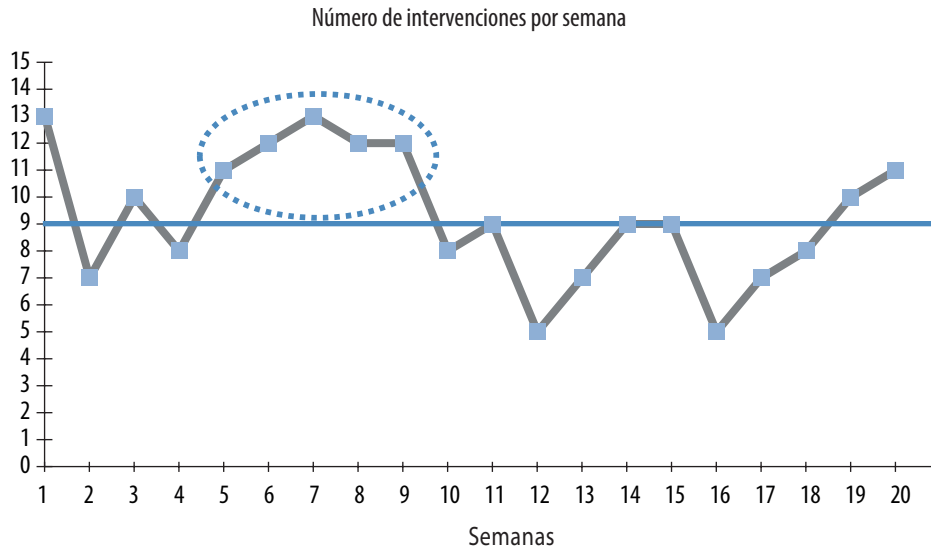
El promedio se calcula con base en 20-30 mediciones sucesivas del indicador, y según la fórmula que corresponda al tipo estadístico de indicador (media, proporción, valor absoluto o ratio).



*: datos del cuadro 4.5

Figura 4.5. Gráfico de rachas (*run chart*) para el indicador "proporción de pacientes que tienen que esperar más de 30 minutos"¹⁸

En los gráficos de desarrollo se aprecian, sobre todo, tendencias significativamente diferentes de la situación de estabilidad.



*: datos del cuadro 4.6

Figura 4.6. Gráfico de rachas (*run chart*) para el indicador "número de intervenciones por semana"*

Los patrones de inestabilidad o de diferencias significativas con el promedio los constituye una serie de puntos por encima o por debajo del promedio, en continuo ascenso o descenso.

7.4. Gráfico de desarrollo para ratios (nº de defectos por unidad de estudio)

En nuestra monitorización de la calidad de la información del documento interconsulta (cuadro 4.7), calculamos el promedio para la línea central del gráfico como la media que resulta de sumar el número total de defectos encontrados y dividirla entre el total de casos (suma de todas las muestras). Para efectos prácticos, es como si fuese un gráfico de medias, si bien la variable que se mide no es cuantitativa continua (como los minutos por visita), sino un conjunto de variables dicotómicas (cumple/no cumple) evaluadas todas a la vez en cada unidad de estudio. Ello hace que la distribución de probabilidades que utilicemos como referencia para analizar los datos no sea la normal, sino la de Poisson, pero en la práctica este matiz pasa inadvertido.

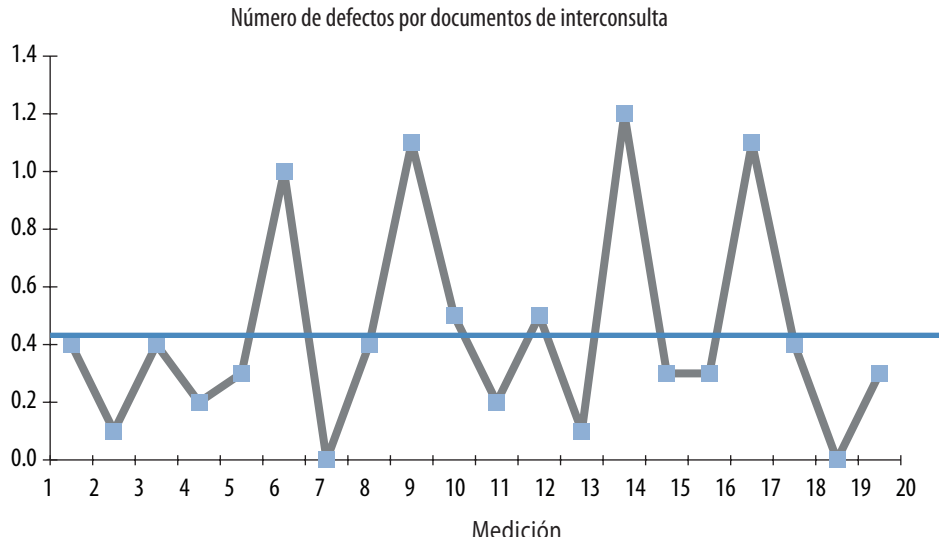
La misma situación, el mismo tipo de indicador, se da aunque sólo sea un criterio de calidad el que midamos, siempre que éste sea categórico (cumple/no cumple) y exista la posibilidad de que puedan aparecer más de un defecto (incumplimiento) en la unidad de estudio. Por ejemplo: números de errores tipográficos por hoja, número de palabras confusas o ilegibles por informe clínico, etc. (figura 4.7).

8. Interpretación de los gráficos de desarrollo

Una vez calculado el promedio y construida la *plantilla* del gráfico de desarrollo, podemos *interpretar los resultados* de nuestras mediciones. Conviene recordar aquí que esta *plantilla* puede utilizarse tanto para analizar la variabilidad de los datos como para monitorizar futuras mediciones (esto último a condición de que no se aprecien causas especiales de variación en el análisis, y la estabilidad estadística incluya conformidad con las especificaciones de calidad).

En nuestros cuatro ejemplos vamos a utilizar el gráfico inicialmente para el primero de los tres usos que apuntábamos en el primer apartado de este capítulo: *analizar la serie de datos* que hemos obtenido para cada indicador. Para ello, hemos llevado al gráfico en su secuencia concreta todas las mediciones efectuadas con el resultado que aparece en las figuras 4.4 a 4.7, y vamos a observar su distribución secuencial en busca de tendencias significativas. Los gráficos de desarrollo sirven básicamente para analizar y detectar tendencias.

Las secuencias o tendencias significativas de cambios o inestabilidad en el proceso son las que figuran en el cuadro 4.9. Obsérvese que son igua-



*: datos del cuadro 4.7

Figura 4.7. Gráfico de rachas (*run chart*) para el indicador "número de defectos por documento de interconsulta"*

| Cuadro 4.9 Interpretación de los gráficos de desarrollo: patrones que denotan situaciones significativamente ($p < 0.01$) diferentes de lo esperable por azar | | | | |
|--|---|---|--|--|
| Tipo de indicador | Patrón significativo de variación | | | |
| Media, proporción, ratio | — Siete o más puntos consecutivos a un mismo lado (por encima o por debajo) del promedio. — Seis o más puntos consecutivos en ascenso o en descenso. | | | |
| | Depende del número de mediciones (puntos) representados en el gráfico: | | | |
| | Número de mediciones | Puntos consecutivos a un mismo lado de la mediana | Puntos consecutivos en ascenso o en descenso | Número mínimo de rachas* a un mismo lado de la mediana |
| Números absolutos | 20 | ≥ 8 | ≥ 7 | 4 |
| | 30 | ≥ 9 | ≥ 7 | 8 |
| | 40 | ≥ 10 | ≥ 7 | 12 |
| | 50 | ≥ 11 | ≥ 7 | 16 |

*: racha: conjunto de puntos consecutivos a un mismo lado de la mediana.

Fuente: Elaboración propia a partir de Pyzdek T y Farnum.

les para los indicadores que se expresan como media, proporción o ratio, mientras que, para los números absolutos, se toma la mediana como referencia y se aplican pruebas no paramétricas.

Si observamos los gráficos (figuras 4.4 a 4.7) buscando la existencia de alguno de estos patrones, se pone en evidencia una situación pro-

blemática en el indicador relativo a la media de minutos por consulta: los últimos seis puntos en constante descenso. Esta situación es clara y significativamente diferente del promedio y merece que investiguemos qué está sucediendo. La aparición de un patrón anormal suele marcarse con una cruz (x) en el gráfico.

Cuando el indicador es de números absolutos, aunque puede calcularse la media, lo más apropiado es utilizar como promedio la mediana y buscar patrones significativos de inestabilidad por medio del test de rachas, de naturaleza no paramétrica.

Cuando se utilizan los gráficos de desarrollo hay que procurar interpretarlos en su justa medida: esencialmente no dar significación a puntos aislados ni a patrones diferentes de los previamente conocidos como significativos.

Los gráficos de desarrollo son muy sencillos de construir e interpretar, pero son menos sensibles que los gráficos de control para detectar situaciones problemáticas.

En los otros gráficos no se observa ningún patrón anormal de los que figuran en el cuadro 4.9. Sin embargo, merece la pena que nos detengamos en la aplicación de la *prueba de rachas* en el gráfico de números absolutos, porque no es tan evidente como las otras. Hay que realizar los siguientes pasos:

1. Contar el número de puntos por encima y por debajo de la mediana, y los que están justo en la mediana. En nuestro gráfico hay nueve puntos por encima, ocho por debajo y tres en la mediana.

Hay que asignar los puntos que están en la mediana al grupo de los de arriba y los de abajo, de forma que quede un número igual de puntos a cada lado. En nuestro ejemplo habría que asignar el número uno a los de arriba y dos a los de abajo, de forma que queden 10 y 10. Si el número de puntos fuese impar, se descartaría para la prueba de rachas uno de los puntos situado justo en la mediana.

2. Se cuentan el número de *rachas* (puntos seguidos a un mismo lado de la mediana) que hay por encima y por debajo de la mediana en todo el gráfico. Los puntos que están justo en la mediana y que hemos asignado se escogen de uno u otro lado, para maximizar el número de rachas (si está a continuación de una racha de lado de abajo, se asigna a los de arriba y viceversa). Las rachas pueden ser únicamente de un punto. En nuestro gráfico hay cinco rachas en el lado de arriba y cuatro en el de abajo, contando con la asignación de los puntos que están en la mediana.
3. Se observa en los cuadros (4.9) si el número de rachas más pequeño de los dos (en nuestro caso, las cuatro rachas del lado de abajo) es igual o mayor que el mínimo esperable. Si resulta *menor*, es un patrón que indica inestabilidad. En el cuadro 4.9 podemos ver que

lo mínimo esperable para un gráfico con 20 puntos son precisamente cuatro rachas, por lo tanto, concluiremos que la evolución de este indicador está dentro de la variabilidad esperable por el azar en un proceso estable.

Realizar el análisis gráfico con gráficos de desarrollo tiene la ventaja de la simplicidad, pero también tiene otras desventajas y peligros que son:

Peligros: (i) que intentemos darle importancia a los *picos* en el gráfico, cuando sólo sirve para detectar tendencias, y (ii) que no creamos que las tendencias existen y son significativas, y esperemos tener *un punto más* para confirmarlo; en este caso, como la nueva situación también tendrá su propia variación aleatoria, pueden aparecer mediciones que produzcan el espejismo de haber vuelto a la normalidad, y el problema queda sin identificar y puede agravarse o repetirse en el futuro.

Desventaja: es bastante menos sensible que los gráficos con límites de control tanto para analizar la variabilidad de los datos como para monitorizar cambios. De hecho hay manuales que no incluyen los gráficos de desarrollo entre el arsenal de gráficos para el control estadístico de la calidad. Sin embargo, sería tremendamente útil para nuestro sistema de salud, su aplicación rutinaria y automatizada en todos los indicadores rutinarios y “obligatorios” que se llevan a cabo o, cuando menos, mucho más informativo que los listados de números y otras representaciones gráficas más comunes.

El anexo 4.1 describe de forma resumida el procedimiento para el análisis de una serie de mediciones de un indicador, utilizando un gráfico de desarrollo.

BIBLIOGRAFÍA

- Deming WE. Causas comunes y causas especiales de la mejora. Sistema estable. En: Calidad, productividad y competitividad. La salida de la crisis. Madrid: Díaz de Santos, 1989.
- Farnum NR. Control chart concepts. En: Modern statistical quality control and improvement. Belmont: Duxbury Press, 1994.
- Methuen MA. Manual de herramientas básicas para el análisis de datos. GOAL/QPC, 1990.
- Ott ER, Schilling EG. Ideas from time sequences of observations. En: Process quality control. 2ª ed. Nueva York, NY: McGraw Hill, 1990.
- Plsek PE. Tutorial: introduction to control charts. Quality Management In Health Care 1992;1(1):65-74.
- Pyzdek T. An SPC Primer. Tucson, AZ: Quality Publishing Inc., 1984.
- Wadsworth HM, Stephens KS y Godpey AB. Statistical process control. En: Modern methods for quality control and improvement. Nueva York: John Wiley and Sons, 1986.
- Walpole RE, Myers RH. Control estadístico de calidad. En: Probabilidad y estadística. 4ª ed. México: McGraw Hill, 1992.

Anexo 4.1

Procedimiento para el análisis de una serie de mediciones con un gráfico de desarrollo

1. Buscar y seleccionar en el propio entorno de trabajo un indicador que se mida o resuma habitualmente con una frecuencia mensual o mayor (quincenal, semanal, diaria, etc.). Este indicador puede ser clínico (proceso o resultado en tipo de paciente, problema de salud) o relacionado con la actividad (pacientes visitados, medición de tiempo por paciente, proporción de consultas programadas, etc.), y expresado como cualquiera de los cuatro tipos de variables vistas en este capítulo (media, proporción, ratio, valores absolutos).
2. Hacer un cuadro con una serie temporal de mediciones. Para la construcción de la plantilla se necesitan, por lo menos, 20 mediciones (lo deseable es 30 o más).
3. Construir la plantilla para el análisis. Calcular el promedio utilizando la fórmula apropiada al tipo de gráfico (\bar{x} , \bar{p} y \bar{u}), según consta en el cuadro 4.8 de este capítulo. Poner atención en no confundir, en su caso, el tamaño de la muestra cada vez que se mide el indicador (n) con el número de mediciones de la serie (k). En caso de ser un gráfico de valores absolutos o individuales, utilizar la mediana como valor de referencia para la línea central.
4. Colocar los puntos correspondientes a la serie de mediciones y trazar el gráfico, uniéndolos con una línea.
5. Observar y señalar, en su caso, si hay algún patrón significativo de falta de estabilidad. Si el gráfico es de mediciones individuales o absolutas, utilizar la prueba de rachas.
6. Conclusiones del análisis:
 - ¿Hay algún patrón de inestabilidad o tendencia significativa?
 - Si es así, ¿cuáles serían las causas hipotéticas que lo explican?
 - ¿Puede utilizarse la plantilla que hemos construido para monitorizar mediciones sucesivas del mismo indicador? (razonar la respuesta, considerando si existe o no estabilidad y el valor promedio es satisfactorio).

Capítulo 5

Gráficos de control

Contenido general y objetivos específicos

1. Contextualización
2. Construcción de la plantilla gráfica para el análisis y monitorización de datos con gráficos de control estadístico
3. Interpretación de los gráficos de control
4. Principales problemas del control estadístico de la calidad

Contenido general

Se describen las características de los gráficos de control, incluyendo la forma de construirlos de acuerdo con los distintos tipos de indicadores (media, proporción, ratio, valores absolutos) y su interpretación. Se muestra cómo y por qué son más sensibles que los gráficos de desarrollo para detectar situaciones problemáticas, y se advierte sobre los principales problemas metodológicos que hay que tener en cuenta al utilizarlos.

Objetivos específicos

1. Describir las características básicas de los gráficos de control
2. Construir una plantilla de gráficos de control
3. Identificar los patrones propios de variabilidad especial en un gráfico de control
4. Determinar el tamaño de muestra necesario para realizar mediciones con gráficos de control
5. Tener en cuenta los principales problemas que dan lugar a interpretaciones erróneas de los resultados

Gráficos de control

1. Contextualización

En el capítulo anterior vimos las bases conceptuales del control estadístico de la calidad, haciendo referencia, desde el punto de vista práctico, a los dos formatos más sencillos (gráfico de desarrollo y gráfico de control) y los cuatro tipos de gráficos (\bar{x} , \bar{x} , p y u) que se relacionan con los cuatro tipos más comunes de indicadores que pueden utilizarse en los servicios de salud. Vimos también los ejemplos prácticos en relación con los gráficos de desarrollo: indicadores, construcción de la plantilla de referencia e interpretación de los gráficos. En este capítulo también analizaremos los datos de los cuatro tipos de indicadores, sólo que mediante gráficos de control.

2. Construcción de la plantilla gráfica para el análisis y monitorización de datos con gráficos de control estadístico

En relación con los gráficos de desarrollo vistos en el capítulo 4, los de control incorporan, por lo menos, dos líneas de referencia más (los límites de control) que representan el promedio ± 3 desviaciones estándar. De esta forma, la probabilidad de que una medición del indicador resulte fuera de los límites de control es <0.01 . La plantilla de base es la representada en la figura 5.1.

Las fórmulas para su cálculo son diferentes según el tipo de indicador desde el punto de vista estadístico. Las correspondientes a los cuatro tipos que hemos seleccionado figuran en el cuadro

La plantilla de los gráficos de control consta del promedio y, por lo menos, de dos límites de control que corresponden al promedio + 3 desviaciones estándar.

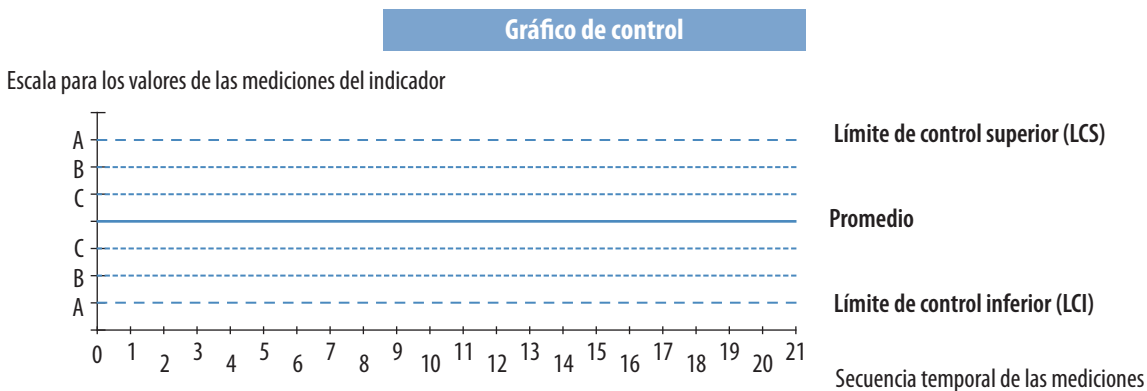


Figura 5.1. Gráfico de control

Los límites de control del gráfico x pueden establecerse calculando la desviación estándar del promedio (sx) o utilizando fórmulas que incluyen el rango medio.

Los límites de control en el gráfico p se calculan en función de n y de la desviación estándar de p. Estos límites son cambiantes si n no es constante.

5.1 y se aplican utilizando los mismos ejemplos que introdujimos en el capítulo anterior.

2.1. Gráfico de medias (X)

Hay dos maneras de proceder para establecer los límites de control: una, calculando la desviación estándar de la media (sx), lo que equivale a calcular la desviación estándar para el conjunto de muestras como si fuese una sola, y dividirlo entre \sqrt{n} ; y otra, más cómoda y común, utilizando el rango medio (R) multiplicado por una constante que varía con el tamaño de la muestra empleada, cada vez que hemos medido el indicador, que se busca en unos cuadros como los que se reproducen en el 5.6. Este cuadro contiene otros factores que se utilizan para calcular los límites de control cuando el promedio es sobre el rango (R) o sobre la desviación estándar (s).

En el capítulo 4 vimos un ejemplo de indicador medido como media: la de minutos por con-

sulta no programada. Utilizando estos mismos datos, en el cuadro 5.2, el valor de A2 es 0.577, que corresponde a las mediciones con muestras de cinco casos como los que hemos utilizado. Los límites de control quedan así establecidos en 11.7 (LCS) y 5.1 (LCI). El gráfico resultante, tras poner en él los puntos correspondientes a las 25 mediciones realizadas, aparece en la figura 5.2.

2.2. Gráfico de proporciones (p)

El cálculo de los límites de control se realiza según la fórmula que figura en el cuadro 5.1. Aplicándola al ejemplo de los datos de la proporción de pacientes que tienen que esperar más de 30 minutos (cuadro 5.3), nos da unos valores para la desviación estándar de la proporción media de nuestro indicador =0.06, y los límites de 0.05 y 0.41. El gráfico resultante está en la figura 5.3. Obsérvese en la fórmula que, al ser la desviación estándar de p dependiente del tamaño de la

Cuadro 5.1
Fórmulas para el cálculo de la línea central y límites de control de los gráficos más comunes

| Gráfico | Línea central (promedio) | Límites de control |
|---------------------------|---|---|
| X Valores individuales | $X = \frac{\text{núm. total de eventos en el periodo estudiado}}{\text{núm. mediciones realizadas}}$ | $X \pm 3 \frac{RM}{1.128}$ |
| X (medias) | = Media de las medias de cada medición del indicador en la serie que analizamos o utilizamos para construir el gráfico = $\frac{\sum X_i}{k}$ | $\pm 3 \text{ desv. estándar de la media}$ $\bar{X} \pm A_2 \bar{R}$ |
| P (proporciones) | P= Proporción media de la serie de mediciones analizada o utilizada para construir el gráfico = $\frac{\sum p_i \cdot n_i}{\sum n_i}$ | $\bar{p} \pm 3 \sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}}$ (*) |
| U (ratios) | $\bar{u} = \frac{\text{Total ocurrencias de las cualidades que se miden}}{\text{Total de casos evaluados}} = \frac{\sum c_i}{\sum n_i}$ | $\bar{u} \pm 3 \sqrt{\frac{\bar{u}}{n_i}}$ |

RM: Rango móvil retrospectivo medio (ver explicación en el texto).

R: Rango medio de las mediciones efectuadas: $\frac{\sum R_i}{k}$

A₂: Una constante dependiente del tamaño de muestra empleado (ver cuadro).

n_i: Tamaño de la muestra en cada medición.

k: n° de mediciones (muestras) realizadas para analizar el indicador o construir el gráfico.

c: Cumplimientos (o incumplimientos) de diversos criterios, evaluados todos ellos simultáneamente en cada unidad de estudio.

i: Secuencia ordinal de las mediciones efectuadas (1, 2, 3,i).

(*): Si el tamaño de la muestra no es constante, los límites son variables, según los diversos tamaños de muestra empleados. Si el tamaño de la muestra no es constante, los gráficos "np" (número absoluto de cumplimiento o incumplimiento de un criterio o indicador) y "c" (número absoluto de cumplimientos o incumplimientos de varios criterios o indicadores evaluados simultáneamente en cada unidad de estudio), son otra posible alternativa.

