

Análisis de R&R por atributos que valora cuantitativamente el desempeño de un sistema de inspección visual

Repeatability and reproducibility (R&R) analysis by attributes which values quantitatively the performance of a visual inspection system

Luisa Carolina Jiménez Jiménez¹ Carmen Elena Patiño Rodríguez^{1*}
Fernando Jesús Guevara Carazas²

Recibido 12 de febrero de 2021, aceptado 05 de mayo de 2022

Received: February 12, 2021 Accepted: May 05, 2022

RESUMEN

Este trabajo presenta el desarrollo de un sistema de valoración cuantitativa de la capacidad de los inspectores para tomar decisiones acertadas, usando el método de análisis de riesgos y el índice Kappa de Cohen. En este artículo se establece un procedimiento para la aplicación del método de análisis de riesgos valorando cuantitativamente el desempeño de inspección visual. Para medir el impacto de diferentes escenarios con variaciones en el nivel de capacitación, experiencia, unificación de criterios, características físicas, métodos y estándares disponibles, se realiza un estudio de simulación, para diferentes inspectores en tres escenarios. Los resultados permitieron obtener niveles de calidad del proceso entre 95% y 99%. Es posible evidenciar que reducir la variabilidad del resultado de los procesos de inspección mediante la estandarización de métodos, unificación de criterios y capacitación del personal mejora el nivel de calidad. El procedimiento presentado y simulado en este artículo puede ser replicable en procesos de inspección cualitativos ya sea procesos de manufactura o de servicios.

Palabras clave: Atributos, repetibilidad y reproducibilidad (R&R), índice Kappa, análisis de riesgos.

ABSTRACT

The present article proposes a system of a quantitative assessment of the ability of inspectors to make correct decisions, applying an R&R study by attributes, with the method of risk analysis, supported by contingency tables and the Cohen Kappa index. Additionally, it establishes a tool to simplify the required calculations and a proposed inspection method that seeks to reduce the variability associated with the inspection process. In this paper, a simulation study is made. We simulate several quality inspectors and three sceneries. Results show a quality level between 95% and 99% was obtained. It is possible to show that reducing the variability of inspection processes results through standardized methods, criteria equalization, and staff training improves the quality level process. The procedure presented and simulated in this article can be applied in the qualitative inspection, such as manufacturing processes as services processes.

Keywords: Attributes, repeatability, reproducibility, R&R, Kappa index, risk analysis.

¹ Universidad de Antioquia. Departamento de Ingeniería Industrial. Grupo de Investigación en Analítica e Investigación para la Toma de Decisiones - ALIADO. Medellín, Colombia. lcarolina.jimenez@udea.edu.co, elena.patino@udea.edu.co

² Universidad Nacional de Colombia. Departamento de Ingeniería Mecánica. Grupo de Investigación en Gestión, Operación y Mantenimiento de Activos - GOMAC. Medellín, Colombia. fjguevarac@unal.edu.co

* Autor de correspondencia: elena.patino@udea.edu.co

INTRODUCCIÓN

En diferentes procesos de producción, la apariencia es una de las principales características de calidad y, en muchas ocasiones, un determinante de compra. Sin embargo, interpretar atributos no es una tarea simple, y menos aún, garantizar que los productos sean fabricados y evaluados con los mismos criterios, cuando están asociados a estas características cualitativas. Por lo tanto, es muy importante que el personal operativo pueda identificar las desviaciones de conformidad que el cliente no tolera, así como las desviaciones que puede pasar por alto [1], por ejemplo, no-conformidades que no afectan la funcionalidad y que no están en la línea de visión del cliente.

Desde el punto de vista de Hoyer and Hoyer, una de las definiciones más clara y completa, para el concepto de calidad, ha sido establecida por Shewhart. Algunos de sus planteamientos principales son [2]:

- “Existen dos características de calidad una subjetiva (lo que el cliente quiere) y la otra objetiva (propiedades del producto que no depende de lo que el cliente quiere)”.
- “Los estándares de calidad, de los productos, deben ser expresados en términos físicos y características cuantitativamente medibles” .
- “La estadística debe ser usada para tomar información sobre el gran potencial que tienen muchos productos y servicios y traducirla en características medibles de un producto específico que satisfaga al mercado”.

En este sentido, la calidad es multidimensional y la apariencia una de las dimensiones más relevantes.