Cuadro 5.2

Indicador: media de minutos por consulta a demanda.
Resultado de 25 mediciones en días sucesivos

| Día | Tamaño de la muestra | Valor máximo | Valor mínimo | Rango | Media |
|-------|----------------------|--------------|--------------|-------|-------|
| 1 | 5 | 20' | 3' | 17 | 10.2 |
| 2 | 5 | 10' | 5' | 5 | 8.4 |
| 3 | 5 | 12' | 5' | 7 | 7.6 |
| 4 | 5 | 8' | 6' | 2 | 7.0 |
| 5 | 5 | 10' | 7' | 3 | 8.0 |
| 6 | 5 | 15' | 5' | 10 | 10.2 |
| 7 | 5 | 9' | 5' | 4 | 7.4 |
| 8 | 5 | 10' | 6' | 4 | 8.0 |
| 9 | 5 | 16' | 8' | 8 | 10.8 |
| 10 | 5 | 7' | 5' | 2 | 6.0 |
| 11 | 5 | 13' | 6' | 7 | 9.2 |
| 12 | 5 | 10' | 5' | 5 | 8.2 |
| 13 | 5 | 7' | 5' | 2 | 6.0 |
| 14 | 5 | 14' | 7' | 7 | 10.8 |
| 15 | 5 | 15' | 6' | 9 | 11.4 |
| 16 | 5 | 6' | 5' | 1 | 5.2 |
| 17 | 5 | 10' | 7' | 3 | 8.6 |
| 18 | 5 | 15' | 6' | 9 | 10.0 |
| 19 | 5 | 8' | 5' | 3 | 6.6 |
| 20 | 5 | 10' | 7' | 3 | 9.3 |
| 21 | 5 | 13' | 6' | 7 | 9.1 |
| 22 | 5 | 12' | 5' | 7 | 8.2 |
| 23 | 5 | 15' | 6' | 9 | 8.1 |
| 24 | 5 | 10' | 5' | 5 | 8.0 |
| 25 | 5 | 9' | 3' | 6 | 7.9 |
| Total | | | | 145 | 210.2 |

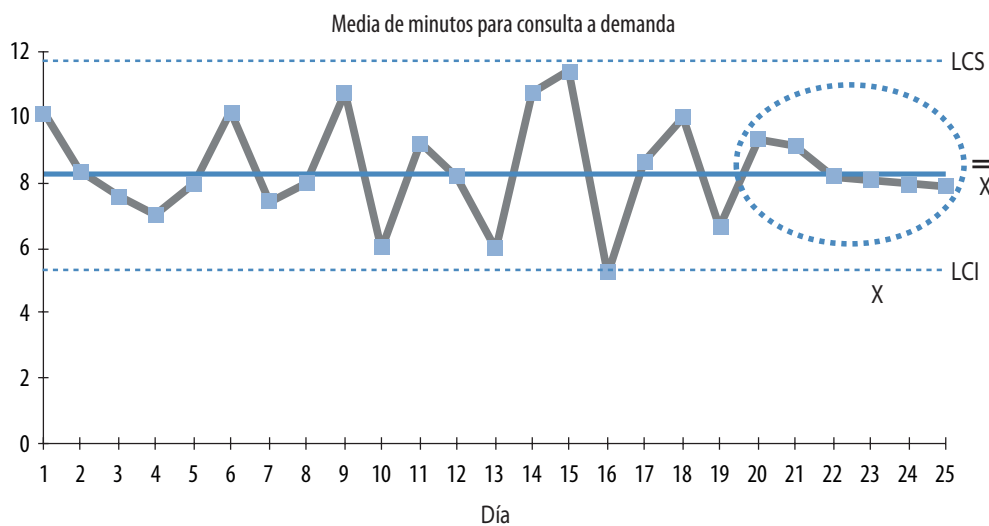
Cálculos para la línea promedio según las fórmulas del cuadro 5.1 para gráfico de medias:

$$\bar{X} = \frac{210.2}{25} = 8.4$$

$$\bar{R} = \frac{145}{25} = 5.8$$

$$LCS = 8.4 + (0.577 \times 5.8) = 11.7$$

$$LCI = 8.4 - (0.577 \times 5.8) = 5.1$$



*: datos del cuadro 5.2

Figura 5.2. Gráfico de control para el indicador "media de minutos para consulta a demanda"*

Cuadro 5.3

Indicador: proporción de pacientes que tienen que esperar más de 30' antes de ser atendidos.
Resultados de las mediciones efectuadas en todos los días de un mes

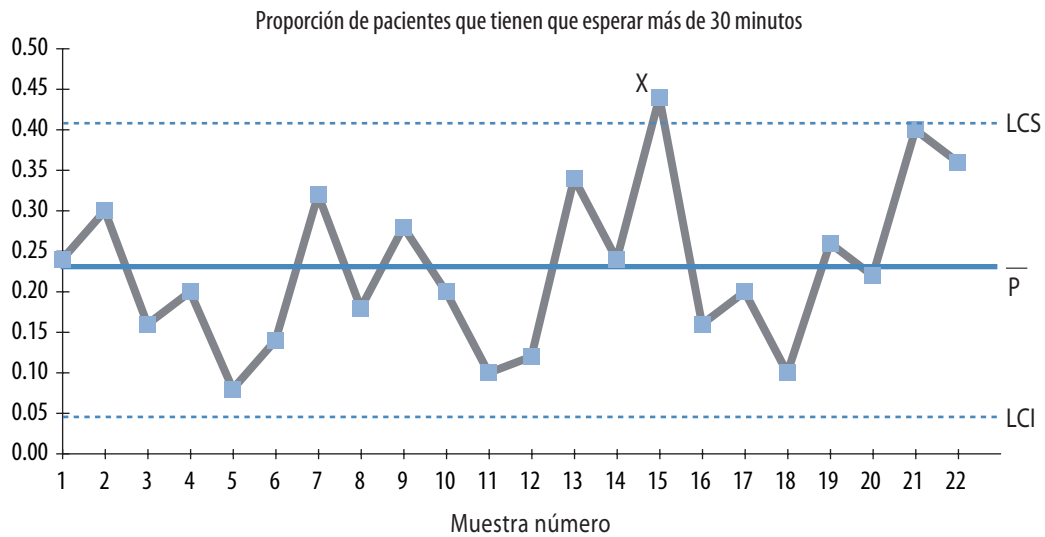
| Muestra número | Tamaño de muestra (n) | Número de incumplimientos | Proporción de incumplimientos |
|----------------|-----------------------|---------------------------|-------------------------------|
| 1 | 50 | 12 | 0.24 |
| 2 | 50 | 15 | 0.30 |
| 3 | 50 | 8 | 0.16 |
| 4 | 50 | 10 | 0.20 |
| 5 | 50 | 4 | 0.08 |
| 6 | 50 | 7 | 0.14 |
| 7 | 50 | 16 | 0.32 |
| 8 | 50 | 9 | 0.18 |
| 9 | 50 | 14 | 0.28 |
| 10 | 50 | 10 | 0.20 |
| 11 | 50 | 5 | 0.10 |
| 12 | 50 | 6 | 0.12 |
| 13 | 50 | 17 | 0.34 |
| 14 | 50 | 12 | 0.24 |
| 15 | 50 | 22 | 0.44 |
| 16 | 50 | 8 | 0.16 |
| 17 | 50 | 10 | 0.20 |
| 18 | 50 | 5 | 0.10 |
| 19 | 50 | 13 | 0.26 |
| 20 | 50 | 11 | 0.22 |
| 21 | 50 | 20 | 0.40 |
| 22 | 50 | 18 | 0.36 |
| TOTAL | 1.100 | 252 | |

Cálculos para la línea promedio y límites de control según las fórmulas del cuadro 5.9 para gráfico de proporciones:

$$\bar{p} = \frac{252}{1.100} = 0.23$$

$$LCS = 0.23 + (3 \times 0.06) = 0.41$$

$$LCI = 0.23 - (3 \times 0.06) = 0.05$$



*: datos del cuadro 5.3

Figura 5.3. Gráfico de control para el indicador "proporción de pacientes que tienen que esperar más de 30 minutos antes de ser atendidos"*

muestra (n), los límites de control variarán según la n de cada medición; por ello es conveniente realizar las mediciones con un tamaño de muestra constante.

2.3. Gráfico de valores individuales (\bar{x})

Para la utilización de la distribución normal como probabilidades de referencia, en lugar de la interpretación no paramétrica revisada en el capítulo 4 para los gráficos de rachas, se calcula como promedio la media de todas las mediciones (cada medición es en realidad como si fuese una muestra de un solo caso), y para los límites de control, el rango móvil medio (MR), que se calcu-

la promediando las diferencias (en valores absolutos) entre cada medición y la inmediatamente anterior. El cuadro 5.4 ofrece un ejemplo de cálculo para el indicador de actividad del programa de cirugía menor que vimos en el capítulo 4. Una alternativa a la fórmula $\frac{MR}{1.128}$ como estimación de la desviación estándar es el cálculo de la propia desviación estándar para el conjunto de mediciones, como si fuese una sola muestra. Sin embargo, la estimación que resulta utilizando MR se acerca más al modelo de distribución normal en los límites establecidos. Los límites de control que calculamos para nuestro ejemplo son 4 y 15 intervenciones, respectivamente. El gráfico de control con estos límites está igualmente en la figura 5.4.

Cuadro 5.4

Cirugía menor en un centro de salud. Número de intervenciones por semana

| Semanas | Número de intervenciones | Rango móvil |
|---------|--------------------------|-------------|
| 1 | 13 | — |
| 2 | 7 | 6 |
| 3 | 10 | 3 |
| 4 | 8 | 2 |
| 5 | 11 | 3 |
| 6 | 12 | 1 |
| 7 | 13 | 1 |
| 8 | 12 | 1 |
| 9 | 12 | 0 |
| 10 | 8 | 4 |
| 11 | 9 | 1 |
| 12 | 5 | 4 |
| 13 | 7 | 2 |
| 14 | 9 | 2 |
| 15 | 9 | 0 |
| 16 | 5 | 4 |
| 17 | 7 | 2 |
| 18 | 8 | 1 |
| 19 | 10 | 2 |
| 20 | 11 | 1 |
| Total | 186 | 40 |

Cálculos para la línea promedio según las fórmulas del cuadro 5.9 para gráfico de valores

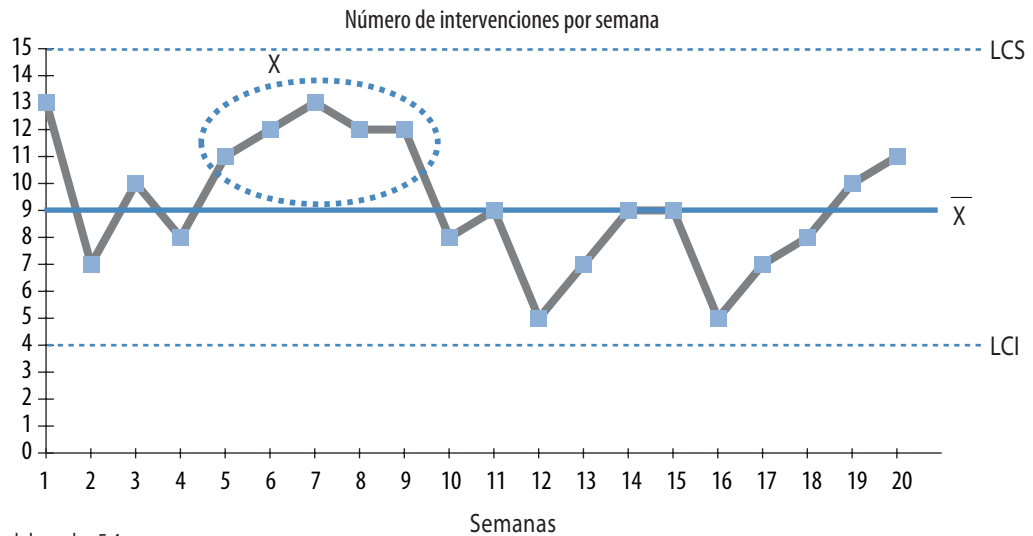
Mediana: 9

Media de intervenciones por semana: $\bar{x} = \frac{186}{20} = 9.3$

Rango móvil medio: $\frac{40}{19} = 2.1$

$$LCS = 9.3 + \left(3 \times \frac{2.1}{1.128}\right) = 15$$

$$LCS = 9.3 - \left(3 \times \frac{2.1}{1.128}\right) = 4$$



*: datos del cuadro 5.4

Figura 5.4. Gráfico de control para el indicador "número de intervenciones por semana"

El gráfico u utiliza la distribución de Poisson y, en su plantilla, los límites de control se establecen únicamente en función del promedio y el tamaño de la muestra.

2.4. Gráfico de ratios (u)

Los límites de control se calculan utilizando la distribución de Poisson, según la cual, la desviación estándar es la propia media dividida entre el tamaño de la muestra cada vez que se mide el indicador. Este es el proceder que podemos aplicar a nuestro ejemplo de número de defectos en los documentos de interconsultas (cuadro 5.5). Los límites, según la fórmula apropiada (cuadro 5.1) son de 0 a 1.1 defectos por documento (figura 5.5).

3. Interpretación de los gráficos de control

¿Qué añaden los límites de control al análisis de los datos en relación con los gráficos de rachas? En primer lugar, una prueba adicional muy evidente: si nos resulta un punto fuera de los límites de control hemos de señalar la situación como problemática, o por lo menos, distinta de lo esperado, significativamente *fuera de control*. Revisando los gráficos de las figuras 5.2 a 5.5 podemos ver que eso nos ocurre en dos de los indicadores: el de la proporción de pacientes que esperan más de 30 minutos (figura 5.3) y el del número de defectos por documento de interconsulta (figura 5.5). Este

último se nos muestra tremendamente inestable, muy lejos de estar bajo control estadístico, con varios picos en o fuera del límite de control. En el caso del tiempo de espera, habrá que averiguar qué ocurrió precisamente el día de la muestra número 15, ya que esa causa *especial* desequilibró significativamente lo que venía siendo un comportamiento dentro de unos límites de variabilidad normales; también debe investigarse a qué se debe la irregularidad de los partes interconsulta (¿depende del tipo de especialista?, ¿de días?, ¿de médicos?, ¿de la demanda?).

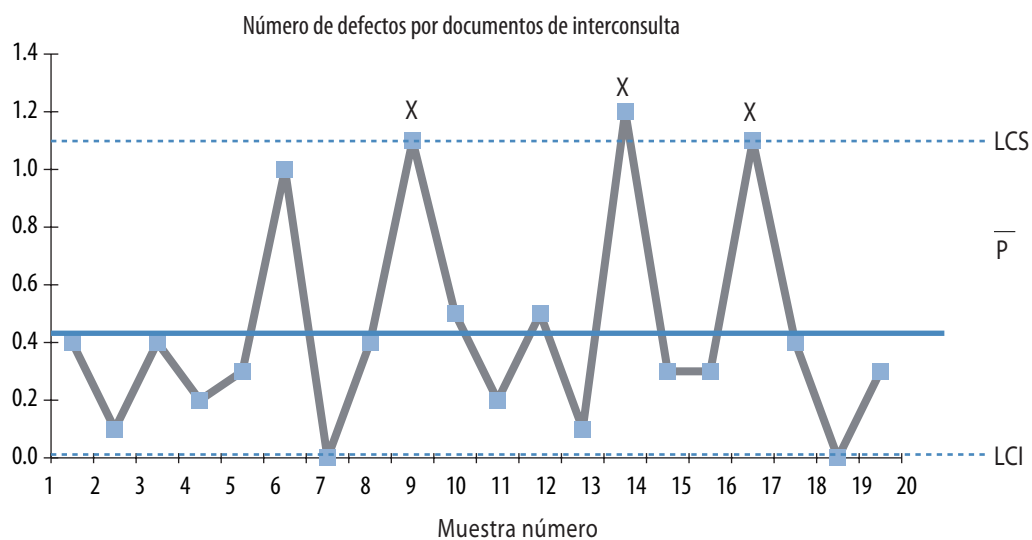
Sin embargo, los gráficos de control no solamente analizan los puntos fuera de los límites. En primer lugar, los mismos patrones de tendencias que vimos para los gráficos de desarrollo también son aplicables a los gráficos de control; además, si los *enriquecemos* señalando en la plantilla dónde se sitúan los límites correspondientes a una y dos desviaciones estándar, hay otra serie de patrones que nos indicarán *descontrol*, utilizando las probabilidades que corresponden a cada una de ellos (figura 4.3). La anotación habitual para estas desviaciones estándar es hablar de *zonas A* (entre dos y tres desviaciones estándar), *B* (entre una y dos) y *C* (entre el promedio y una). Por lo menos, se han establecido ocho patrones signi-

En los gráficos de control cualquier medición del indicador que caiga fuera de los límites de control es signo de problema, inestabilidad del proceso o causa *especial* que conviene investigar.

Cuadro 5.5

Indicador: número de defectos (criterios incumplidos) por documento de interconsulta. Resultados de 20 mediciones sucesivas en muestras de 10 documentos

| Medición | Tamaño de muestra (n) | Núm. total de defectos | Ratio |
|----------|-----------------------|------------------------|-------|
| 1 | 10 | 4 | 0.4 |
| 2 | 10 | 1 | 0.1 |
| 3 | 10 | 4 | 0.4 |
| 4 | 10 | 2 | 0.2 |
| 5 | 10 | 3 | 0.3 |
| 6 | 10 | 10 | 1.0 |
| 7 | 10 | 0 | 0.0 |
| 8 | 10 | 4 | 0.4 |
| 9 | 10 | 11 | 1.1 |
| 10 | 10 | 5 | 0.5 |
| 11 | 10 | 2 | 0.2 |
| 12 | 10 | 5 | 0.5 |
| 13 | 10 | 1 | 0.1 |
| 14 | 10 | 12 | 1.2 |
| 15 | 10 | 3 | 0.3 |
| 16 | 10 | 3 | 0.3 |
| 17 | 10 | 11 | 1.1 |
| 18 | 10 | 4 | 0.4 |
| 19 | 10 | 0 | 0.0 |
| 20 | 10 | 3 | 0.3 |
| Total | 200 | 88 | |



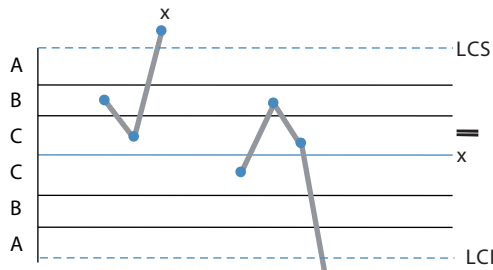
*: datos del cuadro 5.5

Figura 5.5. Gráfico de control para el indicador "número de defectos por documento interconsulta"*

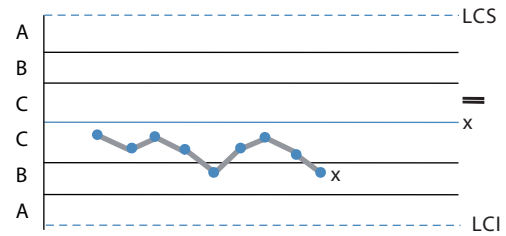
En los gráficos de control, son identificables una serie de secuencias significativas de *descontrol*, además de los patrones de inestabilidad que pueden apreciarse en los gráficos de desarrollo.

ficativamente anómalos (estadísticamente poco compatibles con una situación de estabilidad) (figura 5.6). Como cada una de estas pruebas está pensada para un error α (falsos positivos) en tor-

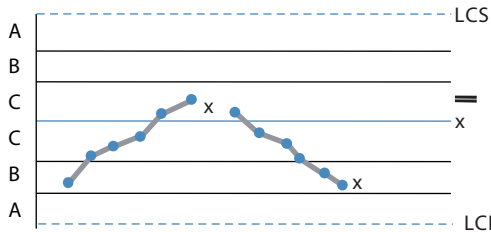
no a 0.3%, la utilización conjunta de todos ellos eleva el error α , como mucho, a $0.3 \times 8 = 2.4\%$; sin embargo, se incrementa dramáticamente el poder de detección de situaciones problemáticas.



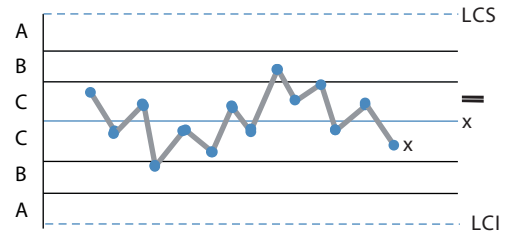
Test 1. Un punto fuera de los límites de control



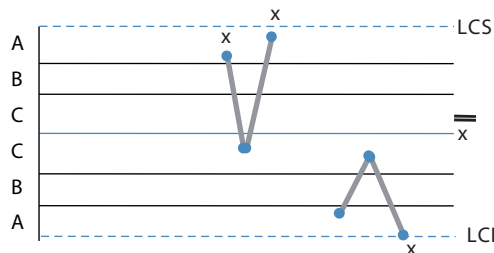
Test 2. Nueve puntos seguidos en la zona C o más allá (de un mismo lado)



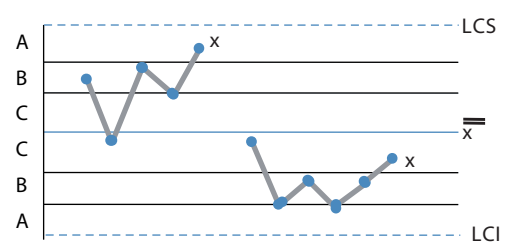
Test 3. Seis puntos seguidos ascendentes o descendentes



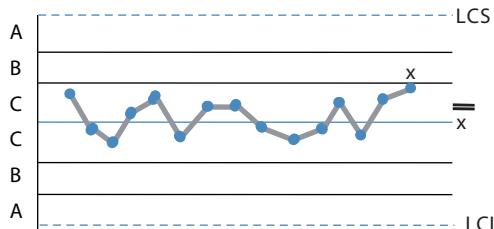
Test 4. Catorce puntos seguidos con alternancia



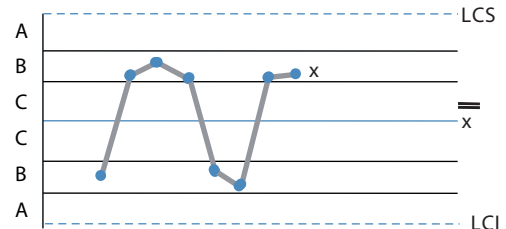
Test 5. Dos de tres puntos seguidos en la zona A (de un mismo lado)



Test 6. Cuatro de cinco puntos seguidos en la zona B o más allá (de un mismo lado)



Test 7. Quince puntos seguidos en la zona C (arriba o abajo)



Test 8. Ocho puntos seguidos a ambos lados de la línea central y ninguno en la zona C

Figura 5.6. Pruebas para identificar situaciones fuera de control

Como ilustración de esta mayor sensibilidad para detectar situaciones *fuera de control*, podemos observar (figura 5.4) cómo el indicador para monitorizar el programa de cirugía menor (el único de los cuatro que hemos venido analizando que parecía estar *bajo control*) muestra un patrón significativamente anómalo según la sexta prueba de la figura 5.6 de la quinta a la novena semanas. Esta situación *fuera de control* es positiva; pero igualmente debería investigarse para *fixar* en el proceso aquel factor o factores que hicieron que se *descontrolara* por arriba; de hecho, que mejorara su promedio.

Obsérvese que con los gráficos de desarrollo (capítulo 4) sólo fue posible detectar como problemático uno de los cuatro indicadores, mientras que los gráficos de control nos han permitido saber que, en realidad, ninguno de los cuatro está realmente controlado, y que no conocemos ni dominamos la variabilidad de los procesos que se miden. Esta situación no es la idónea en una monitorización rutinaria; antes hay que conocer mejor el proceso, intervenir sobre él y llevarlo a control estadístico. Para ello, eliminaremos (o incluiremos permanentemente, según tengan un efecto negativo o positivo) las causas especiales y recalcularemos los límites de control excluyendo los puntos *fuera de control*, si responden a situaciones conocidas, especiales y normales, o, tras la intervención, los cálculos para la plantilla serán sustituidos por lo que resulte de nuevas mediciones y nuevos análisis.