La inspección visual es comúnmente utilizada en la industria para evaluar no-conformidades en un producto y se mide normalmente como un atributo [3]. Los sistemas de inspección tienen un riesgo asociado a rechazar un producto que está en un nivel de calidad aceptable (error Tipo I) y otro riesgo a aceptar un producto que no tiene el nivel de calidad aceptable (error tipo II). En el caso de la evaluación por atributos, la probabilidad de incurrir en cualquiera de estos riesgos aumenta, debido a que la decisión de aceptar o rechazar un producto, se fundamenta en la experiencia, las sensaciones, y los criterios del inspector, en el momento de hacer la evaluación. Para disminuir la subjetividad y por ende el riesgo que existe al realizar la evaluación por atributos es necesario unificar el criterio asociado a la aceptación y rechazo de los productos [4]. Cuando se trata de verificación de atributos y evaluación de no-conformidades, la calibración de criterios y estandarización de métodos se torna fundamental para evitar la adquisición de datos de baja calidad, por la variación debida a la interacción entre el sistema de medición y el ambiente [3, 4]. En este escenario, se requiere de un mecanismo que valore cuantitativamente la aptitud de cada inspector, para garantizar una toma de decisiones acorde a la calidad real que la empresa entrega a sus clientes. En muchas ocasiones se definen índices e indicadores del nivel de concordancia entre las salidas. Para calcular esto índices es necesario realizar estudios R&R para atributos. Al realizar una búsqueda bibliográfica en la base de datos Scopus, se establecieron los tres criterios presentados en la Tabla 1.

Al aplicar conjuntamente los criterios 1 y 2, a la ecuación de búsqueda, se obtuvieron 1774 resultados. Sobre estos 1774 resultados fue aplicado el criterio

Tabla 1. Criterios para búsqueda bibliográfica.

Criterio		Ecuación de búsqueda
1	Palabras claves contendrán términos asociados con los métodos y mediciones.	((KEY ((repeatability AND reproducibility) OR (reproducibilities OR gage OR gauge OR interrater OR capability OR measurement OR measure))).
2	Resumen mencionará el tipo de datos y el tipo de estudio.	(ABS (attribute OR categorical OR discrete OR nominal) AND (gage OR gauge)).
3	Palabras Claves exactas.	“Attribute Data” OR “Attribute Agreement” OR “Attribute gage” OR “Attribute capability” OR “Test Retest Reliability” OR “Categorical Data” OR “Gauge Repeatability” OR “Categorical Attributes” OR “Gauge Capability” OR “Interrater Reliability” OR “Kappa Coefficient” OR “Nominal Data” OR “Reproducibility” OR “Reproducibilities”.

3 para seleccionar sólo artículos que cumplieran con las palabras claves exactas tal como se muestran en la Tabla 1. De esta forma, se obtuvieron 92 documentos.

Al analizar las publicaciones en el horizonte de tiempo, se puede observar que a partir del año 2005 empiezan a crecer el número de publicaciones en torno a este tema, llegando a su máximo en el año 2019. Antes del año 2005, las publicaciones corresponden máximo a una publicación por año y se observa que las primeras publicaciones que involucran el tema de estudios R&R para atributos se realizan en la década de los noventa. Se puede verificar que en el caso de variables discretas lo reportado en la bibliografía no supera el 3% del total de resultados. Lo anterior es un indicador de la dificultad que supone adquirir la información de referencia del verdadero estado de las piezas valoradas, así como obtener procedimientos que sirvan para evaluar la concordancia entre evaluadores. La referencia del verdadero, en el caso de variables continuas, es suministrada por instrumentos de medición calibrados y la concordancia de los resultados es verificada a partir de medidas estadísticas que se sistematizan a partir de estudios R&R para variables continuas [5, 6, 7, 8].

Al analizar los resultados de la búsqueda bibliográfica, en primera instancia se observa que entre las principales áreas de conocimiento que han desarrollado estudios R&R para atributos (ver Figura 1), se encuentran las áreas de la salud e ingeniería. Adicionalmente se puede verificar que, en los cinco últimos años, las investigaciones sobre análisis cuantitativos R&R de datos con variables discretas (atributos) han ido emergiendo en sectores asociados al desarrollo sistemas automatizados que evalúan computacionalmente el grado de concordancia de las decisiones tomadas como lo son las áreas de ciencias de la computación, matemáticas y ciencias de la decisión [9, 10, 11, 12].

Adicionalmente para caracterizar el estado del arte sobre estudios R&R para atributos se realizó un análisis bibliométrico, con el software VOSviewer [13] (ver Figura 2). En este caso se presenta la red asociada con la coexistencia de las palabras claves, lo cual permite identificar los temas de investigación y sus interacciones más importantes, así como, su evolución en el tiempo.

Los temas tratados antes del 2009, estuvieron entorno a analizar la repetibilidad en atributos [1, 3, 14, 15, 16, 17] (ver Figura 2). Con el paso de los años, los

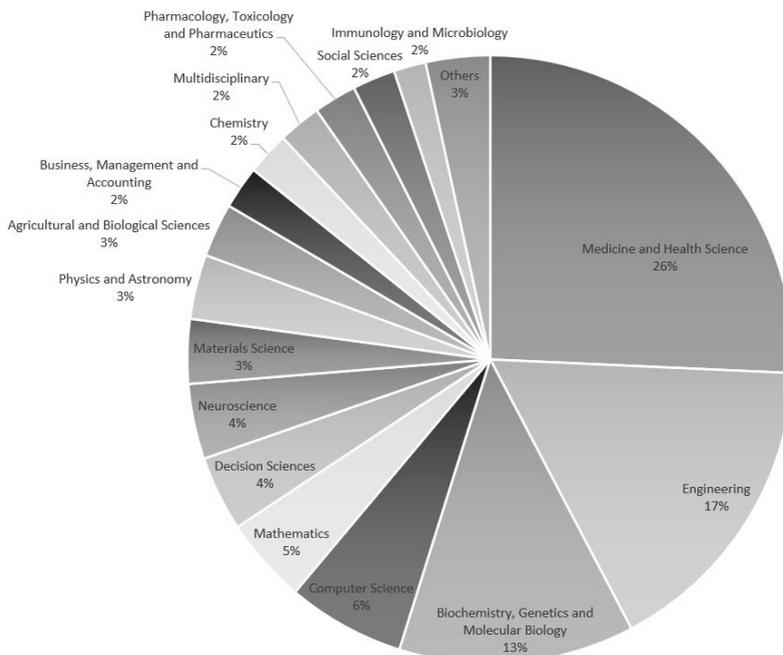


Figura 1. Resultados obtenidos en revisión de bibliográfica.

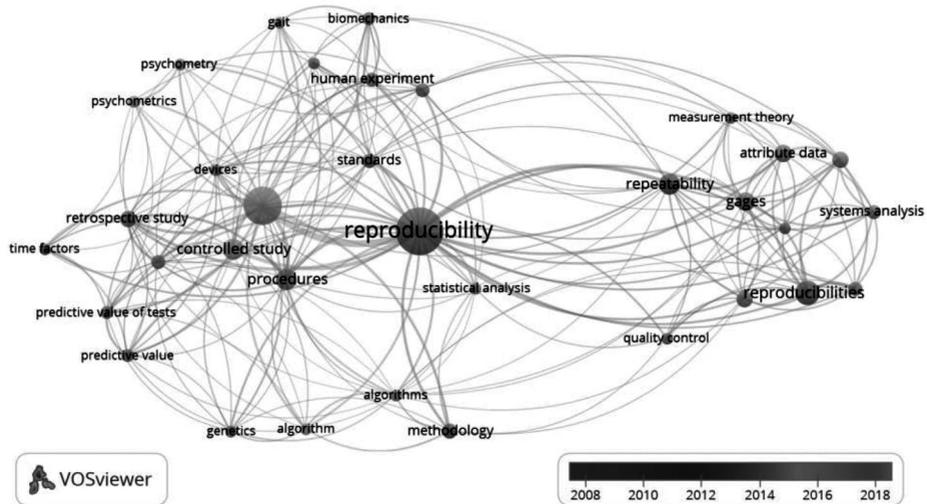


Figura 2. Análisis bibliométrico.

análisis se han ido enfocando hacia el análisis de la reproducibilidad para atributos [18, 19, 20, 21, 22, 23, 24, 25, 26, 27]. Cabe recordar que la repetibilidad busca medir la capacidad del instrumento de obtener medidas con muy poca variabilidad respecto al valor objetivo, es decir en el caso de atributos se asocia con la identificación de la no-conformidad; y la reproducibilidad está directamente asociada con el desempeño del observador, la capacidad que tiene el operario dar un resultado de forma consistente, es decir el resultado puede ser rechazo o aceptación, pero debe ser el mismo.