4. Principales problemas del control estadístico de la calidad

Hay tres aspectos que pueden ocasionar problemas al utilizar los gráficos de control. El primero de ellos, ya mencionado en el capítulo 4, es la confusión entre control estadístico y conformidad con los estándares, requisitos o especificaciones de calidad; el segundo es el tamaño de muestra apropiado para las mediciones del indicador, y el tercero, la definición del marco muestral y los subgrupos (muestras) en los que se realizan las mediciones.

4.1. Confusión entre control estadístico y conformidad con los requisitos de calidad

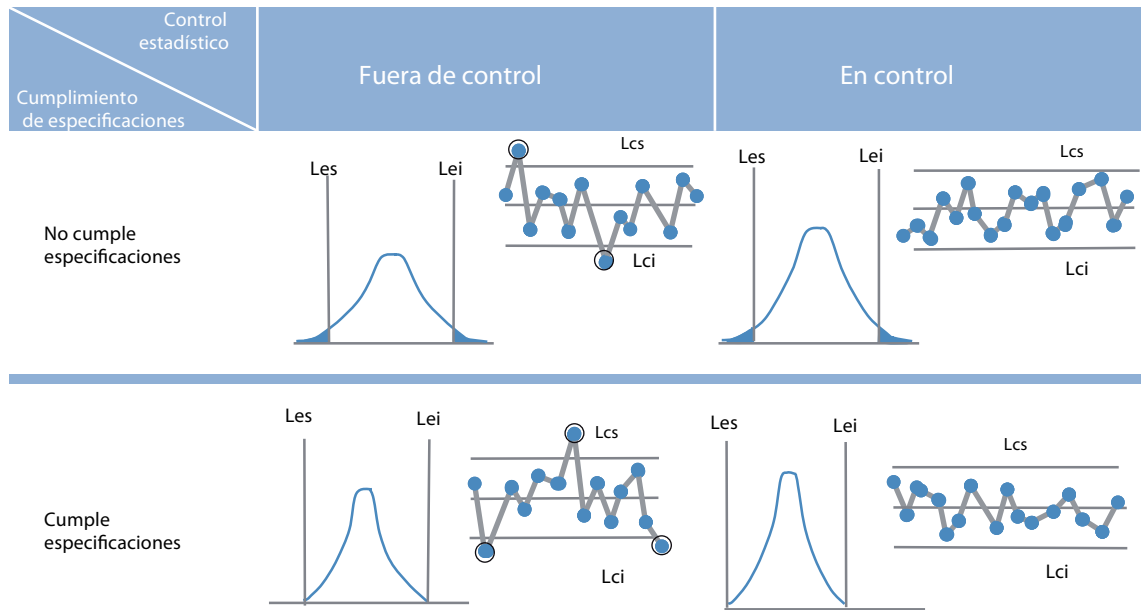
Esta confusión que, de acuerdo con Deming, es uno de los “errores que salen caros”, puede re-

vestir dos formas distintas pero relacionadas. La primera es pensar que los límites de control son especificaciones de calidad u objetivos por cumplir, cuando, en realidad, se determinan estadísticamente, de forma que el gráfico de control nos dice lo que podemos esperar del proceso tal como es, no como quisiéramos que fuese. No es correcto poner directamente las especificaciones de calidad (por ejemplo, el estándar y el umbral del indicador) como promedio y límite de control. La segunda, es hacer equivalente el control estadístico a nivel aceptable de calidad. Existen manuales, algunos de ellos traducidos al español, como el de Hanson y Ghare, que elevan este error a categoría al afirmar, por ejemplo, que “proceso bajo control equivale a proceso de producción capaz de ofrecer una proporción satisfactoria de elementos conformes con las especificaciones”.

Así como los gráficos de control sirven para ver si hay causas que investigar pero no cuáles, la consecución del control estadístico es más que nada un requisito para analizar el grado de conformidad con las especificaciones de calidad. De hecho, pueden surgir cuatro tipos de situaciones (figura 5.7): proceso estable, en control estadístico, que no cumple las especificaciones o estándares de calidad; proceso estable que sí cumple las especificaciones; proceso inestable dentro de las especificaciones; y proceso inestable y no conforme con las especificaciones. La averiguación de la situación concreta en que nos encontramos en relación con las especificaciones de calidad puede realizarse con los gráficos de capacidad de proceso (capítulo 6). Alternativamente, en caso de ser un indicador cuantitativo medido como media, el análisis y control de la variabilidad interna, aun cuando el indicador se mantenga en un promedio aceptable y dentro de los límites de variabilidad esperados (límites de control), puede realizarse con los gráficos de rangos o los de desviaciones estándar (ver fórmulas del cuadro 5.6) en los que se analiza con gráfico de control el rango (gráfico R) o la desviación estándar de cada una de las mediciones (gráfico S), el promedio o línea central del gráfico, el rango medio de todas las mediciones (R) es el gráfico R, y la media de las desviaciones estándar (S) es el gráfico S, recordar qué promedios estables pueden enmarcar variabilidad intragrupo. Como resume Juran (citado por Deming) “los problemas importantes para mejorar comienzan una vez que se ha logrado el control

La división del espacio entre el promedio y los límites de control en zonas que corresponden cada una a una desviación estándar, permite a los gráficos de control tener una mayor sensibilidad para detectar situaciones problemáticas.

Si bien la inestabilidad o falta de predictibilidad estadística es siempre indeseable, la estabilidad o control estadístico no es siempre sinónimo de calidad. Un indicador puede estar estabilizado y ser estadísticamente predecible a niveles de calidad alejados del estándar deseable.



Fuente: Adaptado de Kume H.

Figura 5.7. Control estadístico y cumplimiento de especificaciones: cuatro situaciones posibles

La estabilidad o control estadístico es un requisito previo para plantearse el análisis de la conformidad con los requisitos o estándares de calidad de los indicadores.

estadístico". Esta secuencia de control estadístico → conformidad con especificaciones → monitorización, como tres actividades diferenciadas en las que pueden intervenir los gráficos de control es la reflejada en el algoritmo de la figura 4.2.

En los ejemplos de indicador que hemos utilizado en el capítulo anterior y en éste, una vez que hayamos conseguido que los indicadores analizados sean estables y predecibles, deberíamos preguntarnos si el promedio y la variabilidad obtenidos responden a nuestros deseos en cuanto a nivel de calidad; si es así, podemos utilizar el control estadístico para monitorizar y asegurarnos de que nos mantenemos en ese nivel; si no es así, hay que averiguar qué cambios hay que introducir para lograr el nivel de calidad deseado.

4.2. Tamaño de la muestra para la medición de los indicadores

No parece haber ninguna fórmula, número mágico ni un método mejor que otros para su cálculo; pero sí una serie de recomendaciones de base un tanto empírica y probabilística.

En primer lugar hay que distinguir entre variables cuantitativas y cualitativas. Para las primeras, cuya medición expresamos normalmente como media de la muestra analizada, el número de casos habitual en la industria raramente excede los 10 casos; por lo común, las muestras son de cinco casos. Este número tan pequeño se compensa porque las mediciones son frecuentes, y porque lo que se utiliza para el gráfico es la media y no los valores individuales (con la correspondiente disminución de la variación con respecto a la que podrían tener los valores individuales).

Para las variables cualitativas, dicotómicas, como son la mayoría de indicadores que utilizamos en los servicios de salud, hay varias formas para calcular el tamaño de la muestra. La primera y más sencilla es pensar que lo deseable es que sea lo suficientemente grande como para dar oportunidad de que aparezcan los incumplimientos o defectos en las mediciones. Fijando el número de incumplimientos por detectar en, por lo menos, dos en cada medición del indicador, la muestra necesaria para cada una sería $n = \frac{2}{p}$, donde p es la proporción esperada (promedio estable) de incumplimientos o defectos.

Para variables cuantitativas (gráficos \bar{x}), las mediciones del indicador pueden realizarse con muestras de cinco casos.

Cuadro 5.6
Factores para los cálculos de los límites de control en gráficos de medias, desviaciones estándar y rangos

| Tamaño de muestra | Gráfico para medias | Gráfico para desviaciones estándar | | Gráfico para rangos | |
|-------------------|----------------------------------|------------------------------------|-------|----------------------------------|-------|
| | Factores para límites de control | Factores para límites de control | | Factores para límites de control | |
| | A_2 | B_3 | B_2 | D_3 | D_4 |
| 2 | 1.880 | 0 | 3.267 | 0 | 3.267 |
| 3 | 1.023 | 0 | 2.568 | 0 | 2.574 |
| 4 | 0.729 | 0 | 2.266 | 0 | 2.282 |
| 5 | 0.577 | 0 | 2.089 | 0 | 2.114 |
| 6 | 0.483 | 0.030 | 1.970 | 0 | 2.004 |
| 7 | 0.419 | 0.118 | 1.882 | 0.076 | 1.924 |
| 8 | 0.373 | 0.185 | 1.815 | 0.136 | 1.864 |
| 9 | 0.337 | 0.239 | 1.761 | 0.184 | 1.816 |
| 10 | 0.308 | 0.284 | 1.716 | 0.223 | 1.777 |
| 11 | 0.285 | 0.321 | 1.679 | 0.256 | 1.744 |
| 12 | 0.266 | 0.354 | 1.646 | 0.283 | 1.717 |
| 13 | 0.249 | 0.382 | 1.618 | 0.307 | 1.693 |
| 14 | 0.235 | 0.406 | 1.594 | 0.328 | 1.672 |
| 15 | 0.223 | 0.428 | 1.572 | 0.347 | 1.653 |
| 16 | 0.212 | 0.448 | 1.552 | 0.363 | 1.637 |
| 17 | 0.203 | 0.466 | 1.534 | 0.378 | 1.622 |
| 18 | 0.194 | 0.482 | 1.518 | 0.391 | 1.608 |
| 19 | 0.187 | 0.497 | 1.503 | 0.403 | 1.597 |
| 20 | 0.180 | 0.510 | 1.490 | 0.415 | 1.585 |
| 21 | 0.173 | 0.523 | 1.477 | 0.425 | 1.575 |
| 22 | 0.167 | 0.534 | 1.466 | 0.434 | 1.566 |
| 23 | 0.162 | 0.545 | 1.455 | 0.443 | 1.557 |
| 24 | 0.157 | 0.555 | 1.445 | 0.451 | 1.548 |
| 25 | 0.153 | 0.565 | 1.435 | 0.459 | 4.541 |

Fórmulas:

| | | | |
|-----|-------------------------|---------|---------|
| LCS | $\bar{x} + A_2 \bar{R}$ | $B_4 S$ | $D_4 R$ |
| LCI | $\bar{x} - A_2 \bar{R}$ | $B_3 S$ | $D_3 R$ |

Fuente: Adaptado de Walpole RE y Myers RH.

Obviamente para proporciones esperadas mayores de 0.1 (10%), las muestras son pequeñas (para $p = 0.1 \rightarrow n = \frac{2}{0.1} = 20$); sin embargo, cuanto más pequeña sea p , mayor deberá ser el tamaño de muestra. Algunos autores ven esto como una limitación importante, al pensar, desde luego, en las bajísimas proporciones de defectos

con las que se debe trabajar en la industria (para $p=1\%$, se necesitan muestras de 200 casos); pero no es tan problemático en los servicios de salud, en los que habitualmente nos manejamos con estándares menos extremos. Por ejemplo, para una proporción media esperada de incumplimientos (estándar del indicador) de 0.05 (5%), lo

En variables cualitativas (gráficos p) el tamaño de la muestra en cada medición del indicador puede ser de 20 a 40 casos, siempre que la proporción esperada de defectos o incumplimientos no sea muy pequeña (<0.05).

El tamaño de muestra necesario para los gráficos p también puede calcularse en función de la diferencia que establezcamos entre el estándar y el umbral del indicador. Cuanto más pequeña sea, más casos necesitaremos.

Es conveniente que el tamaño de la muestra en las mediciones del indicador sujeto a control estadístico sea constante. No obstante, los gráficos de control estadístico pueden también utilizarse, ajustándolos, en caso de que el tamaño de la muestra no sea siempre el mismo, e incluso si las mediciones son de toda la población.

El marco muestral para los indicadores que se van a analizar con control estadístico debe ser homogéneo para evitar que el promedio enmascare una probable variabilidad.

La incorporación de los gráficos de control estadístico a la rutina de los programas de gestión de la calidad es factible, fácil y práctica.

cual es un valor casi extremo en nuestro medio, se necesitan muestras de 40 casos.

Otra forma de calcular la muestra es sobre la base de cuál es la diferencia, en relación con la proporción media, que nos interesa identificar, con el nivel de significación que establezcamos para las pruebas al analizar el gráfico. La fórmula, para una significación <0.01 , una proporción de base (promedio o estándar) P_0 y una diferencia por detectar “d”, (por ejemplo la que estableceríamos entre el estándar y su umbral), es: $n = \frac{9 \cdot P_0(1-P_0)}{d^2}$, la cual nos da, en general, tamaños de muestra más grandes que con el otro método y, obviamente, más cuanto menor sea “d”.

Hay otros dos aspectos importantes que debemos considerar. Uno es la conveniencia de que el tamaño de la muestra sea siempre el mismo, sobre todo para gráficos de proporciones y ratios. En caso contrario, con cada medición, habría que recalcular los límites de control y ajustarlos al tamaño de muestra según las fórmulas del cuadro 5.1. Como alternativa, se puede utilizar un tamaño de muestra promedio, siempre que no estemos empleándolo para mediciones con muestras superiores o inferiores a 20% de este promedio. Finalmente, aunque los gráficos de control, sus fórmulas y su práctica están pensados para mediciones realizadas sobre muestras, no hay ningún inconveniente en utilizarlos para analizar y monitorizar indicadores medidos sobre poblaciones completas, siempre que se mida una evolución temporal.

4.3. Definición de las muestras (subgrupos)

Para muchos autores, la definición de los marcos muestrales y las correspondientes muestras (*subgrupos* en la industria) es la parte más impor-

tante de la preparación de los gráficos de control, al grado que una definición o composición incorrecta puede generar gráficos inútiles.

Las mediciones deben realizarse utilizando subgrupos *racionales*, homogéneos, en el sentido de que exista un mínimo de variabilidad dentro del subgrupo, de manera que la monitorización o análisis pueda detectar la evolución de la variabilidad entre subgrupos (muestras). Si, por ejemplo, en un centro o unidad asistencial, el comportamiento de los profesionistas es muy distinto entre sí, juntar todos los pacientes de forma indiscriminada para las muestras puede dar una idea falsa de promedio, y enmascarar la existencia de una amplia variabilidad dentro de cada muestra. La utilización de gráficos estratificados (uno para cada grupo homogéneo) desenmascara en ocasiones esta situación.

Lo que subyace en definitiva es la necesidad de ser especialmente escrupuloso al extraer muestras homogéneas y representativas (no sesgadas). Cualquier agrupación en los marcos muestrales que implique o enmascare posibles diferencias intragrupo en métodos, personas, material o tecnología (las posibles fuentes de variabilidad) debe cuestionarse y aclararse antes de utilizar las mediciones para análisis o monitorización con gráficos de control estadístico.

Una vez que se toman en cuenta todas las precauciones, la incorporación de los gráficos de control estadístico a la rutina de los programas de gestión de la calidad y como forma de analizar y presentar las mediciones sistemáticas que se realizan en el sistema de salud es factible, fácil y práctica. Adicionalmente, existen varios programas de software que posibilitan el análisis rutinario de cualquier indicador, mediante la utilización de las técnicas de los gráficos de control.

BIBLIOGRAFÍA

- Deming WE. Calidad, productividad y competitividad. La salida de la crisis. Madrid: Díaz de Santos, 1989.
- Farnum NR. Control chart concepts. En: Modern Statistical quality control and improvement. Belmont: Duxbury Press, 1994.
- Gitlow H, Gitlow S, Oppenheim A y Oppenheim R. Tools and methods for the improvement of quality. Boston: Irwin, 1989.
- Hanson BL, Ghare PM. Control estadístico de procesos. En: Control de calidad. Teoría y aplicaciones. Madrid: Díaz de Santos, 1990.
- Hanson BL, Ghare PM. Control estadístico de procesos. En: Control de calidad. Teoría y aplicaciones. Madrid: Díaz de Santos, 1990.
- Kume H. Control charts. En: Statistical methods for quality improvement. Tokyo: AOTS, 1992.
- Peña O, Prat A. Técnicas estadísticas de control de calidad. En: ¿Cómo controlar la calidad? 2ª ed. Madrid: IMPI, 1990.
- Plsek P. Responding to variation in health care organizations. París: European Forum on QI in health care, 1997.
- Wadsworth HM, Stephens KS y Godpey AB. Modern methods for quality control and improvement. Nueva York: John Wiley & Sons, 1986.

Capítulo 6

Gráficos de capacidad de proceso. Sistema Seis Sigma para la mejora continua

Contenido general y objetivos específicos

1. Contextualización
2. ¿Qué es la capacidad del proceso?
3. Terminología de los estudios y optimización de la capacidad de los procesos
4. Objetivos de los estudios de capacidad del proceso. Hacia la perfección a través de la mejora continua
5. Fases de un estudio de capacidad del proceso
6. Estimación de la extensión del proceso
7. Estimación de la capacidad del proceso
8. Estimación de la tasa de producción defectuosa
9. Tasa de producción defectuosa en procesos medidos con indicadores cualitativos y mediciones que no siguen la distribución normal
10. Requisitos y precauciones para los cálculos de la capacidad del proceso
11. ¿Cómo reaccionar ante los resultados? Utilidad del análisis de la capacidad del proceso
12. La capacidad del proceso como centro y guía de la gestión de la calidad: el sistema Seis Sigma

Anexo 6.1. Probabilidades en un extremo o cola de la curva de la distribución normal

Contenido general

Se describen los conceptos, características y terminología, y se analiza la utilidad potencial de los estudios de capacidad del proceso, una metodología de gran importancia para la mejora continua de la calidad que aún no se emplea como correspondería en los servicios de salud. La capacidad del proceso analiza la medida en que un determinado proceso es capaz de producir el nivel de calidad deseado en cada una de sus unidades de producto o servicio. Su utilización como centro y guía de los programas de mejora continua de la calidad conforma los procesos y estrategia conocida como Seis Sigma.

Objetivos específicos

1. Describir el concepto y objetivos de los estudios de capacidad de los procesos.
2. Entender los conceptos de capacidad, límites naturales, extensión, especificaciones y tolerancia de un proceso.
3. Describir las fases de un estudio de capacidad del proceso.
4. Estimar los límites naturales de un proceso.
5. Estimar la extensión de un proceso.
6. Estimar el índice de capacidad de un proceso.
7. Estimar los índices de capacidad unilaterales y bilaterales.
8. Interpretar los diversos índices de capacidad del proceso.
9. Estimar la tasa de producción defectuosa de un proceso.
10. Describir los requisitos y precauciones que deben considerarse en la realización de estudios de capacidad del proceso.
11. Describir la utilidad de los estudios de capacidad del proceso en los programas de gestión de la calidad.
12. Describir las características de un programa Seis Sigma para la mejora continua y la gestión de la calidad.
13. Calcular e interpretar los niveles de calidad expresados en Sigma.

Gráficos de capacidad de proceso. Sistema Seis Sigma para la mejora continua

1. Contextualización

En los dos capítulos anteriores revisamos las bases teóricas y la aplicación práctica para diversos tipos de mediciones de los gráficos de control estadístico de calidad más comunes. Hemos visto su utilidad para el análisis de la variabilidad de los indicadores y la monitorización continua, una vez establecida la estabilidad y los niveles de calidad deseados. Sin embargo, si bien los gráficos de control estadístico informan adecuadamente sobre la estabilidad, el razonamiento sobre si se cumplen o no los niveles de calidad deseados es un razonamiento añadido que puede precisar de cálculos adicionales. El análisis de la capacidad del proceso nos da explícita y numéricamente esta información para poder concluir si el indicador está listo o no para monitorizar, o si el diseño ha producido los resultados deseados. En cualquier caso, nos va a decir en qué condiciones se encuentra el proceso analizado (representado por su correspondiente indicador o indicadores), en relación con el nivel de calidad esperado.

Si bien parece que el análisis gráfico y numérico de la capacidad de los procesos es utilizado en la industria a partir de la definición práctica de su medición en el ámbito industrial japonés, no es en absoluto común en la producción de servicios, y menos aún en los servicios de salud. A la vista de su lógica y utilidad potencial, tal como veremos en este capítulo, no parece razonable que sea así. Una vez entendido su funcionamiento, cualquiera que se atreva a utilizarlo en los programas de gestión de calidad de servicios de salud, ofrecerá una experiencia pionera que, sin duda, va a servir de referencia para otros. Una forma particular de utilización es dentro de los sistemas

de mejora continua que se conocen como Seis Sigma, los cuales están ganando popularidad a pasos agigantados desde que se acuñaron con este nombre a finales de los años ochenta.

2. ¿Qué es la capacidad del proceso?

En referencia a cualquier proceso de producción de bienes o servicios, en su totalidad o en las partes donde se considere, la capacidad del proceso es una característica del mismo, su comportamiento natural una vez eliminadas todas las causas de variabilidad especiales. El término capacidad hace referencia a la medida en que este proceso produce los niveles de calidad requeridos, en todas y cada una de las unidades del producto, una vez controlado; es decir, en su comportamiento estable y predecible.

Conviene subrayar que, consecuentemente, el análisis de la capacidad de los procesos para producir calidad va indefectiblemente unido al control estadístico de los mismos, y a la definición explícita y precisa del nivel de calidad que queremos obtener con ese proceso. En los dos capítulos anteriores vimos lo que significa que un proceso (medido a través de su indicador o indicadores) esté bajo control estadístico; también tenemos la idea de que el concepto de nivel de calidad deseado puede ser asimilable al de estándar que queremos para ese indicador. Esos conceptos hacen referencia a valores medios, normalmente medidos en grupos de casos (muestras); sin embargo, al definir y analizar la capacidad de los procesos hay un salto analítico y cualitativo importante: aunque conceptualmente similares a la forma en que los hemos manejado hasta ahora, al hablar de capacidad de proceso, vamos a referir los concep-

El llamado análisis de la capacidad del proceso nos informa sobre si éste produce o no los niveles de calidad deseados.

El análisis de la capacidad del proceso, muy extendido en la industria, es una actividad pionera en los servicios de salud.

Se conoce como capacidad de un proceso a la medida en que éste produce, en su estado natural o estable, los niveles de calidad requeridos en cada unidad de producto.

El valor nominal de una determinada característica de un producto o servicio es aquél que define inequívocamente su calidad, a nivel de unidad de producto o caso individual.

En el caso de criterios o requisitos cualitativos (presencia/ausencia), el valor nominal va a ser siempre la presencia de este criterio o requisito de calidad.

Las especificaciones o tolerancia de los requisitos de calidad de un determinado aspecto de un producto o servicio definen la gama de valores que lo hacen útil, de calidad. Generalmente se expresan como límites en torno al valor nominal.

tos de control estadístico y estándar de calidad a cada unidad de producto o servicio y no a grupos o muestras de casos. De hecho, es perfectamente posible que un proceso con un indicador que cumpla su estándar de calidad no sea capaz de producir el nivel de calidad requerido en todas o, incluso, ni siquiera en la mayoría de las unidades de producto.

La capacidad de los procesos para producir unidades, todas ellas de acuerdo con las características de calidad predefinidas, ha sido tradicionalmente una preocupación de primer orden para la industria, de manera que se ha desarrollado un cuerpo importante de métodos y técnicas en relación con este tema. Paralela y lógicamente se ha ido acuñando una serie de términos que conviene conocer y, en la medida de lo posible, relacionar con otros términos de gestión de calidad que se emplean en los servicios de salud. Antes o después, igual que lenta pero inexorablemente, han ido incorporándose otros métodos y técnicas como el LQAS o los gráficos de control, los métodos y técnicas para estudiar y optimizar la capacidad de los procesos terminarán incorporándose y utilizándose con éxito en los programas de gestión de calidad de los servicios de salud.