Además, en la evolución de los temas relacionados con R&R para atributos se identifican dos enfoques. El primer enfoque se refiere al uso métodos de analítica en R&R con ayuda de la automatización y utilización de sensores, así como con el procesamiento de imágenes [7, 9, 10, 11, 12, 25]. En el segundo se encuentran los temas relacionados con el análisis de riesgo que permiten que los estudios R&R para atributos hagan parte integral de métodos de Lean Manufacturing, como Six Sigma [4, 16, 23, 24]. Los procesos, en los cuales se aplica el segundo enfoque, requieren estudios R&R para atributos que consideran escalas categóricas y estén centrados sobre procedimientos y, que, a su vez, evalúen la capacidad de tomar decisiones consistentes a la hora de rechazar o aceptar un producto acertadamente.

El rápido avance de la tecnología dejó una brecha de conocimiento para las industrias manufacturas y

procesos que no cuentan con tecnologías de última generación o que por su naturaleza no producen la cantidad y la variedad de datos necesarios para aplicar modelos que requieren altos volúmenes de datos. De esta forma, este artículo se enmarca en el segundo enfoque, y es una herramienta útil para estos procesos, ya que permite evaluar las decisiones acerca de la calidad de sus productos desde un procedimiento sistemático y estructurado que incluye el índice Kappa de Cohen, una métrica que permite evaluar cuantitativamente la concordancia de las decisiones.

El índice de Kappa fue propuesto por Cohen[28] como una medida de concordancia que compara la concordancia observada entre la evaluación de dos observadores cuando ambos están calificando el mismo objeto, respecto a la que podría ocurrir por azar (concordancia esperada) [19]. Diferentes autores coinciden que cuando la calidad de un producto es determinada a partir de apreciaciones de los observadores es necesario medir el grado de concordancia de los observadores [1, 16, 20]. Para el caso de más de dos evaluadores, Fleiss generalizó el método de Cohen, conocido también como Kappa de Fleiss [22]. Inicialmente estos métodos fueron desarrollados para la medicina, psicología y otras áreas de la salud. En las últimas dos décadas, después de la publicación realizada por AIAG [3] han aumentado las publicaciones referentes a aplicaciones en áreas industriales [3, 4, 16, 21]. De Mast y Marques mencionan que algunos softwares

comerciales tienen módulos incluidos para analizar el porcentaje de acuerdo entre la apreciación del observador y el estándar, usando el índice Kappa [15, 16, 29].

En este mismo sentido, Altobelli, describen la aplicación del índice Kappa de Cohen buscando responder: 1) ¿Cuál será el nivel de concordancia existente entre los operadores? y, 2) ¿Si se cambia de un clasificador a otro, hasta qué punto los clasificadores coincidirán en su evaluación y si en este caso, el resultado será real o debido al azar? [18]. Bajo el mismo enfoque se encuentra en la literatura los modelos probabilísticos que a través de tablas de contingencia y la teoría asociada con análisis de datos categóricos determinan el grado de concordancia entre observadores y el estándar [17, 20].

Según la Sociedad Andaluza de Medicina Intensiva, Urgencias y Coronarias el índice Kappa se usa con frecuencia en la literatura médica, tanto por su relativa facilidad de cálculo como porque su interpretación es relativamente clara. Sin embargo, solo se debe aplicar en clasificaciones nominales, no está pensada para clasificaciones en que algunas categorías son de distinta importancia que otras, por ejemplo: Leve, moderado, grave, muy grave, extremadamente grave [26].

Con base en todo esto, este artículo presenta el diseño de una metodología para la aplicación de un análisis de R&R por atributos que valore cuantitativamente el desempeño de un sistema de inspección visual, este estará enfocado a la medición de la dimensión de la calidad asociada a la estética de las piezas que generalmente es tratada a través de variables categóricas (discretas).

El artículo está organizado de la siguiente manera: se presenta el desarrollo de la metodología, donde se explica brevemente los conceptos asociados a la aplicación de un estudio R&R por atributos usando el método de análisis de riesgos; luego en Aplicación de la metodología, se construye la herramienta que convierte en un proceso sistemático la aplicación del estudio R&R por atributos en procesos productivos. A continuación, se analizan los resultados obtenidos luego de la puesta en marcha de la herramienta, verificando la aplicabilidad de esta y evaluando la unificación de los criterios de inspección visual. Finalmente, se exponen las conclusiones más relevantes del trabajo.