3. Terminología de los estudios y optimización de la capacidad de los procesos

Aparte del concepto en sí de capacidad del proceso, definido más arriba, hay otra serie de conceptos y términos que vamos a utilizar a lo largo de este capítulo y que van unidos al análisis y optimización de la capacidad de los procesos: valor nominal, especificaciones, tolerancia, límites de tolerancia, límites naturales y extensión del proceso.

Valor nominal. El valor nominal para un determinado producto, servicio o característica concreta de los mismos es aquél que define inequívocamente su calidad. Conceptualmente equivalente al estándar, pero referido a una unidad de producto, a cada caso individual, no a valores medios de grupos de casos o productos.

Cuando se consideran productos manufacturados es un concepto fácil de entender. Por ejemplo, al fabricar llaves para un determinado tipo de cerradura, líquidos con determinada concentración de producto, material de una cierta resis-

tencia para el uso al que va destinado o tuercas para un determinado tipo de tornillo, existe un determinado grosor, concentración, resistencia o forma que determina un uso perfecto, un acople sin problemas. Dicho grosor, concentración, resistencia o forma es el valor nominal de ese producto; aunque otras piezas ligeramente desviadas del valor nominal (hasta ciertos límites) pueden servir también para cerrar y abrir la cerradura, apretar el tornillo o usar el producto. Lo ideal sería que todos tuviesen el valor nominal para que no haya desperdicios ni clientes insatisfechos.

En el sector de servicios, el valor nominal es algo más complicado de entender y concretar, sobre todo cuando hablamos de características cualitativas, como la mayoría de los criterios o indicadores de calidad. Normalmente, para estas características no hay una escala de valores ni desviación posible a nivel individual para que la unidad sea considerada de calidad, de manera que el valor nominal sería siempre que esa característica (criterio de calidad) esté presente. Sin embargo, cuando el indicador que se mide es de tipo cuantitativo o en una escala de valores asimilable a variable cuantitativa puede establecerse con cierta facilidad un valor nominal, siempre que lo refiramos a un mismo y homogéneo tipo de casos. Pensemos, por ejemplo, en valores nominales para minutos por consulta, tiempo de espera o retraso en administración de trombolíticos en infarto de miocardio, por mencionar los tres ejemplos utilizados en los capítulos anteriores. El valor nominal consiste en el objetivo de calidad a garantizar para cada uno de los pacientes, usuarios o clientes. Si queremos utilizar términos ya conocidos, aunque sea de forma heterodoxa, podríamos decir que el valor nominal sería el estándar de calidad a nivel individual.

Especificaciones. Las especificaciones se refieren a las características que debe tener un determinado aspecto de un producto o servicio para que pueda ser útil, es decir, de calidad. En el caso de aspectos cuantitativos, las especificaciones se dan en forma de límites en torno al valor nominal, los cuales también se conocen como tolerancia de los requisitos o criterios de calidad. En los ejemplos arriba citados, los límites de las especificaciones o tolerancia del grosor de la llave serían aquellos valores de grosor que la hacen útil, y por fuera de los cuales no podría servir para la cerradura

para la que están fabricados; de igual manera, los tiempos por encima (ejemplo de tiempos de espera) o por debajo (ejemplo de tiempo de visita) de los cuales habría insatisfacción o mala calidad o un tratamiento inefectivo (ejemplo de retraso en la administración de trombolíticos en el infarto de miocardio) serían también las especificaciones de estos criterios de calidad. Las especificaciones o tolerancia sólo pueden establecerse adecuadamente en variables cuantitativas. En el caso de criterios o indicadores cualitativos dicotómicos, no se establecen tolerancias (recuérdese que estamos siempre hablando de valores individuales).

Algunos autores distinguen entre especificaciones de diseño (lo que se desea que tenga un determinado producto o servicio) y especificaciones de inspección (los patrones que sirven para juzgar si se está obteniendo lo que se desea.). Esta distinción sólo tiene interés para indicar que ambos deben ser concordantes porque, en la industria, un problema frecuente es que se diseñan especificaciones imposibles de cumplir o de medir, que no concuerdan con las especificaciones de inspección, las cuales tienden a adaptarse más a la realidad del día a día.

Límites de tolerancia. Como hemos visto, éstos se corresponden con las especificaciones y, según la ASQC (ANSI/ASQC, Estándar A1), definen valores que delimitan la conformidad de una unidad individual con la calidad requerida. Éstos pueden establecerse en ambos lados del valor nominal o en uno solo, para definir los límites de tolerancia o especificación superior (LES) e inferior (LEI).

Límites naturales o límites de tolerancia naturales son los que corresponden al funcionamiento normal (controlado y predecible) del proceso, referidos también a cada unidad individual, y estadísticamente determinados en una serie de mediciones por la media $+3s$. Subrayaremos de nuevo que la desviación estándar a que se hace referencia no es la desviación estándar de la media (de grupos de casos o muestras) que veíamos en los gráficos de control, sino la desviación estándar de los valores de cada unidad medida, individualmente considerada. Los límites naturales son una característica propia del proceso, independientemente de las especificaciones o límites de tolerancia establecidos.

Extensión del proceso. Gama de valores entre sus límites naturales; es decir, los valores posibles del indicador que pueden resultar del proceso que se está valorando, en su estado *natural*, estable y controlado.

4. Objetivos de los estudios de capacidad del proceso. Hacia la perfección a través de la mejora continua

Los estudios de capacidad del proceso tienen dos objetivos inmediatos y uno final. Los inmediatos son, primero, conocer y estimar la capacidad del proceso analizado (según nuestra definición del primer apartado) y, segundo, estimar la tasa, medida como proporción o porcentaje, de unidades defectuosas que ese proceso está produciendo. El objetivo final es diseñar un proceso, cuyos resultados (o el indicador que se esté considerando) no se aparten del valor nominal establecido o, como mínimo, que reduzcan lo más posible la tasa de productos defectuosos, para que todos se encuentren dentro de las especificaciones de calidad. Dicho de otra manera: optimizar al máximo el proceso analizado. En el sistema Seis Sigma, se establece como máximo 3.4 defectos por millón. Para ello, los estudios de capacidad del proceso se estructuran en una serie de pasos en los que se utilizan las técnicas y mediciones de acuerdo con los objetivos expuestos, como veremos a continuación.

5. Fases de un estudio de capacidad del proceso

Con el objetivo final de optimizar el proceso en estudio, sea dentro de un ciclo de mejora antes de someterlo a monitorización o como parte de la fase de experimentación o monitorización dentro de las actividades de diseño de la calidad, podemos distinguir las siguientes fases en los estudios de capacidad del proceso:

1. Asegurarnos de que el proceso está bajo control estadístico, es decir, que su estabilidad o funcionamiento está dentro de unos límites estadísticamente predecibles. Esto implica, como vimos en los capítulos anteriores, eliminar todas las causas *especiales* hasta reducir la variabilidad a su estado *natural*.
2. Estimar la extensión del proceso, es decir, calcular la media y desviación estándar de

En el caso de indicadores cualitativos no pueden establecerse tolerancias.

Las especificaciones de diseño e inspección deben ser concordantes.

Los límites de tolerancia se corresponden con los límites de las especificaciones, y pueden establecerse en ambos lados o en uno de los lados (superior o inferior) del valor nominal.

Los límites naturales de un proceso son la gama de valores que produce en su funcionamiento normal. Son una característica del proceso, con independencia de la tolerancia establecida, y comúnmente se expresan en términos de desviaciones estándar de los valores individuales con respecto a su promedio.

La extensión de un proceso es el espacio entre sus límites naturales.

Los estudios de capacidad del proceso se realizan con los objetivos inmediatos de estimar tanto su capacidad como la tasa esperable de producción defectuosa. El objetivo final es lograr un proceso perfecto, en el que la producción defectuosa sea inexistente o se minimice al máximo.

Los estudios de capacidad del proceso constan de cinco fases:

1. Asegurarnos de que el proceso está bajo control estadístico.
2. Estimar la extensión del proceso.
3. Estimar la capacidad del proceso.
4. Estimar la tasa de producción defectuosa.
5. Actuar en consecuencia para lograr la optimización del proceso.

los valores individuales en una serie de casos producidos en estado "natural".

3. Estimar la capacidad del proceso en relación con los límites de tolerancia predeterminados. Para ello existe una serie de índices, tal como veremos después, de los cuales aplicaremos los que sean más adecuados al tipo de producto, servicio o indicador que estemos considerando.
4. Estimar la tasa de producción defectuosa.
5. Actuar consecuentemente con los resultados obtenidos, es decir, estudiar más en profundidad el proceso, rediseñarlo para conseguir su optimización, o establecer un plan de monitorización para controlar que siga produciendo dentro de las especificaciones.

A continuación veremos cada una de las fases, con excepción de la primera, puesto que ha sido objeto de los capítulos 4 y 5, e ilustraremos los cálculos y razonamientos con ejemplos de indicadores mencionados en dichos capítulos.

6. Estimación de la extensión del proceso

Como ya hemos mencionado, la extensión del proceso se describe en términos de la desviación estándar de la producción individual, y equivale a la estimación de sus límites naturales.

Como serán calculados una vez que el proceso sea estable, podemos utilizar las últimas 25-30 mediciones del indicador en estado de control, con dos alternativas para los cálculos:

Alternativa 1

Si calculamos los rangos de cada muestra medida (o incluso graficado, como se recomienda en la fase de control estadístico, cuando se pretende analizar la capacidad), la desviación estándar individual es el rango medio de las mediciones del indicador efectuadas, dividido por un factor (d2), cuyo valor depende del tamaño de la muestra empleada cada vez que hemos medido el indicador. El cuadro 6.1 contiene los valores de d2 para diversos tamaños de muestra. En resumen, calcularemos $\hat{\sigma} = s = \frac{\bar{R}}{d_2}$.

La estimación de la extensión del proceso equivale a la estimación de sus límites naturales.

Si las mediciones siguen la distribución normal, los límites naturales del proceso son el promedio $\pm 3\hat{\sigma}$.
 $\hat{\sigma}$ puede calcularse en base al rango medio o con la fórmula estadística habitual.

| Cuadro 6.1 | |
|--|----------------|
| Factores (d2) para estimar σ en los cálculos de los límites naturales de los procesos | |
| n | d ₂ |
| 2 | 1 128 |
| 3 | 1 693 |
| 4 | 2 059 |
| 5 | 2 326 |
| 6 | 2 534 |
| 7 | 2 704 |
| 8 | 2 847 |
| 9 | 2 970 |
| 10 | 3 078 |

Alternativa 2

Calcular la desviación estándar con la fórmula estadística habitual, considerando todas las mediciones del indicador, realizadas como si fuese una sola muestra.

Calcularemos $\hat{\sigma} = s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{\sum n_i - 1}}$ donde:

- x_i = cada uno de los valores individuales
- \bar{x} = la media de toda la serie de mediciones como si fuese una sola. Equivalente a $\bar{\bar{x}}$ si la muestra en cada medición ha sido constante.
- n_i = cada una de las muestras en que hemos medido el indicador.

La hayamos calculado de una manera o de otra, si las mediciones individuales siguen una distribución normal, los límites naturales suelen establecerse como $\bar{x} + 3s$, de forma que 99.73% de las unidades producidas (figura 4.3) tendrá un valor para el indicador dentro de los límites naturales de la extensión del proceso analizado.

Conviene recordar que los límites naturales no deben confundirse con los límites de control. Los límites naturales se refieren a valores individuales y los de control a grupos o muestras en los que se ha medido el indicador. Sin embargo, en el caso de los gráficos \bar{x} (valores absolutos o

individuales) coinciden ambos, límites naturales y de control, al ser las mediciones realizadas cada vez en un solo caso.

Veamos dos ejemplos de estimación de límites naturales: uno sobre un indicador tipo media y otro de valores individuales. Para ello, utilizaremos los ejemplos de la media de minutos por consulta y el retraso en la administración de trombolíticos. En ambos casos, partiremos de una situación de control estadístico. Para la media de minutos por consulta utilizaremos los datos del cuadro 5.2, pero modificando los dos últimos valores para simular una situación bajo control. Para el retraso en la administración de trombolíticos, utilizaremos los de la tercera fase del estudio representados en el gráfico de la figura 6.1, situación controlada con los valores del indicador que se muestran en el cuadro 6.2.

En el primer ejemplo, una vez modificados los dos últimos valores del cuadro 5.2, haciéndolos iguales a las mediciones 19 y 20 respectivamente, nos resulta lo siguiente:

$$\bar{R} = \frac{140}{25} = 5.6 \quad \bar{x} = \frac{217.8}{25} = 8.7 \quad s = \frac{\bar{R}}{d_2} = \frac{5.6}{2.326} = 2.4$$

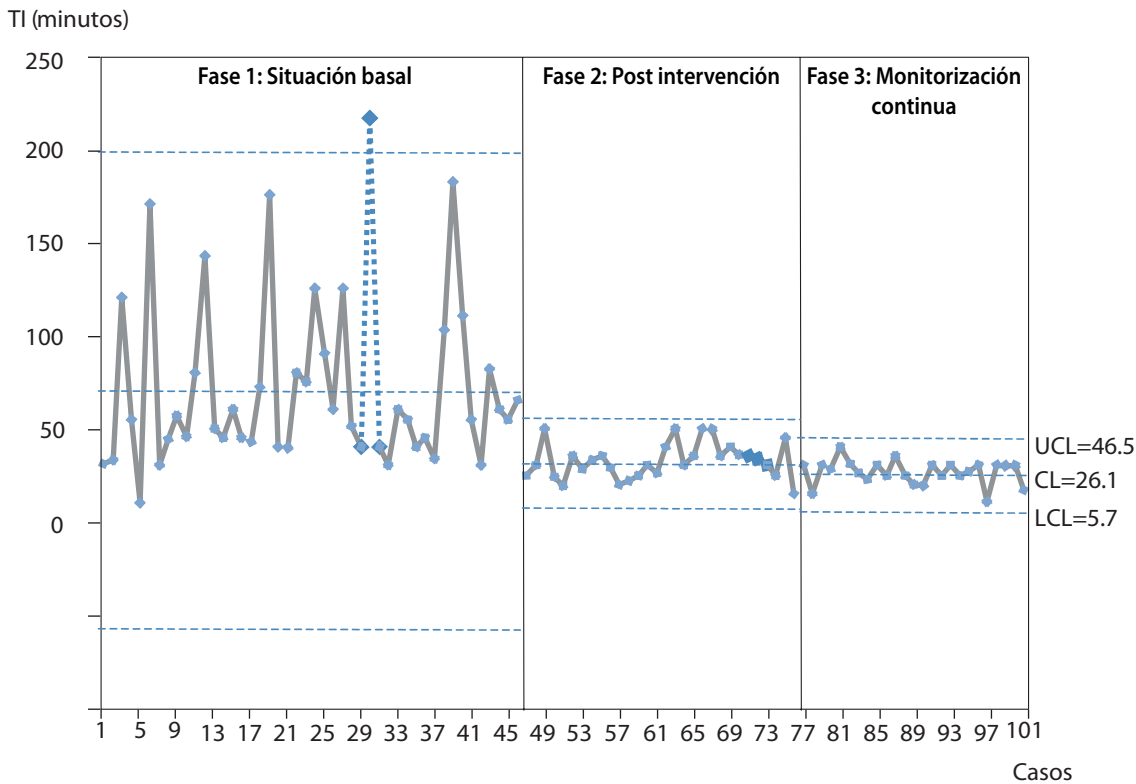
de forma que los límites naturales del proceso serían: $\bar{x} + 3s = 8.7 + 6(3 \times 2.4)$, es decir de 1.5 a 15.9.

En el segundo ejemplo (cuadro 5.2), tendríamos

$$\bar{x} = \frac{653}{25} = 26.1 \quad \overline{MR} = \frac{185}{24} = 7.7 \quad s = \frac{\overline{MR}}{1.128} = 6.8$$

y la extensión del proceso o límites naturales: $26.1 + (3 \times 6.8)$, es decir 5.7 y 46.5.

Los límites naturales se refieren a valores individuales, mientras que los límites de control se refieren a valores medios, calculados con muestras o grupos de casos.



Fuente: Saturno PJ, Felices F, Segura J, Vera A, Rodríguez JJ, Grupo ARIAM. Reducing time delay in the thrombolysis of myocardial infarction: an internal quality improvement project. Am J Med Qual 2000; 15(3):85-93.

Figura 6.1. Gráficos de control del proyecto de mejora para reducir el retraso en la administración de trombolíticos a pacientes con infarto

Cuadro 6.2

Retraso en la administración de trombolíticos en pacientes con infarto agudo de miocardio: tiempo en minutos para cada paciente en una serie de 25 mediciones. Indicador bajo control estadístico

| Paciente núm. | Minutos desde la llegada al hospital | Rango móvil |
|---------------|--------------------------------------|-------------|
| 1 | 30 | — |
| 2 | 12 | 18 |
| 3 | 29 | 17 |
| 4 | 26 | 3 |
| 5 | 40 | 14 |
| 6 | 33 | 13 |
| 7 | 25 | 8 |
| 8 | 23 | 2 |
| 9 | 31 | 8 |
| 10 | 26 | 5 |
| 11 | 33 | 7 |
| 12 | 25 | 8 |
| 13 | 20 | 5 |
| 14 | 20 | 0 |
| 15 | 30 | 10 |
| 16 | 26 | 4 |
| 17 | 31 | 5 |
| 18 | 25 | 6 |
| 19 | 26 | 1 |
| 20 | 28 | 2 |
| 21 | 10 | 18 |
| 22 | 27 | 17 |
| 23 | 28 | 1 |
| 24 | 30 | 2 |
| 25 | 19 | 11 |
| | 653 | 185 |

¿Cómo interpretamos estos resultados? Tal como hemos dicho, si los datos siguen la distribución normal (lo cual debe comprobarse antes de sacar conclusiones), en su forma estable y actual, 99.73% de los pacientes son atendidos en visitas que duran, redondeando, de 2 a 16 minutos (primer ejemplo), y 99.73% de los pacientes que llegan al hospital en estudio y se les diagnostica de infarto de miocardio susceptible de trombolisis (segundo ejemplo) esperan, redondeando, de 6 a 47 minutos hasta recibir este tratamiento beneficioso.

En caso de que los datos no sigan una distribución normal, siempre se puede intentar normalizarlos mediante conversión matemática (recíproca, logarítmica, etc.) o utilizarse la fórmula del teorema o desigualdad de Tchebychev

(capítulo 4), según la cual, la proporción de datos entre k desviaciones estándar a ambos lados de la media es al menos $1 - \frac{1}{k^2}$. Esta fórmula es válida para cualquier tipo de distribución, y da valores de 88.9% de los datos entre la media $+3s$ y 93.7% entre la media y $+4s$.

Una vez caracterizados los límites naturales del proceso, la siguiente pregunta es cómo se comporta, es decir, ¿hasta qué punto es capaz de producir calidad en cada una de las unidades?, o, lo que es lo mismo, ¿cuál es la capacidad de este proceso?

7. Estimación de la capacidad del proceso

Existen varios índices que resumen numéricamente la relación entre las características de la extensión del proceso y la tolerancia o los límites de las especificaciones de calidad deseables; en esencia, estos índices comparan el comportamiento actual del proceso con lo que se espera de él en términos de calidad: se ponen en relación los límites naturales con los de las especificaciones.

Los índices utilizados para estimar la capacidad del proceso son el índice de capacidad del proceso (C_p), índices C_p unilaterales (C_{pi} , C_{ps}) y bilaterales con referencia al valor medio \bar{y} o nominal (C_{pk}), y la ratio o proporción de capacidad $\frac{1}{C_p}$.

7.1. Índice de capacidad del proceso

El primero y más global de todos es el llamado índice de capacidad del proceso (C_p , también llamado *process capability ratio*, CPR por algunos autores) que relaciona ambos límites, naturales y de las especificaciones, de la forma más general y simple posible:

$C_p = \frac{LES - LEI}{6\hat{\sigma}}$ donde LES y LEI son los límites de tolerancia o especificaciones superior e inferior, y $\hat{\sigma}$ ó s la desviación estándar estimada para calcular los límites naturales, de forma que $6\hat{\sigma}$ es la extensión del proceso.

Como es lógico, cualquier valor de $C_p < 1$ denota un proceso incapaz de producir de acuerdo con las especificaciones establecidas. $C_p = 1$ corresponde a un proceso marginalmente capaz; sin embargo, lo deseable es que sea > 1 . El objetivo en los procesos de producción industriales es obtener un $C_p > 1.33$.

Si los datos no siguen la distribución normal, puede utilizarse la fórmula de Tchebychev o realizar una conversión de los datos para acercarlos a la normalidad.

Cuando las mediciones siguen una distribución normal, hay una relación entre el valor de C_p y el porcentaje esperable (o partes por millón, como a veces se mide en la industria) de incumplimientos del indicador (equivalente a producto de calidad defectuosa), tal como se refleja en el cuadro 6.3. Obviamente, si cada producto defectuoso es un desperdicio y/o causará insatisfacción en los clientes que lo consumen, no pueden aceptarse valores de $C_p < 1$.

Esta interpretación del C_p es posible, únicamente, en el caso de que los datos sigan una distribución normal; si no es así, recuérdese que puede hacerse una transformación de los datos (por ejemplo, recíproca, logarítmica o de cualquier otro tipo que sea adecuado) para acercarlas a la distribución normal, o utilizar otro tipo de distribuciones de probabilidad.

Calculemos ahora el C_p en nuestros ejemplos para ver los límites naturales; para ello, necesitamos establecer los límites de tolerancia (especificaciones) aceptables. Podemos considerar, por ejemplo, en el caso de la media de minutos por consulta, que una visita de duración inferior a cinco minutos será insatisfactoria, y que raramente vamos a necesitar más de 20 minutos. De esta forma, teniendo en cuenta que 2.4 es la $\hat{\sigma}$ que calculamos anteriormente, y aplicando la fórmula correspondiente:

$$C_p = \frac{LES - LEI}{6\hat{\sigma}} = \frac{20 - 5}{6 \times 2.4} = \frac{15}{14.4} = 1.04$$

Este valor debe interpretarse como marginalmente satisfactorio para nuestras especificaciones, aun cuando el indicador está bajo control, situación que, de acuerdo con lo dicho en el capítulo anterior, podría presentarse (figura 5.7).

En el segundo ejemplo (tiempo para la *trombolisis*), el establecimiento de especificaciones puede ser más controvertido porque, en realidad, queremos que el estándar de 30 minutos sea el valor máximo, es decir el LES, y el mínimo no está determinado (cuanto menos tiempo, mejor), por lo que, de una forma un tanto arbitraria, podemos establecer el LEI en el tiempo mínimo que obtuvimos en nuestra serie de mediciones con el indicador bajo control. Este valor mínimo es 10 minutos (cuadro 6.2). Con estos datos, y puesto que $\hat{\sigma} = 6.8$, según calculamos en el apartado 5, $C_p = \frac{30 - 10}{6 \times 6.8} = \frac{20}{40.8} = 0.5$. Éste es un valor claramente insatisfactorio, pues denota un proceso incapaz de producir de acuerdo con las especificaciones establecidas. $C_p < 1$ como éste es excesivamente frecuente; abundan los casos individualmente considerados, en los que el indicador no se cumple dentro de las especificaciones deseadas.

El C_p es un índice que mide de forma global la capacidad potencial del proceso para producir calidad; pero no informa de hasta qué punto tiene su media centrada o no en torno al valor nominal (en caso de que nos interesen especificaciones bilaterales), o la medida en que excede o no uno de los límites en concreto (el superior o el inferior), que, además, puede ser el único que nos interesa. Para valorar estas situaciones se han desarrollado los otros índices que vemos a continuación.

La capacidad del proceso se estima con una serie de índices que relacionan los límites naturales con las especificaciones; es decir, la extensión del proceso con la tolerancia en cuanto a calidad.

El índice de capacidad del proceso (C_p) es la relación simple entre la tolerancia y la extensión del proceso. $C_p < 1$ denota un proceso incapaz de producir de acuerdo con las especificaciones.

Cuadro 6.3
Relación entre C_p y tasa esperable de defectos

| C_p | Estándar esperable (% Incumplimientos) | Partes por millón (Incumplimientos) |
|-------|--|-------------------------------------|
| 0.5 | 13.36 | 130000 |
| 0.75 | 2.44 | 24400 |
| 1.00 | 0.26 | 2500 |
| 1.25 | 0.02 | 200 |
| 1.33 | 0.003 | 30 |
| 1.50 | 0.001 | 10 |
| 1.63 | 0.0001 | 1.0 |
| 1.75 | 0.00001 | 0.1 |
| 2.00 | 0.0000002 | 0.002 |

El C_p mide la capacidad potencial, de forma que $C_p < 1$ son incapaces, pero $C_p > 1$ pueden no serlo también si el proceso no está centrado y se excede en alguno de los límites de la tolerancia.

Fuente: Adaptado de Ott ER, Schilling EG. Process Quality Control, Nueva York, 1990.

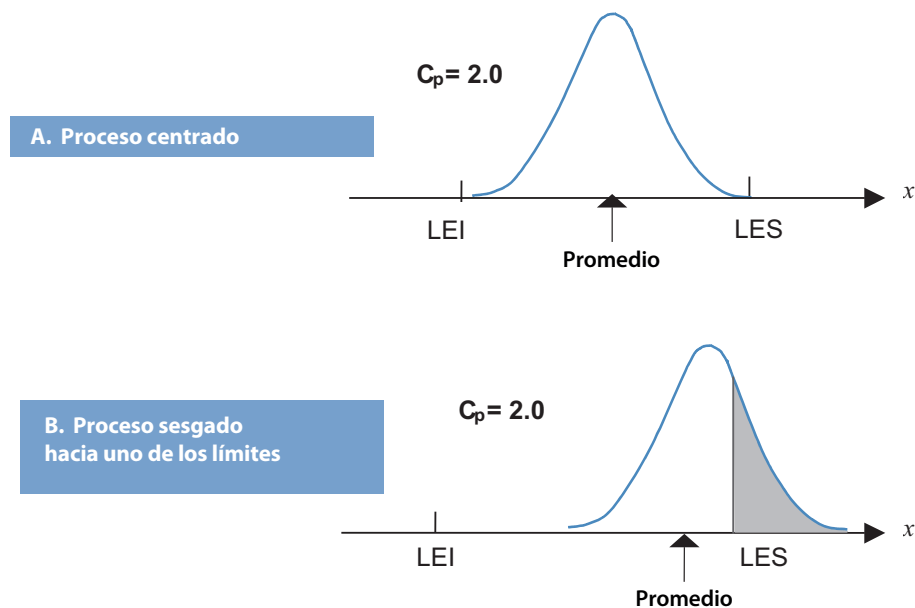
7.2. Índices de capacidad del proceso unilaterales y bilaterales con referencia al promedio y valor nominal

Así como el control estadístico del proceso es un requisito, aunque no garantice una correcta capacidad del proceso, un $C_p > 1$ indica que el proceso es potencialmente capaz, pero no es suficiente para concluir que, efectivamente, lo es. Para ello hay que tener en cuenta dónde se sitúa el valor medio del proceso, en relación con el valor nominal y las especificaciones. En la figura 6.2 pueden verse dos procesos con $C_p > 1$ (en ambos casos la

tolerancia, diferencia entre los límites de especificación, es mayor que la extensión del proceso), pero el proceso B, puesto que la media está sesgada a uno de los extremos, el proceso está produciendo unidades defectuosas con valores por encima del límite de especificación superior.

Los índices de capacidad superior (C_{ps}), inferior (C_{pi}) y bilateral (C_{pk}) dan una visión más completa y real de la capacidad del proceso, al tener en cuenta la ubicación del valor medio, y sirven para evaluar sólo uno de los lados, en caso de que sólo consideremos como defectuosas las unidades que superen uno de los extremos (figura 6.3).

Los índices de capacidad superior, inferior y bilateral (C_{ps} , C_{pi} , C_{pk}) dan una información más completa y real, al tener en cuenta la ubicación del promedio.



Fuente: Adaptado de Farnum, NR

Figura 6.2. Importancia de la ubicación del promedio para la capacidad del proceso

LEI: límite de especificación inferior
 LES: límite de especificación superior
 LNI: límite natural Inferior
 LNS: límite natural superior

$$C_p = \frac{LES - LEI}{6\hat{\sigma}} = \text{índice de capacidad de proceso}$$

C_{pi}, C_{ps} = índices de capacidad del proceso inferior y superior

$$C_{pi} = \frac{\bar{X} - LEI}{3\hat{\sigma}} \quad C_{ps} = \frac{LES - \bar{X}}{3\hat{\sigma}}$$

$$C_{pk} = \min\{C_{ps}, C_{pi}\}$$

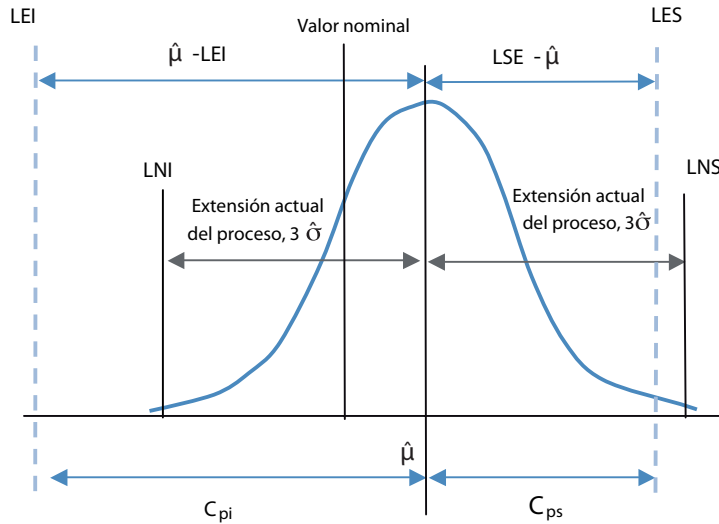


Figura 6.3. Modelo de análisis gráfico de la capacidad del proceso

$$C_{pi} = \frac{\bar{X} - LEI}{3\hat{\sigma}}$$

$$C_{ps} = \frac{LES - \bar{X}}{3\hat{\sigma}}$$

$$C_{pi} = \frac{\bar{x} - LEI}{3\hat{\sigma}} = \frac{8.7 - 5}{3 \times 2.4} = 0.51$$

$$C_{ps} = \frac{LES - \bar{x}}{3\hat{\sigma}} = \frac{20 - 8.7}{3 \times 2.4} = 1.57$$

$$C_{pk} = \min\{C_{pi}, C_{ps}\} = 0.51$$

donde $\hat{\mu}$ es la media estimada del proceso (\bar{x} en indicadores tipo media y \hat{x} en indicadores de valores individuales), y $\hat{\sigma}$ es la desviación estándar individual estimada (s), LEI es el límite de especificación inferior y LES es el límite de especificación superior.

Para que el proceso sea capaz, ambos (C_{pi} y C_{ps}), o el del lado que nos interese, han de ser >1 ; 1.33 es un valor comúnmente utilizado como objetivo o valor de referencia en el ámbito industrial.

Como resumen de la capacidad del proceso en caso de que estemos interesados por ambos límites de especificaciones y teniendo en cuenta dónde se sitúa el promedio, se utiliza el índice bilateral C_{pk} , cuyo valor es el más pequeño de los dos índices unilaterales: $C_{pk} = \min\{C_{pi}, C_{ps}\}$. Naturalmente, C_{pk} ha de ser >1 (deseable: un mínimo de 1.33) para que aceptemos el proceso como capaz.

En los dos ejemplos que venimos utilizando tendríamos los siguientes valores. Para los minutos por visita, con los datos que veíamos para el cálculo de la extensión del proceso y del C_p :

La conclusión sería que estamos ante un proceso incapaz de cumplir las especificaciones (al ser $C_{pk} < 1$), y que además está descentrado, con un sesgo hacia el LEI (al ser $C_{pi} < 1$), en cuyo extremo se están produciendo la mayoría, si no todos, los casos que incumplen las especificaciones.

Para el caso del retraso de la administración de trombolisis, podemos igualmente calcular C_{pk} , aunque lo que en realidad nos interesa es considerar el límite de especificación superior, porque el inferior puede sobrepasarse sin que el caso sea de calidad "defectuosa"; en realidad, cuanto menos tiempo mejor. De este modo, con los datos ya vistos en los apartados anteriores, calcularemos C_{ps} :

$$C_{ps} = \frac{LES - \bar{x}}{3\hat{\sigma}} = \frac{30 - 26.1}{3 \times 6.8} = 0.19$$

Es decir, un proceso incapaz y descentrado, sesgado, hacia el límite superior.

Para que un proceso sea capaz, tanto C_{pi} como C_{ps} (o el del lado que nos interese, en caso de que sólo nos interese uno) deben ser >1 .

C_{pk} resume la capacidad del proceso, si estamos interesados en que no se excedan las especificaciones de ambos lados. Para considerarlo capaz, debe ser >1 .

Si los datos no siguen la distribución normal, C_{pk} se calcula en función de C_p y del llamado factor k .

En ambos casos, hemos asumido que el proceso sigue una distribución normal. Si no fuese así, C_{pk} se calculará en función del llamado factor k , cuya fórmula es :

$$k = \frac{(LES + LEI)/2 - \hat{\mu}}{(LES + LEI)/2} \quad \text{y} \quad C_{pk} = (1 - k)C_p$$

De esta forma, k , siempre utilizado como valor absoluto, toma valores entre 0 y 1, y C_{pk} es igual a C_p precisamente cuando el proceso está centrado entre los límites de especificación.

7.3. Ratio de capacidad y tercera generación de indicadores de capacidad del proceso

Otra forma relativamente común de caracterizar la capacidad del proceso, una vez que sea aceptable, es calcular la ratio de capacidad (C_{pr}) o proporción (porcentaje si lo queremos así) de la tolerancia (espacio entre especificaciones) utilizado para el proceso. La idea es cuantificar la relación entre la extensión de este último y el espacio entre los límites de las especificaciones, que sería el máximo de extensión permitida. La fórmula sería, entonces, la inversa de C_p :

$$C_{pr} = \frac{\text{Extensión del proceso}}{\text{Tolerancia}} = \frac{6 \hat{\sigma}}{LES - LEI} = \frac{1}{C_p}$$

Obsérvese que cuanto mayor sea C_{pr} , menor es C_p , lo cual es lógico y mejor, pues nos indica que el proceso "utiliza" una menor proporción de la tolerancia o gama de valores permitida. Obsérvese también que se trata de un índice que sólo tiene sentido calcular cuando la extensión del proceso sea menor, o cuando mucho igual, que la tolerancia, lo cual corresponde con $C_p/1$.

Sin embargo, en todos los índices que hemos visto para caracterizar la capacidad del proceso, no hemos considerado un elemento relevante: su valor nominal. Su consideración explícita hizo surgir, a finales de los ochenta y principios de los noventa, lo que algunos autores llaman la *tercera generación* de índices (C_p sería la 1ª, C_{pk} la segunda), con fórmulas en las que intervienen los límites de especificación, la media del proceso y su desviación respecto al valor nominal: los índices C_{pkm} o PCR_{km} . Estos proporcionan información con mayor sensibilidad sobre la presencia de

procesos no centrados y de su grado de desviación respecto al valor nominal. Una explicación somera de estos nuevos índices, aunque suficientemente detallada, puede encontrarse en el capítulo 9 del libro de Montgomery.

8. Estimación de la tasa de producción defectuosa

Una vez que sabemos numéricamente hasta qué punto el proceso es capaz de producir según las especificaciones de calidad, podemos estimar de forma explícita qué fracción de unidades producidas tiene calidad defectuosa. Se trata de realizar los cálculos cuyo resultado vimos en el cuadro 6.3, para cada uno de los posibles valores de C_p . Esta información nos da una idea más exacta de las consecuencias de la incapacidad o de los diversos niveles de capacidad del proceso analizado. Para calcularlo, lo que hemos de hacer es averiguar la probabilidad de obtener unidades de producto fuera de los límites de especificación; esta probabilidad equivale a nuestra estimación de la tasa esperable de producción defectuosa en el proceso que estamos analizando. Veámoslo más en detalle.

En el caso más sencillo y frecuente para indicadores tipo media o valores individuales en los que podemos asumir la distribución normal, lo que hacemos es buscar en las tablas de esta distribución la probabilidad de la z correspondiente a la diferencia (o distancia) estandarizada entre los límites de especificación y el promedio del proceso. La anotación matemática es la siguiente:

Para el LEI: $\text{Prob} \left(z \geq \frac{LEI - \hat{\mu}}{\hat{\sigma}} \right)$

Para el LES: $\text{Prob} \left(z \geq \frac{LES - \hat{\mu}}{\hat{\sigma}} \right)$

Primeramente, debemos calcular las diferencias estandarizadas (que sería el valor de z para cada uno de los extremos) y después buscar la probabilidad que corresponde a ese valor en los cuadros de la distribución normal.

En nuestros ejemplos, la tasa de unidades defectuosas sería como sigue:

Para la media de minutos por consulta, $LEI=5$, $\hat{\mu}=8.7$, $\hat{\sigma} = 2.4$

$$z = \frac{LEI - \hat{\mu}}{\hat{\sigma}} = \frac{5 - 8.7}{2.4} = -1.54$$

La ratio de capacidad (C_{pr}) indica la proporción de tolerancia usada por el proceso. Es inversamente proporcional a C_p , de forma que cuanto mayor es la capacidad del proceso, más pequeño ha de ser C_{pr} .

La tercera generación de índices de capacidad del proceso incluye en los cálculos la diferencia entre el promedio del proceso y el valor nominal.

Buscamos ahora en los cuadros de la distribución normal (anexo 6.1) la probabilidad de $z = -1.54$ (recuérdese que se busca la probabilidad de valores absolutos de z , sin que importe el signo, que sólo nos va a indicar a qué lado corresponde la probabilidad). Según el cuadro (anexo 6.1) la probabilidad de $z = 1.54$ es 0.062, lo cual significa que 6.2% de los casos (unidades de producto) tiene una duración de la visita por debajo del LEI, en las condiciones actuales. Si miramos ahora los casos fuera de especificaciones por exceso (lo que debemos valorar si nos interesa o no), tendríamos, para $LES = 20$

$$z = \frac{LES - \hat{\mu}}{\hat{\sigma}} = \frac{20 - 8.7}{2.4} = 4.7$$

, a lo que corresponde una tasa de unidades por encima del límite de especificación superior (probabilidad de la z calculada), inferior a uno por ciento. Si nos interesara el valor exacto, utilizaríamos tablas más detalladas.

En el ejemplo de los trombolíticos, nos interesaremos exclusivamente por la tasa de casos por encima de los 30 minutos del LES (ya establecimos que el LEI no interesa, porque cuando menos es el retraso mejor). En este caso:

$$z = \frac{LES - \hat{\mu}}{\hat{\sigma}} = \frac{30 - 26.1}{6.8} = -0.17$$

Para este valor corresponde una probabilidad de 0.433, lo cual significa que es esperable 43.3% de casos con retraso superior a los 30 minutos.

9. Tasa de producción defectuosa en procesos medidos con indicadores cualitativos y mediciones que no siguen la distribución normal

En el caso de indicadores cualitativos como los ejemplos de los gráficos p y u de los capítulos 4 y 5, la capacidad del proceso es directamente el valor de la línea central de los gráficos de control, \bar{p} y \bar{u} , y este mismo valor (una vez el proceso está bajo control) nos indica la proporción esperable de casos que no cumplen las especificaciones de calidad (gráfico p), y la media de defectos por unidad (gráfico u). Recuérdese que en estos casos de indicadores cualitativos, la calidad individual (valor nominal) es que cumpla siempre el indicador (gráfico p) y que no tenga defectos (gráfico u).

Para calcular la tasa de producción defectuosa en caso de indicadores de variable cuantitati-

va, cuyas mediciones no siguen la distribución normal, utilizaremos el mismo procedimiento que vimos para la distribución normal, pero aplicando los cuadros de probabilidad de distribución correspondiente.

10. Requisitos y precauciones para los cálculos de la capacidad del proceso

A lo largo de los apartados anteriores hemos mencionado una serie de requisitos y precauciones para que el análisis de la capacidad de los procesos se realice e interprete correctamente. Conviene recordarlos y subrayarlos.

1. El proceso tiene que ser predecible, es decir, estar bajo control estadístico. Los cálculos relativos o cualquiera de los índices de la capacidad del proceso, así como los de los límites naturales y de la extensión del proceso, no tienen ningún sentido si el proceso analizado está sujeto a variación especial, por causas asignables.

2. Es obvio, pero hay que subrayar que el segundo requisito tras el control estadístico del proceso, es haber definido de forma clara, explícita y realista lo que entendemos por calidad en el proceso analizado. En otras palabras: cuáles son las especificaciones que vamos a aceptar como producto (o servicio o aspecto concreto de ambos) de calidad. Estas especificaciones deben ser para cada unidad o caso, no como promedio o tasa general para un grupo, muestra o conjunto de casos.

3. La mayoría de los cálculos, índices e interpretaciones de estos índices en los análisis de capacidad del proceso asumen una distribución normal de los datos. Si no es así, hay que hacer los ajustes necesarios. En caso de índices medidos con variables cuantitativas, estos ajustes pueden consistir en la *normalización* de los datos mediante una conversión matemática de los valores (recíproca, logarítmica o de otro tipo), la utilización de fórmulas no dependientes de distribución de probabilidad alguna (como el teorema de Tchebychev) o, si se conoce la distribución de probabilidad a la que se ajustan los datos, la utilización de esta distribución para las fórmulas y cálculos oportunos. En el caso de indicadores medidos con variable cualitativa, el análisis de capacidad del proceso se simplifica, pero también pierde muchos matices.

La tasa de producción defectuosa del proceso analizado equivale a la probabilidad de que se obtengan unidades de producto fuera de los límites de especificación.

Si los datos siguen la distribución normal, la tasa de producción defectuosa es la probabilidad correspondiente a la z calculada como diferencia estandarizada entre los límites de especificación y el promedio del proceso.

Si los indicadores son de tipo cualitativo, la capacidad del proceso y la tasa de producción defectuosa es el valor promedio (\bar{p} y \bar{u}), una vez que el proceso está bajo control estadístico.

Si el indicador se mide con variable cuantitativa, pero no sigue la distribución normal, se puede utilizar la distribución correspondiente para calcular la tasa de producción defectuosa.

Para analizar la capacidad del proceso se necesitan unos requisitos, y tener en cuenta una serie de precauciones para los cálculos y la interpretación de los resultados de los mismos.

Entre los requisitos: 1) es preciso que el proceso sea predecible, y 2) hay que tener perfectamente definido qué entendemos por calidad en cada unidad individual para el indicador que se mide.

Entre las precauciones: 1) no confundir los límites de control con los límites naturales del proceso ni con las especificaciones de calidad; 2) ajustar los datos y cálculos si los datos no siguen la distribución normal, y 3) considerar que lo que calculamos son estimaciones puntuales que deberían acompañarse de su correspondiente intervalo de confianza.

El análisis de la capacidad del proceso es útil tanto en programas internos como en la relación con clientes y proveedores.

En programas internos, los estudios de capacidad del proceso son útiles en los ciclos de mejora, en la monitorización y en el diseño de la calidad.

ces: el valor promedio de estos indicadores, una vez en situación de control estadístico, expresa a la vez la capacidad del proceso y la tasa esperable de producción defectuosa.

4. En ningún caso deben confundirse los límites de los gráficos de control estadístico con los límites naturales del proceso (estos últimos se refieren a valores individuales), ni con los de las especificaciones o tolerancia. Los límites de control y los límites naturales coinciden en el caso particular de los gráficos \bar{x} , en los que el indicador se mide en valores individuales, pero sólo en este caso particular. Los gráficos que representan los análisis de la capacidad del proceso son del tipo que contiene la figura 3. En éstos se representa la extensión del proceso, su promedio, el valor nominal y los límites de las especificaciones, así como los valores de los índices que puedan resultar de interés.

5. Los cálculos que se efectúan en el análisis de la capacidad del proceso son estimaciones, y así las hemos llamado en este capítulo. Sin embargo no hemos mencionado explícitamente lo que ello implica: para expresarlos correctamente, habría que acompañarlas de su consiguiente intervalo de confianza, tal como ya sabemos que hay que hacer cuando medimos algo mediante una muestra. Todavía no es muy común hacerlo así en los estudios de capacidad del proceso, pero se viene insistiendo en ello últimamente por parte de algunos autores y, sin lugar a dudas, deben terminar por expresarse de esta manera. Quienes estén interesados en las fórmulas para los intervalos de confianza pueden encontrarlas en el texto de Montgomery, ya citado.

11. ¿Cómo reaccionar ante los resultados? Utilidad del análisis de la capacidad del proceso

Los estudios de capacidad del proceso han demostrado ser extremadamente útiles en el ámbito industrial. Su realización y los resultados que brindan tienen aplicación tanto en los programas de gestión de calidad internos como en las relaciones con clientes y proveedores.

11.1. Análisis de la capacidad de los procesos en programas internos

Al igual que las técnicas de control estadístico, el estudio de la capacidad del proceso es útil en diversas fases del ciclo de mejora y en la monitorización de los indicadores. En el ciclo de mejora, el estudio de la capacidad del proceso es útil para un análisis más detallado del problema (sobre todo si éste se resume en un aspecto o indicador), y también para comprobar el efecto de la intervención para mejorar. Una vez obtenida la mejora y la capacidad de proceso deseada, ésta puede monitorizarse periódicamente para asegurarnos que se mantiene en ese nivel. En todos los casos, estos estudios deben considerarse como un complemento de las técnicas de control estadístico vistos en los capítulos 4 y 5. De hecho, el control estadístico de los indicadores puede contemplarse, como vimos en el apartado 4 de este capítulo, como la primera e ineludible fase de los estudios de capacidad del proceso (única fase en realidad si los indicadores son de tipo cualitativo).

En relación con lo que indican los valores en sí de C_p que podemos obtener, según varios autores, un proceso no está listo para monitorización continua hasta que su C_p sea <1.3 ; valores <1 son claramente objeto de ciclo de mejora, y entre 1 y 1.3 deberían ser objeto de estudio y rediseño para mejorar su capacidad. Serían susceptibles de mediciones (controles) esporádicas, si el C_p es >1.6 .