METODOLOGÍA

Según de Mast [16, 20, 21] los estudios de R&R por atributos pueden mejorar el rendimiento de los procesos y reducir los costos de no calidad. La repetibilidad de un sistema de medición hace referencia a la variabilidad de sus mediciones cuando se obtienen varias mediciones del mismo objeto en condiciones similares (mismo operador); mientras que la reproducibilidad es la precisión o variabilidad de las mediciones del mismo objeto, pero con diferentes operadores [17]. Los métodos habituales para estudios R&R por atributos se pueden clasificar en: Análisis de riesgos, Teoría de Señales, Método Analítico y Método Bayesiano. Existen diferentes métodos para la aplicación de estudios R&R por atributos, no obstante, la aplicabilidad de uno u otro depende de la información de referencia de la cual se disponga (ver Tabla 2). En este artículo se aplicará el método de riesgos.

El riesgo de tomar una decisión incorrecta puede ser evaluado usando el método de análisis de riesgo, que se analiza cuando el resultado de la prueba de calidad es un atributo y no es factible obtener suficientes partes con valores de referencia. El método de riesgo está basado en el análisis de hipótesis y tablas de contingencia. Es importante tener en cuenta que el método no cuantifica la variabilidad del sistema de medición y debe usarse únicamente con el consentimiento del cliente. La selección y uso de esta técnica deberá basarse en una buena práctica estadística, y en un entendimiento de las fuentes potenciales de variación que pueden afectar el producto o la medición de procesos, y el efecto de una decisión incorrecta en el cliente final [15, 16].

Metodología para aplicar el método de análisis de riesgos

La Automotive Industry Action Group -AIAG expone el procedimiento de aplicación del método de análisis de riesgos que incluye cinco pasos: 1) Método de Inspección, 2) Clasificación de piezas, 3) Resumen de los datos, 4) Distribución de los datos y 5) Determinación del índice Kappa y análisis de los resultados [3].

Es importante resaltar que en esta metodología resulta primordial desarrollar un método estandarizado para realizar la inspección. Este procedimiento representa la información de entrada.

Tabla 2. Métodos para la aplicación de estudios R&R por atributos.

Método	Características
Análisis de riesgos	Se debe contar con la referencia del verdadero estado de las piezas a evaluar. m operadores evalúan n piezas r veces, clasificándolas como aceptable o inaceptable. Se valora la concordancia entre operadores con el estándar. El estadístico usado es Índice Kappa, proporción de acuerdo corregido, retirando el acuerdo debidos al azar. Nota: En algunos casos se usa como estadístico el porcentaje de acuerdo, sin ninguna corrección del acuerdo dado por el azar.
Teoría de señales	Se requiere contar con una referencia en escala continua para cada pieza y un código para establecer el estado de las piezas respecto a la referencia. Se hace un conteo de las piezas que se ubican dentro de las especificaciones. El objetivo del análisis es estimar el ancho promedio de las zonas de incertidumbre que aparecen alrededor de cada especificación.
Método analítico	Se requiere contar con una referencia en escala continua. n piezas son evaluadas r veces y se registra el número de veces que se acepta cada una de ellas respecto a la especificación. El objetivo es valorar el desempeño del sistema a partir de los valores de probabilidad acumulada de la distribución normal.
Método bayesiano	Procedimiento probabilístico capaz de clasificar un inspector en calificado o no calificado para situaciones en las que exista o no información sobre los errores de clasificación. El esquema propuesto utiliza una distribución de probabilidad de los errores de clasificación que evita su especificación exacta, además de incorporar una variación entre inspectores.

Método de inspección

El método de un proceso de inspección visual en tres etapas: 1) Exploración, 2) Evaluación y 3) Decisión [15] (ver Figura 3). En la primera etapa el inspector explora la superficie del producto para verificar la presencia de desviaciones y el tipo de anomalía. En la exploración se pueden aplicar la exploración aleatoria o la exploración sistemática. Es importante resaltar que la exploración sistemática es más eficiente ya que el tiempo de inspección es menor.

Una vez identificada la no-conformidad, se pasa a la etapa de evaluación, donde se establece la intensidad de la no-conformidad. En esta etapa se usa un perfil sensorial para evaluar cada uno de los atributos de la anomalía detectada, teniendo en cuenta el tipo y los atributos de la anomalía, así como los estándares definidos por la compañía [26].

Finalmente, está la etapa de decisión, donde el inspector establece la intensidad global de la anomalía basada en la correlación entre la evaluación de los atributos y la decisión de expertos. Esta decisión se establece en función de los límites que se han fijado en la compañía, los cuales pueden variar según la gama del producto o el cliente [17].