Sin embargo, los estudios de capacidad del proceso son particularmente útiles en las actividades de diseño de la calidad, sobre todo en la fase de experimentación para comprobar los efectos de algunos diseños o factores aislados dentro de los procesos diseñados. En las actividades de diseño, los experimentos se realizan para estudiar de una forma sistemática los efectos de variaciones en inputs o variables controlables, de forma que se identifique qué combinación de inputs o variables es el más adecuado para optimizar la producción de calidad (capacidad) del proceso diseñado. De esta forma, los índices de capacidad del proceso se convierten en el mejor referente para comparar y concluir sobre la bondad de las diversas alternativas del diseño. En la bibliografía, se recomienda como objetivo alcan-

zar y garantizar unos valores mínimos de C_p y C_{pk} tanto más exigentes cuanto más importante es el efecto, característica o indicador analizado, y más aún en procesos de nuevo diseño. Estos valores recomendados son los que se reflejan en el cuadro 6.4.

11.2. Análisis de la capacidad del proceso en las relaciones con clientes y proveedores

Al igual que el estándar, el umbral y los riesgos de “productor” y “consumidor” para la inspección con LQAS, es perfectamente factible también exigir y ofrecer productos o servicios con índices de capacidad de proceso, previa y explícitamente pactados. Este hecho ha sido utilizado en numerosas ocasiones dentro de ámbito de la industria como una característica positiva, una ventaja competitiva, para las empresas capaces de ofrecer y garantizar productos con un $C_p/1.33$.

Al parecer, es una práctica cada vez más común en las grandes empresas exigirles a los proveedores que demuestren una cierta capacidad del proceso como parte de los contratos de compra. Esta circunstancia puede, incluso, comprobarse empíricamente, con un acercamiento de comprobación de hipótesis en la cual la hipótesis nula sería que el proceso no es capaz y la alternativa que sí lo es. Para ello habría que identificar un cierto valor crítico (C) por encima del cual aceptaríamos la hipótesis alternativa y rechazaríamos la nula. Este valor va a ser diferente en función del tamaño de muestra que empleamos en la medición de C_p . Kane ha construido unos cuadros especialmente para estos efectos, publicados por la ASQC, en los cuales se indica el tamaño de la muestra y los valores críticos para aceptar que el proceso es capaz, con unos riesgos

α y β prefijados de forma similar a como veíamos para la aplicación del LQAS. No parece muy descabellado pensar que podrían exigirse también en los contratos de compra de servicios de salud determinados niveles de capacidad del proceso para determinados indicadores de calidad.

A la vista de su gran utilidad, demostrada y potencial, y de la variedad de situaciones y objetivos para los que está plenamente justificado, es de esperar la aplicación, a un plazo más bien corto que largo, de experiencias pioneras que muestren el beneficio de la utilización de las técnicas de análisis de la capacidad del proceso para la mejora continua de los servicios de salud. De momento, las aplicaciones que se conocen han venido de la mano de programas Seis Sigma, cuyas características revisaremos a continuación.

12. La capacidad del proceso como centro y guía de la gestión de la calidad: el sistema Seis Sigma

Los gráficos de capacidad de proceso son conocidos desde hace décadas y su utilización en los sistemas de gestión de calidad en la industria, si bien no puede decirse que sea generalizada, tampoco parece que sea poco frecuente. Sin embargo, a mediados de los ochenta, se volvieron básicas e imprescindibles para la mejora continua de los procesos, dentro de los programas Seis Sigma, ingenieros y lanzados desde la división de electrónica de Motorola Inc., por M. Harry y G. Fisher. Desde la presidencia de esta compañía, se anunció un programa de calidad cuyo objetivo era “lograr para 1992, un nivel de calidad Seis Sigma, equivalente a sólo 3.4 defectos por millón”, en sus productos y servicios (citado por Barba, Boix y Cuatrecasas). El sistema se aplicó a los productos

Valores de $C_p < 1$ nos indican que el indicador no es susceptible de monitorización continua, sino de ciclo de mejora.

Los estudios de capacidad de proceso son especialmente útiles en las actividades de diseño de la calidad, en las cuales son el mejor referente para ver los efectos de diseño alternativos.

Los índices de capacidad del proceso pueden utilizarse, incluso contractualmente, como garantía de unos determinados niveles de calidad tanto en los productos y servicios que se ofrecen como en los productos y servicios que se adquieran.

Es de esperar que, más bien pronto que tarde, empiecen a utilizarse los estudios de capacidad del proceso en la mejora continua de los servicios de salud.

Cuadro 6.4

Valores mínimos recomendados de C_p y C_{pk} , si el proceso es de nuevo diseño, y la importancia del aspecto analizado

| | C_p | C_{pk} |
|---|-------|----------|
| ▶ Procesos ya existentes en general. | 1.33 | 1.25 |
| ▶ Procesos de nuevo diseño. | 1.50 | 1.45 |
| ▶ Aspectos críticos* de procesos ya existentes. | 1.50 | 1.45 |
| ▶ Aspectos críticos* de procesos de nuevo diseño. | 1.67 | 1.60 |

*: por ejemplo, relativas a seguridad, relación directa con complicaciones, malos resultados, etcétera.

Fuente: Adaptado de Montgomery DC.

Los gráficos de capacidad de proceso fueron convertidos en algo básico e imprescindible para la mejora continua, dentro de los programas Seis Sigma, ideados en la segunda mitad de los ochenta en Motorola Inc.

El sistema Seis Sigma es básicamente un método de mejora continua, mediante la reducción de la variabilidad, que se inicia y cuyo progreso se controla analizando la capacidad del proceso sometido a mejora.

En el modelo Seis Sigma, la capacidad del proceso se mide calculando el nivel (nº) de sigmas dentro de los límites de tolerancia. Un nivel de calidad 6s equivale a la probabilidad de producir 3.4 dpm (una $C_p = 2$).

La estrategia Seis Sigma pretende implantar en todos los niveles de la empresa las acciones oportunas para lograr una calidad con una tasa de defectos de sólo 3.4 dpm.

finales de los clientes tanto externos como internos. Una de las repercusiones más inmediatas en muchas industrias fue la generalización de contabilizar los defectos, ya no en porcentajes, sino por millón de piezas, productos o, como veremos más adelante, oportunidades. Por otra parte, el objetivo Seis Sigma representaba un avance gigantesco en los objetivos de calidad de las empresas, cercano al “cero defectos” que preconizaba Crosby. La propia Motorola partía de una situación de 35 000 defectos por millón (¡más de 10 000 veces el objetivo Seis Sigma!). Muchas otras empresas se han ido sumando a la iniciativa, y algunas, como General Electric, tienen una actividad importante para ayudar a su puesta en práctica en servicios de salud.

¿En qué consiste el sistema Seis Sigma? Básicamente es un método de mejora continua mediante la reducción de la variabilidad, que se inicia y cuyo progreso se controla analizando la capacidad del proceso sometido a mejora. No supuso el descubrimiento de ninguna metodología nueva, puesto que combina técnicas y métodos ya conocidos en los sistemas de gestión de calidad. Lo original es precisamente el formato y la combinación que propone, con la capacidad de procesos como eje central. A pesar de no ser nada nuevo, M. Harry registró el método como propio, con el nombre de Six Sigma, y lo *vendió* como nuevo paradigma para la gestión de la calidad, rodeado además de toda una parafernalia de términos y propuestas de estructura que fueron y son la base de una floreciente industria de formación y consultoría en todo el mundo. Veamos lo esencial de ambos aspectos, método y terminología, de los programas Seis Sigma.

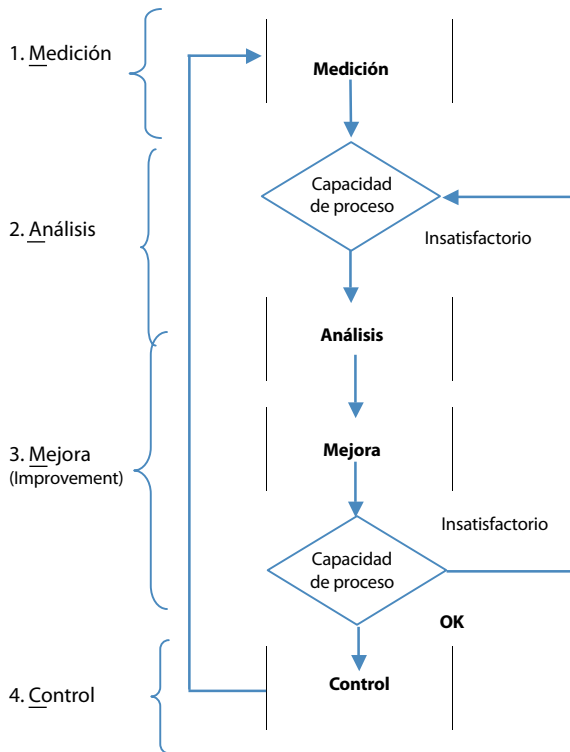
12.1. Conceptos y método de los programas Seis Sigma

Seis Sigma supone tanto un sistema de mejora continua de base estadística como una filosofía de la propia gestión de la calidad. Sigma (s) es, naturalmente, la desviación estándar del indicador en su estado natural. Esta desviación se compara con los límites de tolerancia establecidos para la calidad del producto o servicio en consideración, y lo que se propone es calcular cuántas sigmas caben dentro de los límites de tolerancia. Si las mediciones siguen una distribución normal y el valor medio está centrado en el valor nominal,

cuanto más pequeño sea s, mayor será el número de s que cabe dentro de los límites de tolerancia, y menor la probabilidad de producir piezas, productos o servicios defectuosos, como vimos al describir e interpretar los significados de los diversos índices de capacidad de proceso. En el modelo Seis Sigma, la capacidad del proceso se mide calculando el nivel (nº. de sigmas dentro de los límites de tolerancia), admitiendo un desplazamiento de m de 1.5 s con respecto y a ambos lados del valor nominal, de forma que un nivel de calidad 6s equivale a la probabilidad de producir 3.4 defectos por millón (una $C_p = 2$). Más que los índices de capacidad del proceso, los parámetros que interesan en el sistema Seis Sigma son los defectos por millón (dpm) y el nivel de sigmas que se alcanza. La estrategia Seis Sigma podría definirse como la implantación en todos los niveles de la empresa de un modelo de mejora de procesos en el que, utilizando el conocimiento estadístico de los datos, se implementan las acciones oportunas para lograr una calidad que represente un índice de defectos de sólo 3.4 dpm.

El proceso (ciclo) de mejora se inicia con el conocimiento de la capacidad del proceso (nivel de sigmas), seguido de su análisis en caso de no ser satisfactorio, y las intervenciones para mejoras que se vean oportunas, cuya efectividad se comprueba con un nuevo análisis de la capacidad del proceso. El producto o servicio sometido a mejora queda listo para control, cuando se alcanza el nivel de Seis Sigma. Como puede verse, el enfoque y esquema en sí no es diferente de lo que hemos visto para los ciclos de mejora, con la particularidad de utilizar el análisis de la capacidad de proceso para dimensionar el problema y controlar el progreso logrado con las intervenciones para mejorar. No obstante, el método Seis Sigma ha propuesto su propio esquema (MAIC: *measurement, analysis, improvement, control*: medición, análisis, mejora, control) que no deja de ser un esquema completo de ciclo de mejora y monitorización, y que ya ha empezado a tener modificaciones para incorporar el diseño para convertirse en DMAIC (define, measure, analyse, improve, control). El método o ciclo MAIC está representado en la figura 6.4.

Otra característica distintiva del sistema Seis Sigma es la necesidad de contar con un amplio colectivo de empleados con conocimiento y dominio de los métodos estadísticos, así como de las



Adaptado de: Barba E. Bpox F. Cuatrecasas L: Seis Sigma. Una iniciativa de calidad total. Gestión 2000. Barcelona 2001.

Figura 6.4. Fases del ciclo MAIC

herramientas para el diseño y los ciclos de mejora necesarios para la implementación del programa. Adicionalmente, según el nivel de su puesta en práctica, se utiliza una terminología jerárquica similar a la de los maestros de karate: *cinturón negro*, *cinturón verde*, *maestro cinturón negro* o *primer Dan* y los *campeones*.

Los cinturones negros (*black belts*) son expertos en el programa y le dedican tiempo completo; identifican oportunidades de mejora y lideran, inspiran y entrenan a los equipos; deben tener conocimientos sólidos en estadística, resolución de problemas y gestión de calidad en general. Se recomienda que constituyan el uno por ciento de los empleados, de modo que haya entre dos y cuatro en cada una de las áreas importantes de la empresa.

Los cinturones verdes (*green belts*) están formados en la metodología Seis Sigma y sirven de apoyo a los cinturones negros en los equipos de mejora. Deben manejar todas las herramientas

para la mejora continua, aunque no conozcan a profundidad la base teórica de las mismas.

Los maestros cinturón negro sirven de entrenadores, mentores y consultores de los cinturones negros; deben poseer, por lo tanto, gran conocimiento y experiencia en todo lo concerniente a Seis Sigma y a las características concretas de la empresa y sector industrial en los que laboran.

Los campeones inician y patrocinan a los cinturones negros y son los directivos responsables del programa Seis Sigma en la empresa.

En una empresa, el entrenamiento de un grupo de cinturones negros por parte de la Six Sigma Academy de M. Harry, o la obtención de una licencia para hacerlo es considerablemente caro, con cifras de un mínimo de un millón de dólares. Por lo que, hasta ahora, ha sido un sistema restringido a las grandes empresas, que no han dudado en pagarlos, debido a los enormes beneficios en términos de disminución de los costes de la falta de calidad y el incremento de la satisfacción de los clientes que, al parecer, se ha logrado con su implementación.

Las herramientas utilizadas en los procesos de mejora de los programas Seis Sigma incluyen todas las que hemos visto para los ciclos de mejora (diagrama de Ishikawa, pareto, histograma, etc.), además de, naturalmente, las de control estadístico de la calidad y las propias del diseño de la calidad, como pueden ser el diseño de experimentos y el análisis del modo de fallos y efectos.

Merece la pena, sin embargo, que nos detengamos en la revisión del cálculo del nivel de sigmas de un proceso, puesto que complementa y simplifica, resumiendo, el análisis estadístico de la capacidad del proceso revisados en este capítulo, además de servir de base y punto de partida para los proyectos de mejora del programa Seis Sigma y para monitorizar su avance y mantenimiento.

12.2. Cálculo e interpretación del nivel en sigmas de la calidad de un producto o servicio

El primer paso es identificar y definir los indicadores o aspectos que deben medirse para valorar la calidad del producto o servicio. En la terminología Seis Sigma se conocen como los elementos críticos para la calidad (CTQs: *critical to quality elements*).

A continuación se miden los defectos por millón que se encuentran en una muestra sufi-

En terminología Seis Sigma los ciclos de mejora se conocen como MAIC (*Measurement-Analysis-Improvement-Control*) y se inician y controlan con el conocimiento de la capacidad del proceso (nivel de sigmas) sometido a mejora.

El sistema Seis Sigma precisa de un amplio colectivo de empleados con dominio de los métodos de control estadístico y las demás herramientas para el diseño y la mejora de la calidad.

La estructura de los recursos humanos para el sistema Seis Sigma adopta la terminología y la jerarquía de las artes marciales.

La implantación de un sistema Seis Sigma precisa de una importante inversión en formación, aunque compensada por los enormes beneficios en términos de disminución de costes de la falta de calidad que se obtienen al implementarlo.

El primer paso para averiguar el nivel en sigmas de la calidad de un producto o servicio es identificar los indicadores o elementos críticos para la calidad que deben medirse.

El nivel de calidad en número de sigmas se calcula a partir de los dpm encontrados en una muestra del producto o servicio por mejorar.

Cuando se miden varios aspectos, indicadores o elementos críticos para la calidad en cada caso, el resultado se expresa en defectos por millón de oportunidades.

Explicar la calidad en niveles de sigma facilita la comparación entre indicadores, productos o servicios, por muy diferentes que sean.

cientemente amplia, con el proceso bajo control estadístico, y finalmente se calcula el nivel de desempeño o calidad en sigma. Los defectos por millón pueden referirse a un solo aspecto del producto o servicio, o a varios de ellos considerados conjuntamente (a modo de *indicador compuesto* de tipo porcentaje simple), midiéndose entonces los *Defectos por millón de oportunidades*.

Según el valor de dpm encontrado, puede calcularse el nivel de sigma que corresponde en el modelo Seis Sigma (que incluye una posibilidad de desviación de m respecto al valor nominal de $1.5 s$), utilizando la fórmula:

$$\text{Nivel en sigmas} = 0.8406 + \sqrt{29.37 - 2.22 \ln(\text{dpm})}$$

o bien utilizando los cuadros ya construidos, en los que consta el nivel de calidad en sigmas, según el número de dpm. El cuadro 6.5 es una versión simplificada de estos cuadros.

Aplicándolo a nuestros ejemplos, tendríamos lo siguiente:

En el caso de la media de minutos por consulta, en el apartado siete vimos que tiene una probabilidad de 6.2% de casos defectuosos, lo que equivale a 62 000 dpm y un nivel de calidad entre 3s y 3.5s, según el cuadro 6.5. Lejos del nivel 6s, que sería el objetivo del programa.

En el caso del retraso en la administración de trombolíticos, con 43.3% de casos esperables con retraso superior a los 30 minutos (dpm = 433 000), el nivel en sigma está entre 1.5 y 2.

En caso de medir más de un elemento crítico, los dpm se calculan sobre el total de oportunidades de calidad. Por ejemplo, si se miden 10 factores críticos (como subindicadores de un indicador compuesto) en una muestra de 500 casos, y se encuentran 25 defectos de calidad, el número de dpm sería: $(25 / 500 \times 10) \times 106 = 5\ 000$ con un nivel entre 4 y 4.5s, según el cuadro 6.5.

Una ventaja adicional de valorar la calidad en niveles de sigma es que facilita la comparación entre diferentes indicadores, productos y servicios, al utilizar una misma unidad de medida-resumen: el número o nivel de sigma.

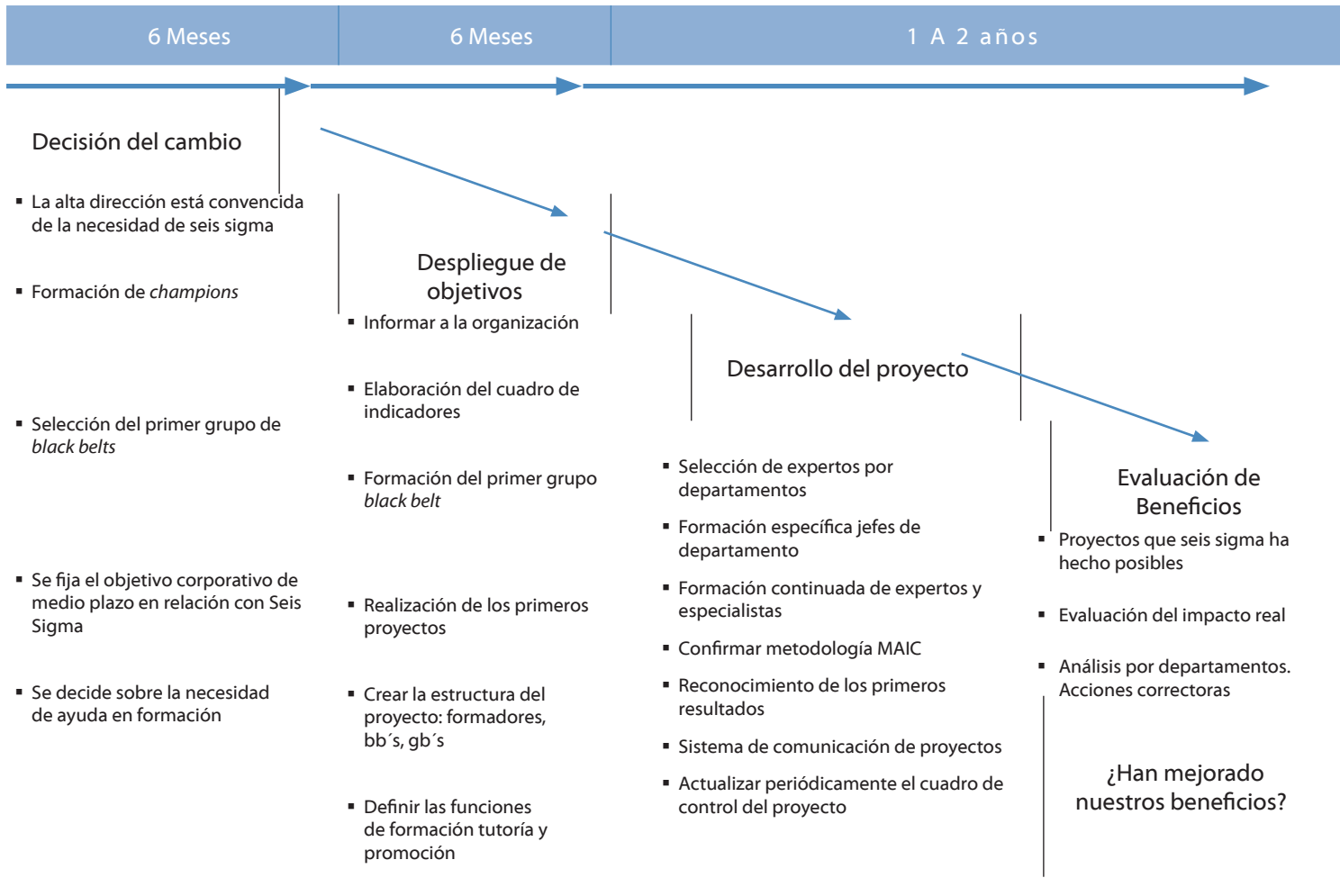
12.3. Estrategia de implantación de un sistema Seis Sigma

La estrategia de implantación de un sistema Seis Sigma no es, a grandes rasgos, diferente al de cualquier otra innovación. Naturalmente, va a ser más fácil en la medida en que haya más experiencia y compromiso en y con la gestión de la calidad en la organización. La figura 6.5 contiene un esquema, con marco temporal estimado, del proceso de implantación.

| DPMO | Núm. de sigmas* | % Buena calidad** |
|--------|-----------------|-------------------|
| 691500 | 1 | 30.85 |
| 500000 | 1.5 | 50 |
| 308500 | 2 | 69.15 |
| 158700 | 2.5 | 84.13 |
| 66800 | 3 | 93.32 |
| 22700 | 3.5 | 97.73 |
| 6200 | 4 | 99.38 |
| 1300 | 4.5 | 99.87 |
| 230 | 5 | 99.977 |
| 30 | 5.5 | 99.997 |
| 3.4 | 6 | 99.99966 |

*: Calculado según la fórmula: $\text{núm. de sigmas} = 0.8406 + \sqrt{29.37 - 2.22 \ln(\text{dpm})}$, asumiendo distribución normal y con desviación máxima de 1.5 respecto al valor nominal.

** : Equivalente al nivel de cumplimientos, cuando se trata de un solo indicador o *variable crítica* de tipo cualitativo.



Fuente: Barba E, Boix F y Cuatrecasas L. Seis Sigma. Una iniciativa de calidad total. Barcelona: Gestión 2000, 2001.

Figura 6.5. Estrategia de implantación de Seis Sigma

12.4. Seis Sigma en los servicios de salud

Como señalamos arriba, las experiencias de aplicación de los gráficos de capacidad del proceso en

los servicios de salud se han producido de la mano de proyectos Seis Sigma. Aún no son abundantes, pero el campo es obviamente muy prometedor.*

Las todavía escasas experiencias de aplicación de los gráficos de capacidad de proceso en servicios de salud se han efectuado dentro de proyectos Seis Sigma, mayoritariamente.

* Algunas de estas experiencias pueden encontrarse en el sitio <http://www.isixsigma.com>, que además contiene otro tipo de materiales de apoyo y divulgación sobre el tema, así como enlaces de interés para quien quiera profundizar en ello.