Método de análisis de riesgos

El resultado obtenido en el método de inspección será base para la aplicación del método de análisis de riesgos (ver Figura 4). En el desarrollo de este método se pueden identificar tres pasos: 1) Ingreso de los datos, 2) Análisis de datos, y 3) Evaluación de resultados. En el primer paso, cada uno de los inspectores, toma dos decisiones en cada parte observada. Una decisión aceptable (Pasa) es designada como uno (1) y una decisión inaceptable (Nº Pasa) con cero (0). La decisión es tomada teniendo como base el método de inspección, con el objetivo de reducir la variabilidad asociada a las condiciones de inspección. En el segundo paso se resumen los datos observados y se estima la distribución de datos esperados.

Para lograr el objetivo en el segundo paso, se evalúan los datos por par de observadores (Experto e Inspector), contando el número de veces que coincidan en la decisión de aceptar o rechazar cada una de las piezas en las diferentes observaciones (ver Tabla 3).

La probabilidad de que un par de observadores estén de acuerdo o en desacuerdo en una observación únicamente por el azar, es una proporción que resume el número de acuerdos y desacuerdos entre

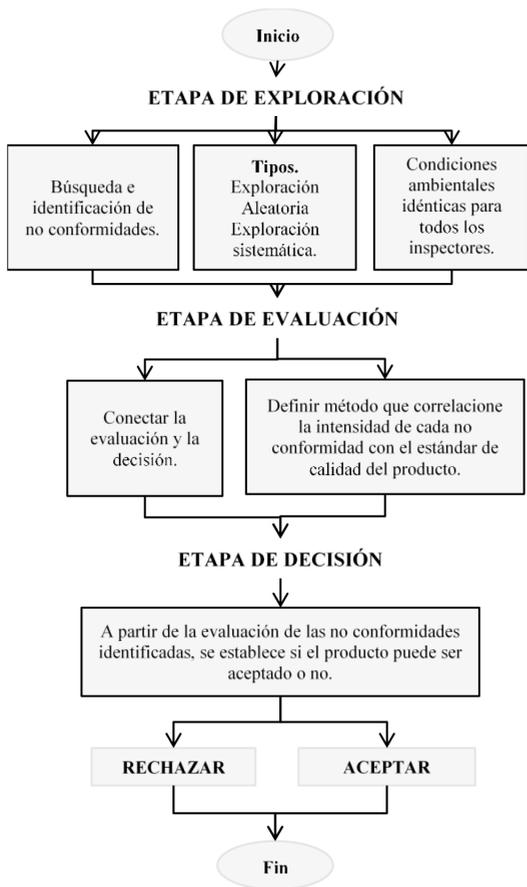


Figura 3. Etapas proceso de inspección.

el inspector y el estándar. La probabilidad observada (P_0) y esperada (P_e), se calculan como se muestra en la ecuación (1) y ecuación (2).

$$P_0 = \frac{\text{Observaciones}_{0,0} + \text{Observaciones}_{1,1}}{n} \quad (1)$$

$$P_0 = \frac{\# \text{acuerdos No conformes} + \# \text{acuerdos Conformes}}{\text{Total piezas}} \quad (2)$$

$$P_e = \frac{E[\text{acuerdos No conformes}] + E[\text{acuerdos Conformes}]}{\text{Total piezas}}$$

El número esperado de veces en que el Observador A y el Observador B estén de acuerdo en rechazar una pieza es estimado por la multiplicación de la probabilidad combinada por el número de observaciones, como se muestra en la Tabla 3.

El tercer paso, consiste en determinar el índice Kappa y analizar resultados. Está claro que una