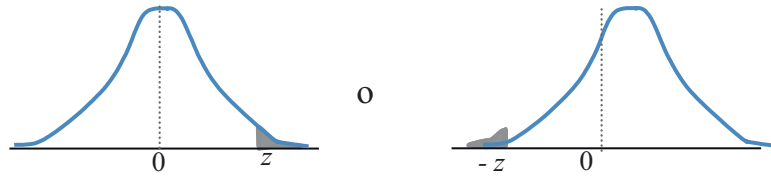
BIBLIOGRAFÍA

- Barba E, Boix F, Cuatrecasas LI. Seis Sigma, una iniciativa de calidad total. Barcelona: Gestión 2000, 2001.
- Farnum NR. Capítulo 8. En: Modern statistical quality control and improvement. Belmont: Dubbury Press, 1994.
- iSixSigma. Six Sigma Resources for Six Sigma Quality [sitio de internet] [consultado en febrero de 2015]. Disponible en: <http://www.isixsigma.com>
- Montgomery DC. Capítulo 9. En: Introduction to statistical quality control. 2ª ed. Nueva York: J. Wiley & Son, 1996.
- Ott ER, Schilling EG. Capítulo 8. En: Process quality control. 2ª ed. Nueva York: McGraw Hill, 1990.

Anexo 6.1

Probabilidades en un extremo o cola de la curva de la distribución normal

Esta tabla da el valor de la zona sombreada



| Z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|-----|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0.0 | 0.500 | 0.496 | 0.492 | 0.488 | 0.484 | 0.480 | 0.476 | 0.472 | 0.468 | 0.464 |
| 0.1 | 0.460 | 0.456 | 0.452 | 0.448 | 0.444 | 0.440 | 0.436 | 0.433 | 0.429 | 0.425 |
| 0.2 | 0.421 | 0.417 | 0.413 | 0.409 | 0.405 | 0.401 | 0.397 | 0.394 | 0.390 | 0.386 |
| 0.3 | 0.382 | 0.378 | 0.374 | 0.371 | 0.367 | 0.363 | 0.359 | 0.356 | 0.352 | 0.348 |
| 0.4 | 0.345 | 0.341 | 0.337 | 0.334 | 0.330 | 0.326 | 0.323 | 0.319 | 0.316 | 0.312 |
| 0.5 | 0.309 | 0.305 | 0.302 | 0.298 | 0.295 | 0.291 | 0.288 | 0.284 | 0.281 | 0.278 |
| 0.6 | 0.274 | 0.271 | 0.268 | 0.264 | 0.261 | 0.258 | 0.255 | 0.251 | 0.248 | 0.245 |
| 0.7 | 0.242 | 0.239 | 0.236 | 0.233 | 0.230 | 0.227 | 0.224 | 0.221 | 0.218 | 0.215 |
| 0.8 | 0.212 | 0.209 | 0.206 | 0.203 | 0.200 | 0.198 | 0.195 | 0.192 | 0.189 | 0.187 |
| 0.9 | 0.184 | 0.181 | 0.179 | 0.176 | 0.174 | 0.171 | 0.169 | 0.166 | 0.164 | 0.161 |
| 1.0 | 0.159 | 0.156 | 0.154 | 0.152 | 0.149 | 0.147 | 0.145 | 0.142 | 0.140 | 0.138 |
| 1.1 | 0.136 | 0.133 | 0.131 | 0.129 | 0.127 | 0.125 | 0.123 | 0.121 | 0.119 | 0.117 |
| 1.2 | 0.115 | 0.113 | 0.111 | 0.109 | 0.107 | 0.106 | 0.104 | 0.102 | 0.100 | 0.099 |
| 1.3 | 0.097 | 0.095 | 0.093 | 0.092 | 0.090 | 0.089 | 0.087 | 0.085 | 0.084 | 0.082 |
| 1.4 | 0.081 | 0.079 | 0.078 | 0.076 | 0.075 | 0.074 | 0.072 | 0.071 | 0.069 | 0.068 |
| 1.5 | 0.067 | 0.066 | 0.064 | 0.063 | 0.062 | 0.061 | 0.059 | 0.058 | 0.057 | 0.056 |
| 1.6 | 0.055 | 0.054 | 0.053 | 0.052 | 0.051 | 0.049 | 0.048 | 0.048 | 0.046 | 0.046 |
| 1.7 | 0.045 | 0.044 | 0.043 | 0.042 | 0.041 | 0.040 | 0.039 | 0.038 | 0.038 | 0.037 |
| 1.8 | 0.036 | 0.035 | 0.034 | 0.034 | 0.033 | 0.032 | 0.031 | 0.031 | 0.030 | 0.029 |
| 1.9 | 0.029 | 0.028 | 0.027 | 0.027 | 0.026 | 0.026 | 0.025 | 0.024 | 0.024 | 0.023 |
| 2.0 | 0.023 | 0.022 | 0.022 | 0.021 | 0.021 | 0.020 | 0.020 | 0.019 | 0.019 | 0.018 |
| 2.1 | 0.018 | 0.017 | 0.017 | 0.017 | 0.016 | 0.016 | 0.015 | 0.015 | 0.015 | 0.014 |
| 2.2 | 0.014 | 0.014 | 0.013 | 0.013 | 0.013 | 0.012 | 0.012 | 0.012 | 0.011 | 0.011 |
| 2.3 | 0.011 | 0.010 | 0.010 | 0.010 | 0.010 | 0.009 | 0.009 | 0.009 | 0.009 | 0.008 |
| 2.4 | 0.008 | 0.008 | 0.008 | 0.008 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.006 |
| 2.5 | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 |
| 2.6 | 0.005 | 0.005 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 |
| 2.7 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
| 2.8 | 0.003 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |
| 2.9 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 |
| 3.0 | 0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |

Capítulo 7

Gráficos CUSUM, gráficos EWMA y otros métodos de análisis y monitorización estadística de la calidad

Contenido general y objetivos específicos

1. Contextualización
2. Gráficos de control secuenciales
3. Principales diferencias entre los gráficos de control tradicionales y los de mediciones secuenciales
4. Elaboración e interpretación de un gráfico CUSUM (*cumulative sum*, suma acumulada)
5. Elaboración e interpretación de un gráfico EWMA (*exponentially weighted moving-average*, media móvil exponencialmente ponderada)
6. Otros métodos gráficos de análisis y monitorización estadística de la calidad

Contenido general

En este capítulo se describen dos tipos de gráficos de control estadístico especiales: los gráficos CUSUM y EWMA, que se basan en el análisis de un índice que se calcula para cada una de las mediciones del indicador, incorporando la información de mediciones anteriores. Con este proceder se detectan cambios más pequeños que con el gráfico de control tradicional y se hacen evidentes con menos mediciones. Se mencionan finalmente otros tipos de gráficos de control estadístico que se construyen para monitorizar algún aspecto de la capacidad del proceso, y no solo la estabilidad estadística.

Objetivos específicos

1. Describir las características de los gráficos de control secuenciales.
2. Diferenciar los gráficos CUSUM y EWMA.
3. Construir los índices CUSUM de una serie de mediciones de un indicador de valores individuales.
4. Analizar los datos CUSUM en la forma tabular y con gráfico de control.
5. Construir los valores EWMA de una serie de mediciones de un indicador de valores individuales.
6. Analizar una serie de mediciones con gráfico EWMA.
7. Comprender el concepto de ARL y su relación con los parámetros de los gráficos CUSUM y EWMA.
8. Describir las características y objetivos de los gráficos de precontrol, gráfico de control con límites modificados y gráfico de control de la aceptación.

Gráficos CUSUM, gráficos EWMA y otros métodos de análisis y monitorización estadística de la calidad

1. Contextualización

Después de revisar en los capítulos 4 y 5 las más conocidas y sencillas herramientas de control estadístico de la calidad (gráficos de desarrollo y gráficos de control), las únicas sobre las que hay ya una cierta experiencia de aplicación en los servicios de salud, hemos iniciado, con el análisis de la capacidad del proceso (capítulo 6), el estudio de otras herramientas de control estadístico de procesos mucho menos conocidas en el ámbito sanitario, pero de un gran interés potencial. En este capítulo completamos esta visión de las principales técnicas de control estadístico con los gráficos que se basan en análisis secuenciales: los gráficos CUSUM y EWMA. Estas herramientas son también muy poco conocidas en el ámbito de gestión de calidad en nuestro ámbito. Sin embargo, así como el análisis de la capacidad del proceso aventaja a otros métodos alternativos dentro de las actividades de diseño de la calidad, los gráficos CUSUM y EWMA serían para muchos las técnicas de elección para la monitorización continua, al permitir advertir de cambios significativos en el proceso antes que los gráficos de control tradicionales, sobre todo si los cambios son de pequeña magnitud, además de que incorporan en sus cálculos y análisis el estándar deseable para el indicador que se monitoriza. El capítulo finaliza con una breve reseña de los gráficos que combinan las nociones y objetivos del control estadístico y la capacidad del proceso, diseñados precisamente para monitorizar esta última mediante la aplicación de los principios de los gráficos de control.

2. Gráficos de control secuenciales

2.1. ¿Qué son?

Lo que analizan los gráficos de control secuenciales, aun basándose en los mismos principios estadísticos que los gráficos de control normales, no son directamente las mediciones del indicador que se pretende monitorizar, sino un índice estadístico que se crea para tal efecto, que se basa y tiene en cuenta mediciones anteriores del mismo indicador. Al aprovechar para cada punto del gráfico toda la información anterior, este tipo de gráficos es capaz de detectar cambios en la estabilidad del indicador más pequeños y de forma precoz que con los gráficos de control tradicionales. Los dos representantes más genuinos de este tipo de gráficos son los CUSUM (cumulative-sum) y EWMA (exponentially weighted moving-average).

2.2. ¿Para qué sirven?

Aunque en teoría podrían tener las mismas aplicaciones que los gráficos de control, fueron diseñados y tienen su máxima utilidad en la monitorización continua de procesos estables que, además, están en el nivel de calidad deseado, puesto que, además de aprovechar la información de toda la secuencia de mediciones que se vayan realizando, utilizan un punto de referencia, independiente de las mediciones, que normalmente debe ser el estándar en que se desea mantener ese indicador.

En este capítulo se estudian los gráficos secuenciales (CUSUM y EWMA), aún poco utilizados en la gestión de la calidad en los servicios de salud.

En los gráficos secuenciales no se analiza el resultado de la medición de los indicadores, sino un índice estadístico creado a tal efecto, que resume la información de mediciones anteriores.

Los gráficos secuenciales están indicados sobre todo en la monitorización continua de indicadores estables y en el nivel de calidad deseado.

Los gráficos secuenciales detectan cambios más pequeños antes que los gráficos de control tradicionales.

Los gráficos CUSUM analizan la suma acumulada de las diferencias entre los resultados de la medición del indicador y el estándar o valor de referencia en el que se han estabilizado.

Los gráficos EWMA analizan la media exponencialmente ponderada de la secuencia de mediciones realizadas.

En los gráficos EWMA, el estándar del indicador o valor estable deseado del indicador se utiliza como valor de referencia o primera medición de la serie.

Los límites de control se establecen en función de la desviación estándar estimada del proceso, y unos parámetros que determinan la probabilidad de detectar cambios más o menos grandes en el indicador monitorizado.

3. Principales diferencias entre los gráficos de control tradicionales y los de mediciones secuenciales

Como mencionamos líneas arriba, la principal diferencia de los gráficos de control secuenciales de los clásicos es que lo que se grafica no es el resultado de la medición del indicador, sino el resultado de la aplicación de una fórmula que incluye la medición actual y un resumen de las anteriores, con base en unos parámetros que determinan el diseño del gráfico. Tienen también unos límites de control peculiares, que no son siempre horizontales, y una mayor capacidad para detectar valores más pequeños y de forma precoz.

3.1. Índices estadísticos en vez de estimaciones del indicador

Tanto en los gráficos CUSUM como en los EWMA, el valor que se obtiene al medir el indicador se introduce en una fórmula para calcular el índice estadístico analizado.

Los gráficos CUSUM grafican la suma acumulada de las desviaciones (diferencias) de los valores que se obtienen al medir el indicador, con respecto a un valor que se considera el objetivo o estándar de ese indicador. La fórmula general para el valor analizado (anotado como C) sería:

$$C_i = \sum_{j=1}^i (\bar{x}_j - \mu_0)$$

donde C_i es el valor acumulado para la medición i (suma acumulada de las diferencias hasta esa medición), \bar{x}_j representa la media (valor del indicador) cada vez que se mide, y μ_0 es el valor estándar que queremos para el indicador. De esta forma, si el proceso permanece en control y al nivel μ_0 , los valores de C deben seguir una distribución que ha de tener cero como media.

En los gráficos EWMA, lo que se analizan son los valores que toma la fórmula:

$$\hat{x}_i = \lambda \bar{x}_i + (1 + \lambda) \hat{x}_{i-1}$$

donde \hat{x}_i es la media móvil exponencialmente ponderada, \bar{x}_i representa el valor que se obtiene en la medición del indicador, y l es una constante

(con valor entre 0 y 1), que se determina en función de la magnitud del campo que se va a detectar y de la sensibilidad que queremos que tenga el gráfico. Se establece además un valor de referencia (\hat{x}_0) para el cálculo de \hat{x} en la primera medición analizada (\hat{x}_1), que puede ser el estándar u objetivo de calidad que queremos para el indicador. Es decir, $\hat{x} = \mu_0$, siendo μ_0 el estándar, tal como vimos para el gráfico CUSUM. La fórmula descrita más arriba para calcular la EWMA para cada medición del indicador es una derivación matemática de apariencia y cálculo más simple de la siguiente fórmula, la cual es más explícita y de apariencia bastante más complicada:

$$\hat{x}_i = \lambda \sum_{j=0}^{i-1} (1 - \lambda)^j \bar{x}_{i-j} + (1 - \lambda)^i \hat{x}_0$$

en la que es más obvio que \hat{x} es una media ponderada de todas las medias anteriores, por lo que es necesario establecer un valor inicial \hat{x}_0 que, como hemos dicho, puede ser el estándar del indicador.

3.2. Límites de control para los gráficos de control secuenciales

Tanto en los gráficos CUSUM como en los EWMA, los límites de control, y por tanto la regla de decidir si el proceso está o no fuera de control, se establecen en función de la desviación estándar estimada del indicador en situación de estabilidad estadística y de unas constantes, cuyo valor determina la probabilidad de equivocarse al decidir que el proceso está fuera de control, y que se deben elegir por tanto conscientemente.

En el gráfico CUSUM las dos constantes que determinan los límites y si el indicador está o no fuera de control son: 1) la desviación permitida respecto al estándar o valor de referencia (anotada como K), y 2) el llamado intervalo de decisión (anotado como H), por fuera del cual se concluirá que el proceso se ha descontrolado. Un valor comúnmente utilizado para K es la mitad de la diferencia entre el estándar (μ_0) y el valor fuera de control que queremos detectar con rapidez si aparece (μ_F); es decir:

$$K = \frac{|\mu_F - \mu_0|}{2}, \text{ que suele expresarse como unidades desde s: } K=k.s.$$

Una vez establecido K , se calcula C para cada medición del indicador, distinguiendo entre diferencias por encima de μ_0 ($C+$) y por debajo de μ_0 ($C-$), llamados respectivamente CUSUM superior ($C+$) y CUSUM inferior ($C-$), de la siguiente manera:

$$C_i^+ = \max \left\{ 0, x_i - (\mu_0 + K) + C_{i-1}^+ \right\}$$

$$C_i^- = \max \left\{ 0, (\mu_0 - K) - x_i + C_{i-1}^- \right\}$$

Es decir, para cada valor de x_i tanto C_i^+ como C_i^- van a ser 0 o el valor que resulte de las operaciones que se indican, si este valor es mayor que cero. Cuando $C+$ o $C-$ exceden el intervalo de decisión (H), el proceso se considera fuera de control. ¿Cómo se elige H ? Se recomienda un valor de 4σ o 5σ . En cuanto a su utilización, ya veremos después que hay dos formas de dibujar y analizar los gráficos CUSUM, aunque sólo revisaremos en detalle la más sencilla.

Para los gráficos EWMA, además de λ , hay que decidir el valor de otro parámetro conocido como L , que determina la amplitud de los límites de control, cuya fórmula es:

$$LCS = \mu_0 + L \sigma \sqrt{\frac{\lambda}{(2-\lambda)}} \left[1 - (1-\lambda)^{2i} \right]$$

$$LCI = \mu_0 - L \sigma \sqrt{\frac{\lambda}{(2-\lambda)}} \left[1 - (1-\lambda)^{2i} \right]$$

Donde i es el número de medición del indicador, en su secuencia de mediciones, con lo cual, los límites serán cambiantes conforme avancemos en mediciones; pero también, según puede deducirse de la propia fórmula, la expresión $\left[1 - (1-\lambda)^{2i} \right]$ se aproxima a la unidad conforme aumenta i , de manera que, tras una serie de mediciones, los límites de control se estabilizan en:

$$\mu_0 \pm L \sigma \sqrt{\frac{\lambda}{(2-\lambda)}} \left[1 - (1-\lambda)^{2i} \right]$$

¿Cómo se eligen λ y L ? Aunque lo discutiremos más en detalle en un próximo apartado, podemos

adelantar que una combinación posible es $\lambda=0.2$ y $L=3$.

3.3. Capacidad para detectar situaciones fuera de control

Tanto el CUSUM como el EWMA están diseñados para advertir la aparición de cambios significativos respecto a la media estable del proceso, de una magnitud $<1.5s$. Recuérdese que en los gráficos de control, para que un solo punto sea considerado fuera de control, éste debe ser mayor que el promedio $+3s$; recuérdese también que, para aumentar la capacidad de detectar cambios, se tienen en cuenta en los gráficos de control series de puntos, dando lugar a los patrones de secuencias de puntos significativos de descontrol (capítulo 5).

Esta manera de proceder se incorpora a los gráficos CUSUM y EWMA, pero al tener en cuenta la información anterior, los patrones de descontrol los denota siempre un solo punto, lo cual es más sencillo de interpretar que los patrones de los gráficos de control, además de permitir actuaciones más precoces. Sólo en el caso de desviaciones relativamente grandes respecto al promedio van a ser preferibles los gráficos de control tradicionales, en cuanto a prontitud en la detección de cambios significativos en un indicador, cuando éste es sometido a monitorización continua. Por tal motivo, en la industria no es inusual la combinación de gráficos de control tradicionales y CUSUM o EWMA en la monitorización de indicadores. Parece ser, además, que estos últimos son especialmente útiles y necesarios para indicadores que se miden como valores absolutos o individuales (es decir de $n=1$), correspondientes a gráficos \bar{x} en los gráficos de control tradicionales.

4. Elaboración e interpretación de un gráfico CUSUM (cumulative sum, suma acumulada)

Los gráficos CUSUM aparecieron en la gestión de calidad en los años cincuenta de la mano de E. S. Page. Desde esa fecha han surgido múltiples modificaciones, matices y aplicaciones nuevas de gran interés. En la actualidad, hay dos formas de analizar numérica y gráficamente los datos CUSUM: con plantilla V (V -mask), y de forma tabular (o algorítmica) con o sin representación gráfi-

En el gráfico CUSUM se calcula el valor acumulado superior ($C+$) e inferior ($C-$), según las diferencias por encima y por debajo de valor de referencia (m_0).

Los dos parámetros de diseño del gráfico CUSUM son K (relacionado con el tamaño de la diferencia respecto al valor de referencia que queremos detectar) y H (intervalo decisional, que establece los límites de control). Ambos se expresan en unidades de s .

En el gráfico EWMA hay que elegir los valores de dos parámetros, λ y L , el primero de ellos relacionado con la magnitud del cambio que queremos detectar, y el segundo que determina los límites de control.

λ tiene valores entre 0 y 1, mientras que L son unidades de s .

Los gráficos CUSUM y EWMA son útiles para detectar cambios significativos del indicador, de magnitud $<1.5s$.

Los patrones de descontrol son siempre un punto por fuera de los límites de control.

Es normal e incluso conveniente, combinar los gráficos de control tradicionales con gráficos secuenciales.

Los datos CUSUM pueden analizarse con la llamada plantilla V, o en forma algorítmica y un gráfico de control estructuralmente semejante al gráfico de control tradicional.

En la plantilla V, el vértice de la V se alinea con las últimas mediciones del indicador y los brazos de la V son los límites de control.

ca que puede tener una forma similar a un gráfico de control clásico.

La plantilla V, llamada así por su forma, consiste en una V con vértice alineado con el último valor de C, y los límites de control representados por los dos brazos de la V. Esta plantilla se va re-locando a medida que se añaden nuevas mediciones; la situación fuera de control se detectará cuando aparezca un punto fuera de los límites de control (los brazos de la V). Este proceder, aunque es el más tradicional, es más complicado que la forma tabular, y tiene además una serie de desventajas (por ejemplo, aparte de otras cuestiones, la plantilla V es siempre de dos lados y está calculada así, siendo poco aplicable si estamos interesados en las desviaciones sólo hacia un lado o queremos establecer límites diferentes, no simétricos, en los dos lados) por lo que se empieza a recomendar hoy día utilizar preferentemente la forma tabular, misma que veremos a continuación.

Para elaborar los datos CUSUM con arreglo a las fórmulas expuestas más arriba (apartado 2 de este capítulo) y ser analizados en términos de monitorización de la estabilidad del indicador estandarizado, primeramente es necesario definir los parámetros H y K, siendo $H=h.s$ y $K=k.s$. La selección de uno u otro valor determina las características de CUSUM en cuanto a capacidad de detección de cambios en el indicador, y se deciden en función de qué tan grande es el cambio que quiera detectarse y qué tan pronto. Esta mayor o menor precocidad en la detección de cambios se mide en términos del número de muestras (mediciones del indicador) necesario para que el cambio se detecte, un concepto conocido en la industria como ARL del gráfico o del sistema de monitorización de que se trate ($ARL=average run$

length, longitud o extensión media de las rachas o serie de mediciones seguidas). Hay tablas construidas en las que se han simulado diversas combinaciones de h y k, y se expresa el ARL que corresponde a cada cambio por detectar, de manera que de antemano se sepan las consecuencias de elegir uno u otro valor de h y k. Una combinación de valores bastante común es $k=0.5$ y $h=5$ o 4. Los ARL que implican diversos tamaños de desviaciones o cambios por detectar en el valor del indicador se muestran en el cuadro 7.1. Así, por ejemplo, para detectar un cambio de una magnitud equivalente a 1s, hace falta un promedio de 8.38 mediciones seguidas del indicador, si fijamos k en 0.5 y h en 4; este promedio de mediciones por realizar para detectar el cambio sería de 10.4 si elegimos $h=5$.

Una vez fijados h y k podemos proceder al análisis y monitorización del indicador con CUSUM. Ilustraremos este proceder aplicándolo al ejemplo del retraso a la administración de trombolíticos, con los mismos datos del capítulo 6.

Para nuestro ejemplo, establecemos como valor objetivo o estándar del indicador mantenernos en el nivel medio alcanzado y estable (26.1 minutos) y aplicamos las fórmulas para el cálculo de C+ y C- vistas más arriba, llevando los cálculos a una tabla como la que se muestra en el cuadro 7.2, que contiene los resultados del ejemplo que vamos a utilizar. Las dos primeras columnas del cuadro son simplemente la secuencia de mediciones realizadas (en este caso, 29) con sus respectivos resultados (xi) en la misma secuencia en la que los hemos obtenido. Las tres columnas siguientes representan los cálculos para el CUSUM superior (C+) y las tres últimas para el CUSUM inferior (C-).

Cuadro 7.1
ARL del análisis CUSUM de un indicador, con $k=0.5$ y diversos valores de h y del cambio por detectar

| Cambio del promedio (múltiplos de s) | h=4 | h=5 |
|--------------------------------------|------|------|
| 0.75 | 13.3 | 17.0 |
| 1.00 | 8.38 | 10.4 |
| 1.50 | 4.75 | 5.75 |
| 2.00 | 3.34 | 4.01 |
| 2.50 | 2.62 | 3.11 |
| 3.00 | 2.19 | 2.57 |

Actualmente se aconseja analizar los datos CUSUM con la forma tabular (algorítmica).

Para el CUSUM superior (de fórmula $C_i^+ = \max \{0, X_i - (\mu_0 + K) + C_{i-1}^+ \}$) se calcula en primer lugar $x_i - (\mu_0 + K)$, es decir la diferencia entre el resultado de la medición del indicador (x_i) y el valor estándar u objetivo por mantener (μ_0 , que hemos establecido en 26.1 minutos) más K (que hemos establecido en $0.5 \hat{\sigma}$, es decir $(0.5) \cdot (6.8) = 3.4$). La primera columna del CUSUM superior contiene de esta forma los sucesivos resultados de calcular $x_i - (\mu_0 + K) = x_i - (26.1 + 3.4) = x_i - 29.5$. La siguiente columna es el valor C_i^+ , que será cero cuando la diferencia $x_i - (\mu_0 + K)$, es negativa, y será el valor de esta diferencia cuando sea positiva o la suma (acumulación) de todas las diferencias positivas que vayan apareciendo de forma sucesiva e ininterrumpida, de forma que cuando aparezca de nuevo una diferencia negativa se empieza de nuevo en cero. La siguiente y última columna del CUSUM superior (N^+) es un conteo de las veces sucesivas en que hay, de forma ininterrumpida, un CUSUM positivo; N^+ también se reinicia en cero, cuando vuelve a aparecer un CUSUM cero.

Para el CUSUM inferior, se calcula $(\mu_0 - K) - x_i$, que en nuestro caso es $(26.1 - 3.4) - x_i = 22.7 - x_i$. Cuando el resultado de este cálculo es negativo se procede igual que para el CUSUM superior, asignándole a C_i^- el valor cero, y, en cambio, el valor de esta diferencia cuando aparezca un valor positivo por primera vez, y se irá igualmente acumulando todos los valores positivos hasta que aparezca de nuevo un $C_i^- = 0$. La columna N_i^- es igualmente el conteo sucesivo de los C_i^- positivos, que se interrumpe al aparecer un $C_i^- = 0$.

Como puede apreciarse, tanto C_i^+ como C_i^- acumulan las desviaciones del indicador respecto al valor estándar u objetivo fijado que sean mayores que K ; en ambos casos, cuando estas desviaciones son negativas, se reinician en cero. Ahora, para interpretar si aparece o no una situación fuera de control, hay que observar si aparece algún valor de C^+ o C^- que sea mayor que H . Hemos apuntado más arriba que un valor de H frecuentemente recomendado es $4 \cdot \hat{\sigma}$ o $5 \cdot \hat{\sigma}$. Si elegimos $4 \cdot \hat{\sigma}$, el valor límite para C en nuestro ejemplo sería $(4) \cdot (6.8) = 27.2$. Viendo los valores de C^+ en el cuadro, no aparece ninguno por encima de ese valor. Sin embargo, en la columna de C^- , encontramos un valor > 27.2 en la medición núm. 28. El cuadro también nos dice en qué momento es probable que haya ocurrido el cambio, lo cual

correspondería al número de medición en la que se ha detectado, menos el N que corresponde a esa medición. Así, en nuestro caso, puesto que el cambio se ha detectado en la medición núm. 28, y el correspondiente N es 4, el cambio se produjo probablemente en algún momento tras la medición $28 - 4 = 24$. Si hubiéramos fijado el intervalo decisional en $H = 5 \cdot \hat{\sigma}$, el valor límite habría sido $(5) \cdot (6.8) = 34$. En este caso, el punto fuera de control sería el núm. 29, y la primera medición a partir de la cual ocurrió el cambio sería igualmente la 24. Obsérvese que estos cambios detectados podrían pasar desapercibidos en un gráfico de control normal (hacen falta seis puntos seguidos descendentes, y a la altura de la medición 28 sólo hay cinco) o, más probablemente, se habrían detectado en mediciones posteriores. Con cualquiera de los dos intervalos decisionales (los dos valores que hemos elegido de H), lo que nos encontramos es que el proceso ha disminuido significativamente su promedio.

Si queremos hacer el análisis CUSUM gráficamente, podemos utilizar una plantilla semejante a las de los gráficos de control, con valor promedio de cero y límites de control $0 \pm H$. En esta plantilla se grafican los valores de C^+ y C^- en ambos lados del promedio, y puede detectarse visualmente cuando aparece el punto significativo de descontrol. Este tipo de gráfico es el que contiene la figura 7.1, donde se representan los datos del cuadro 7.2. Si se considera oportuno, se añaden los valores que va tomando el indicador para visualizar su evolución.

5. Elaboración e interpretación de un gráfico EWMA (exponentially weighted moving-average, media móvil exponencialmente ponderada)

Los gráficos EWMA fueron introducidos por Roberts en 1959, tomando como base técnicas de proyección de series temporales desarrolladas en la Marina de Estados Unidos. De hecho, parece que se utiliza en la actualidad tanto para la gestión de calidad (dentro de la monitorización de indicadores, con el objetivo de detectar cambios relativamente pequeños) como para modelar y hacer proyecciones de series temporales. Para construir un gráfico EWMA, hay que decidir qué valores darles a λ y a L , lo cual se hace, como en

H y K (parámetros de diseño del análisis de los datos CUSUM) se eligen en función del ARL que se quiera tener, para detectar una determinada magnitud de cambio en el indicador.

El ARL (*average run length*) es el número medio de mediciones seguidas de un indicador que hay que hacer para detectar un determinado cambio en el promedio de ese indicador.

Para el análisis CUSUM en su forma tabular se calculan el CUSUM superior (C^+) y el CUSUM inferior (C^-). Estos índices expresan diferencias acumuladas por encima y por debajo del valor (estándar) de referencia.

La acumulación de diferencias se reinicia en cero (tanto para C^+ como para C^-) cuando el valor de C es cero.

En la tabla CUSUM se incluye también una columna con el núm. de veces sucesivas en las que C es positivo, en cualquiera de los sentidos (N^+ y N^-). Este conteo sirve para identificar entre qué mediciones se produjo el cambio detectado.

Para el análisis gráfico de los valores CUSUM se puede utilizar una plantilla semejante a la de los gráficos de control, con promedio cero y límites de control + H. Los valores C+ y C- se grafican en ambos lados del promedio.

| Cuadro 7.2 CUSUM tabular para el retraso en la administración de trombolíticos Análisis de 29 mediciones | | | | | | | |
|--|-------------------------|----------------|-----------------------------|----------------|----------------|-----------------------------|----------------|
| Medición núm. | Retraso en minutos (Xi) | CUSUM superior | | | CUSUM inferior | | N ⁻ |
| | | Xi-29.5 | C ⁺ _i | N ⁺ | 22.7-Xi | C ⁻ _i | |
| 1 | 30 | 0.5 | 0.5 | 1 | -7.3 | 0 | 0 |
| 2 | 12 | -17.5 | 0 | 0 | 10.7 | 10.7 | 1 |
| 3 | 29 | -0.5 | 0 | 0 | -6.3 | 0 | 0 |
| 4 | 26 | -3.5 | 0 | 0 | -3.3 | 0 | 0 |
| 5 | 40 | 10.5 | 10.5 | 1 | -17.3 | 0 | 0 |
| 6 | 33 | 3.5 | 14 | 2 | -10.3 | 0 | 0 |
| 7 | 25 | -4.5 | 0 | 0 | -2.3 | 0 | 0 |
| 8 | 23 | -6.5 | 0 | 0 | -0.3 | 0 | 0 |
| 9 | 31 | 1.5 | 1.5 | 1 | -8.3 | 0 | 0 |
| 10 | 26 | -3.5 | 0 | 0 | -3.3 | 0 | 0 |
| 11 | 33 | 3.5 | 3.5 | 1 | -10.3 | 0 | 0 |
| 12 | 25 | -4.5 | 0 | 0 | -2.3 | 0 | 0 |
| 13 | 20 | -9.5 | 0 | 0 | 2.7 | 5.0 | 1 |
| 14 | 20 | -9.5 | 0 | 0 | 2.7 | 7.7 | 2 |
| 15 | 30 | 0.5 | 0.5 | 1 | -7.3 | 0 | 0 |
| 16 | 26 | -3.5 | 0 | 0 | -3.3 | 0 | 0 |
| 17 | 31 | 1.5 | 1.5 | 1 | -8.3 | 0 | 0 |
| 18 | 25 | -4.5 | 0 | 0 | -2.3 | 0 | 0 |
| 19 | 26 | -3.5 | 0 | 0 | -3.3 | 0 | 0 |
| 20 | 28 | -1.5 | 0 | 0 | -5.3 | 0 | 0 |
| 21 | 10 | -19.5 | 0 | 0 | 12.7 | 12.7 | 1 |
| 22 | 27 | -2.5 | 0 | 0 | -4.3 | 0 | 0 |
| 23 | 28 | -1.5 | 0 | 0 | -5.3 | 0 | 0 |
| 24 | 30 | 0.5 | 0.5 | 1 | -7.3 | 0 | 0 |
| 25 | 19 | -10.5 | 0 | 0 | 3.7 | 3.7 | 1 |
| 26 | 15 | -14.5 | 0 | 0 | 7.7 | 11.4 | 2 |
| 27 | 14 | -15.5 | 0 | 0 | 8.7 | 20.1 | 3 |
| 28 | 12 | -17.5 | 0 | 0 | 10.7 | 30.8 | 4 |
| 29 | 19 | -10.5 | 0 | 0 | 3.7 | 34.5 | 5 |

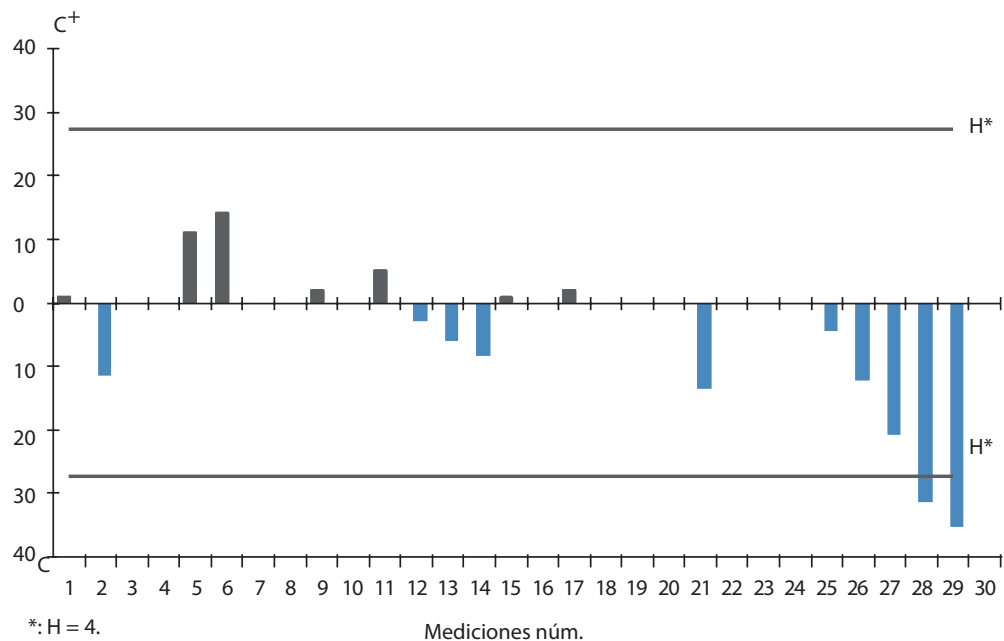


Figura 7.1. Gráfico CUSUM. Datos del cuadro 7.2

el CUSUM, con base en la experiencia de acuerdo con el ARL que nos parezca aceptable. En general se manejan valores de λ entre 0.05 y 0.25 (eligiendo un λ tanto más pequeño, cuanto más pequeños sean los cambios que se quieran detectar) y una L que puede ser de $3\hat{\sigma}$, igual que en los gráficos de control tradicionales. Sin embargo, se aconseja reducir L a 2.8 o 2.6 en los valores de l más pequeños. El cuadro 7.3 expresa diversos valores de ARL para distintas combinaciones de λ y L y cambios en el promedio que queremos detectar.

Elegiremos para nuestro ejemplo $\lambda=0.2$ y $L=3$, en combinación con un ARL semejante (cuadro 7.1) al proceder CUSUM (con $k=0.5$ y $H=4.s$) que hemos utilizado antes. Construiremos una tabla de datos EWMA y su correspondiente gráfico, utilizando nuevamente los datos del retraso en la administración de trombolíticos del CUSUM. Estableceremos también en este caso como valor de referencia (objetivo estándar del indicador) el promedio en que conseguimos estabilizarlo, que cumple las expectativas de calidad en relación con ese indicador. En nuestro ejemplo, ese valor para \hat{X}_0 es 26.1 minutos, y $\hat{\sigma}$ es 6.8.

Teniendo en cuenta estos valores, construiremos el cuadro 7.4, con los resultados de la fórmula para \hat{X}_i que vimos en el apartado anterior: $\hat{X}_i = lX_i + (1-l)X_{i-1}$ que sería $\hat{X}_i = (0.2)X_i + (1-0.2) X_{i-1}$.

El gráfico EWMA resultante es el de la figura 7.2, en la cual puede verse que el proceso no es estable, ya que aparecen tres puntos fuera de control (mediciones núm. 3, 6 y 22) no advertidas en el CUSUM, además de la tendencia significativa descubierta al final de la serie que también detectó el CUSUM. Todos estos cambios podrían haber

pasado desapercibidos en los gráficos de control tradicionales.

6. Otros métodos gráficos de análisis y monitorización estadística de la calidad

Tanto los análisis CUSUM como EWMA pueden adaptarse y utilizarse, prácticamente, para todo tipo de indicadores; sin embargo, las técnicas y herramientas de control estadístico de procesos no se agotan con estos dos tipos de gráficos, además de que éste es un campo en continua evolución. De entre los métodos de control estadístico no vistos hasta aquí, merece la pena mencionar brevemente las características y objetivos de los llamados análisis precontrol y de los gráficos para monitorizar, con control estadístico, ciertas características de la capacidad del proceso.

6.1 Precontrol

En un proceso que se supone que funciona razonablemente bien, puede establecerse un sistema de control exclusivamente para comprobar periódicamente si el producto cumple o no las especificaciones. Entonces, para la monitorización, se establece lo que se conoce como límites estrechos de conformidad (narrow limits gages) que se sitúan en ambos lados a un cuarto de la diferencia entre los límites de especificación. De esta forma no hay cálculo estadístico de ninguna clase, y lo único que se hace es comprobar si el producto está dentro de los límites establecidos (que son más exigentes que los de las especificaciones). Con dos mediciones que caigan fuera de los límites estrechos, podrá concluirse (con base en cálculo

Los gráficos EWMA se basan en técnicas de proyección de series temporales.

Los parámetros de diseño del gráfico EWMA (l y L) se deciden en función del ARL que queremos tener para detectar una determinada magnitud de cambio en el indicador analizado.

Los gráficos CUSUM y EWMA tienen su aplicación más genuina para indicadores de valores absolutos ($n=1$), pero pueden adaptarse para otro tipo de indicadores.

Hay muchos otros tipos de gráficos de control estadístico de procesos; destacan aquellos que tienen en cuenta las especificaciones de calidad: análisis precontrol y gráficos para monitorizar la capacidad del proceso.

Cuadro 7.3

ARL del gráfico EWMA para diversos valores de λ y L

| Cambio del promedio (múltiplos de s) | L=3.054 $\lambda=0.40$ | 2.998 0.25 | 2.962 0.20 | 2.814 0.10 | 2.615 0.05 |
|--------------------------------------|---------------------------|---------------|---------------|---------------|---------------|
| 0.75 | 28.4 | 20.1 | 18.2 | 15.9 | 16.4 |
| 1.00 | 14.3 | 11.1 | 10.5 | 10.3 | 11.4 |
| 1.50 | 5.9 | 5.5 | 5.5 | 6.1 | 7.1 |
| 2.00 | 3.5 | 3.6 | 3.7 | 4.4 | 5.2 |
| 2.50 | 2.5 | 2.7 | 2.9 | 3.4 | 4.2 |
| 3.00 | 2.0 | 2.3 | 2.4 | 2.9 | 3.5 |

| Cuadro 7.4 Valores EWMA para el retraso en la administración de trombolíticos (29 mediciones) | | | | |
|--|-------------------------|----------------------|-------|-------|
| Medición núm. | Retraso en minutos (Xi) | EWMA (\hat{X}_i) | LCS* | LCI* |
| 1 | 30 | 26.88 | 30.07 | 22.13 |
| 2 | 12 | 26.40 | 31.19 | 21.01 |
| 3 | 29 | 15.40 | 31.79 | 20.41 |
| 4 | 26 | 28.40 | 32.14 | 20.06 |
| 5 | 40 | 28.80 | 32.35 | 19.85 |
| 6 | 33 | 38.60 | 32.49 | 19.71 |
| 7 | 25 | 31.40 | 32.57 | 19.63 |
| 8 | 23 | 24.60 | 32.62 | 19.58 |
| 9 | 31 | 24.60 | 32.66 | 19.54 |
| 10 | 26 | 30.00 | 32.68 | 19.52 |
| 11 | 33 | 27.40 | 32.69 | 19.51 |
| 12 | 25 | 31.40 | 32.70 | 19.50 |
| 13 | 20 | 24.00 | 32.71 | 19.49 |
| 14 | 20 | 20.00 | 32.71 | 19.49 |
| 15 | 30 | 22.00 | 32.71 | 19.49 |
| 16 | 26 | 29.20 | 32.72 | 19.48 |
| 17 | 31 | 27.00 | 32.72 | 19.48 |
| 18 | 25 | 29.80 | 32.72 | 19.48 |
| 19 | 26 | 25.20 | 32.72 | 19.48 |
| 20 | 28 | 26.40 | 32.72 | 19.48 |
| 21 | 10 | 24.40 | 32.72 | 19.48 |
| 22 | 27 | 13.40 | 32.72 | 19.48 |
| 23 | 28 | 27.20 | 32.72 | 19.48 |
| 24 | 30 | 28.40 | 32.72 | 19.48 |
| 25 | 19 | 27.80 | 32.72 | 19.48 |
| 26 | 15 | 18.20 | 32.72 | 19.48 |
| 27 | 14 | 14.80 | 32.72 | 19.48 |
| 28 | 12 | 13.60 | 32.72 | 19.48 |
| 29 | 19 | 13.40 | 32.72 | 19.48 |

*: Límite de control superior y límite de control inferior, para $l=0.2$ y $L=3$; $m_0=26.1$.

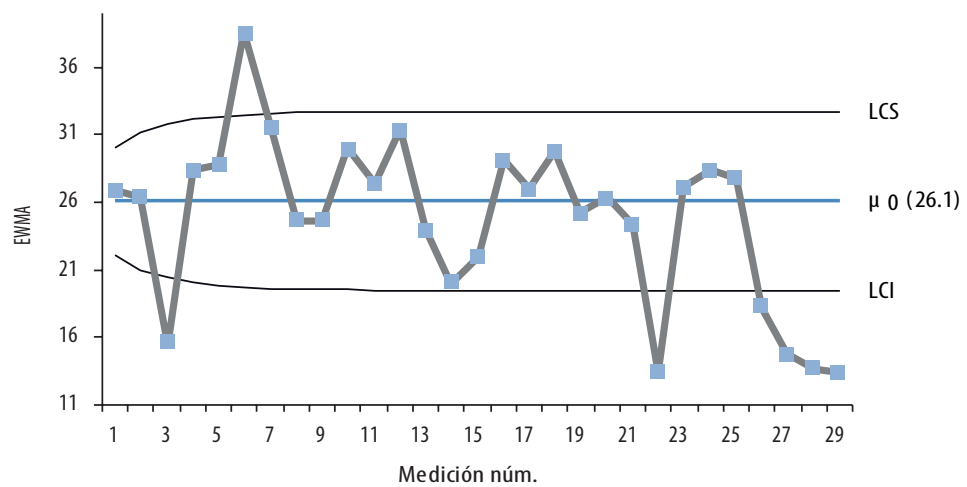


Figura 7.2. Gráfico EWMA. Datos del cuadro 7.4

los probabilísticos establecidos) que el proceso ya no funciona tan bien como era de esperarse.

6.2. Gráficos para monitorizar la estabilidad de la capacidad del proceso

Es posible construir y utilizar gráficos para controlar la fracción de producción defectuosa; dos de los cuales están incluidos en los estándares de la ASQC (American Society for Quality Control): gráfico de control con límites modificados y gráficos de control de la aceptación.

La modificación a que hace referencia el nombre de gráficos de control con límites modificados es la que resulta de tener en cuenta los límites de especificación, que se utilizan para calcular límites de control superior e inferior y líneas *centrales* superior e inferior, en torno a las cuales se espera que se distribuyan las mediciones del indicador.

Los gráficos de control de la aceptación están diseñados para monitorizar tanto la conformidad con la tolerancia o límites de especificación como

la estabilidad estadística del indicador. Para ello se establecen dos niveles, uno de aceptación y otro de rechazo, teniendo en cuenta los riesgos α y β de forma semejante a como se hace en el LQAS. El nivel de aceptación es el promedio de la proporción aceptable de unidades no conformes, defectuosas o fuera de los límites de especificación. El nivel de rechazo es el promedio de lo que se define como proporción inaceptable de unidades defectuosas. En ambos, se establecen límites de control.

Como puede verse, las técnicas y herramientas de control estadístico de la calidad son muy variadas y sus aplicaciones casi infinitas o, como mínimo, de lo más versátil y polivalente para toda la gama de actividades a realizar dentro de los programas de gestión de calidad. El control estadístico de procesos es un campo fascinante, en continuo perfeccionamiento y tremendamente útil, y no solamente, aunque sí de manera muy destacada, en los sistemas de monitorización.

El análisis de precontrol consiste en comprobar si el producto cumple o no con las especificaciones de calidad dentro de los llamados límites estrechos, que dejan fuera 50% (1/4 a cada lado) de la tolerancia. Es una técnica útil para monitorizar procesos bien diseñados y fiables.

Los gráficos de control con límites modificados utilizan dos líneas *centrales* (superior e inferior) que se establecen tomando como referencia los límites de las especificaciones.

Los gráficos de control de la aceptación controlan la fracción de producción defectuosa, estableciendo un nivel de aceptación y otro de rechazo de forma parecida al LQAS.

Las técnicas y herramientas de control estadístico de la calidad son muy versátiles y aplicables en prácticamente toda la gama de actividades de los programas de gestión de la calidad.

BIBLIOGRAFÍA

- Farnum NR. Capítulo 7. Modern statistical quality control and improvement. Belmont: Duxbury Press, 1994.
- Montgomery DC. Capítulo 7. Introduction to statistical quality control. 2ª ed. Nueva York: John Wiley & Sons, 1996.
- Wadsworth HM, Stephens KS y Godfrey AB. Capítulo 8. Modern methods for quality control and improvement. Nueva York: John Wiley & Sons, 1986.

Métodos y herramientas
para la **monitorización**
de la **calidad**
en **servicios de salud**

Se terminó de imprimir en octubre de 2015.
La edición consta de 1 000 ejemplares
y estuvo al cuidado de la
Subdirección de Comunicación Científica y Publicaciones
del Instituto Nacional de Salud Pública