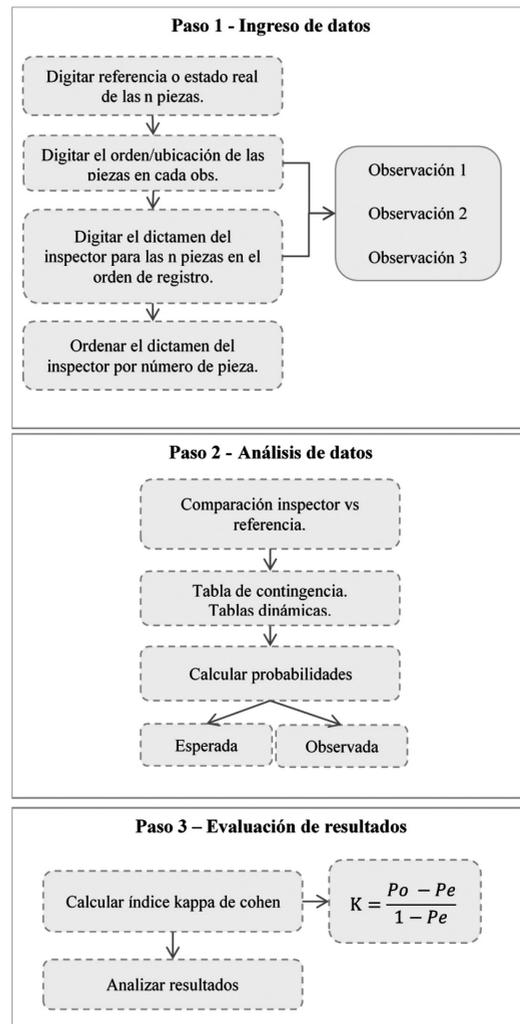


Figura 4. Descripción metodológica de la herramienta.

medida simple de concordancia sería la proporción de coincidencias frente al total de sujetos [16]. Esta etapa se apoya en el Valor observado y el Valor esperado estimado en el paso previo.

Aun cuando, no hubiera ninguna relación entre los dos observadores, o entre las dos escalas de evaluación, podría haber algún grado de coincidencia por azar. En estudios R&R se quiere eliminar de alguna forma la concordancia esperada por azar.

Para eliminar el azar, es necesario tener en cuenta que, si se denomina P_0 a la proporción de la concordancia observada y P_e , a la proporción de

Tabla 3. Tabla de contingencia y cálculo de valores esperados.

		Observador B		Total
		0	1	
Observador A	0	(Acuerdos) $\frac{\text{Total No conformes B} * \text{Total no conformes A}}{\text{Total piezas}}$	(Desacuerdos) $\frac{\text{Total conformes B} * \text{Total No conformes A}}{\text{Total piezas}}$	No conformes de A
	1	(Desacuerdos) $\frac{\text{Total No conformes B} * \text{Total conformes A}}{\text{Total piezas}}$	(Acuerdos) $\frac{\text{Total conformes B} * \text{Total conformes A}}{\text{Total piezas}}$	Conformes de A
Total		No conformes de B	Conformes de B	Piezas evaluadas

concordancia que se esperaría por azar, el índice Kappa (k) estaría dado por la cantidad en que el acuerdo observado excede lo esperado por azar, dividido el valor máximo de esta diferencia [30], como se muestra en la ecuación (3).

$$k = \frac{(P_0 - P_e)}{(1 - P_e)} \quad (3)$$

Es difícil determinar el valor del índice Kappa para indicar un buen acuerdo, ya que lo que constituye un buen acuerdo dependerá del uso que se le dé a la evaluación.

Aunque siempre es una escala subjetiva, Landis y Koch [19] propusieron unos límites para el grado de concordancia estimado con el resultado del cálculo de Kappa.

Los valores de validación de los niveles de acuerdo se pueden adaptar valores de referencia para cada caso de aplicación de acuerdo con las necesidades del proceso, como muestra la Tabla 4. $K = 0$, significa que la concordancia observada coincide con la que ocurriría por puro azar. $K > 0$ señalan mayor concordancia que la que se esperaría por el puro azar. $K = 1$, se trata de una concordancia perfecta. $K < 0$, significa existencia de discordancia [27].

El índice Kappa no es fácil de interpretar en términos de la precisión de una sola observación. El problema es el mismo que surge con coeficientes de correlación para el error de medición en los datos continuos. Los rangos en que se debe mover el índice Kappa serán entre la probabilidad de que los inspectores estén de

acuerdo, debido a que ambos están en lo correcto (qxq) y la probabilidad de que ambos están equivocados, $((1 - q) * (1 - q))$ [30]. De esta forma, la proporción observada de acuerdo para los inspectores es la suma de la proporción observada en la diagonal principal (P_0) está definida en la ecuación (4)

$$P_0 = q^2 + (1 - q)^2 \quad (4)$$

Incluyendo P_e y los valores de desacuerdo entre el Observador y el valor real de las piezas se obtiene la Tabla 5.

El Acuerdo Observado (Acobs) para los inspectores es la suma de las proporciones observadas en la diagonal principal tal como lo define la ecuación (5).

$$Ac_{obs} = p1 * q + (1 - p1) * (1 - q) \quad (5)$$

La probabilidad esperada representa la proporción de acuerdo que debería existir dadas las condiciones de calidad del proceso teniendo en cuenta sólo el azar. Adicionalmente, la probabilidad esperada en concordancia perfecta estaría dada por la ecuación (6).

$$P_e = q^2 + (1 - q)^2 \quad (6)$$

Tabla 4. Interpretación índice de Kappa.

Kappa (K)	Evaluación del grado de concordancia
< 0,6	Requiere reforzar capacitación.
≥ 0,6	Nivel aceptable de concordancia en el proceso.

Fuente: Adaptado de Landis y Kotch [31].

En el caso de la Tabla 5, la diagonal principal representa la proporción de que un Observador esté en lo correcto (q) ya que está coincidiendo con el valor real de las piezas y la diagonal secundaria representa, la proporción de que el Observador esté equivocado (1-q). Reemplazando los valores de las diagonales en la ecuación (6) se tiene que la proporción esperada estaría dada por la ecuación (7).

$$P_e = (p1 * q + (1 - p1) * (1 - q))^2 + ((1 - p1) * q + p1 * (1 - q))^2 \tag{7}$$

$$P_e = q^2 + (1 - q)^2 - 2 * (1 - 2q)^2 * p1 * (1 - p1)$$

De esta forma el índice Kappa (k) en términos de p1 y q se expresa como muestra la ecuación (8).

$$k = \frac{(q^2 + (1 - q)^2) - (q^2 + (1 - q)^2 - 2 * (1 - 2q) * p1 * (1 - p1))}{1 - (q^2 + (1 - q)^2 - 2 * (1 - 2q)^2 * p1 * (1 - p1))}$$

$$k = \frac{(2(1 - 2q)^2 * p1 * (1 - p1))}{(1 - q^2 - 1 + 2q - q^2 + 2 * (1 - 2q)^2 * p1 * (1 - p1))}$$

$$k = \frac{p1 * (1 - p1)}{\frac{q * (1 - q)}{(1 - 2q)^2} + p1 * (1 - p1)} \tag{8}$$

La ecuación (8) muestra que sólo en caso de que q sea igual a: 1,0; 0,0 o 0,5, todas las observaciones correctas corresponderán a una variable aleatoria y el índice Kappa va a depender de p1. Teniendo un máximo cuando p1 = 0,5. Por lo tanto, el índice Kappa será específico para una población dada, de acuerdo con la calidad del proceso (q), como se muestra en la Figura 5.

En la Figura 5 se observa que las líneas corresponden muy de cerca a las categorías de Landi y Koch [30], donde Kappa varía de acuerdo con p1 y el grado de acuerdo va a depender de la calidad del proceso. Además, cuando la probabilidad de un si verdadero (p1) se acerca a cero o a uno, el índice Kappa esperado

Tabla 5. Contingencia con desacuerdo entre el Observador y el valor real de las piezas.

		Observador B	
		Aceptar	Rechazar
Observador A	Aceptar	p1 * q	p1 * (1 - q)
	Rechazar	(1 - p1) * q	(1 - p1) * (1 - q)

se hace más pequeño, se evidencia un cambio notable en los extremos cuando el acuerdo es muy bueno. Cuando el acuerdo no es perfecto, si una de las dos categorías es pequeña en comparación con la otra, el índice Kappa será pequeño, no importa lo bueno que el acuerdo sea.

Cuando existe desbalanceo importante entre la proporción de piezas defectuosas y la proporción de piezas buenas, el valor del índice Kappa será pequeño. Esto debido a que el proceso no se comporta bajo la influencia sólo del azar. En este caso, se estaría interpretando que falta concordancia, cuando realmente la selección de piezas no permite juzgar adecuadamente el nivel de concordancia entre los inspectores.

APLICACIÓN DE LA METODOLOGÍA DE ANÁLISIS DE RIESGOS EN UN SISTEMA PRODUCTIVO

La Figura 6 expone el contexto inicial en el que se evidencia la necesidad de aplicar esta herramienta y a continuación se definen las etapas para ejecutar un estudio R&R en un sistema productivo. La aplicación del método implica definir las condiciones de entrada, es decir seleccionar n piezas del proceso y establecer el estado de referencia.

Del mismo modo, R&R implica: definir los inspectores que serán evaluados de acuerdo con su rol y al impacto que tenga en la selección de piezas; exponerles brevemente en qué consiste la prueba para evitar errores en la valoración y suministrar y; explicar claramente el método de inspección a aplicar. Posteriormente, es necesario definir el estado de las piezas, lo que implica iniciar la revisión de las n piezas y el registro de la decisión (Aceptar o

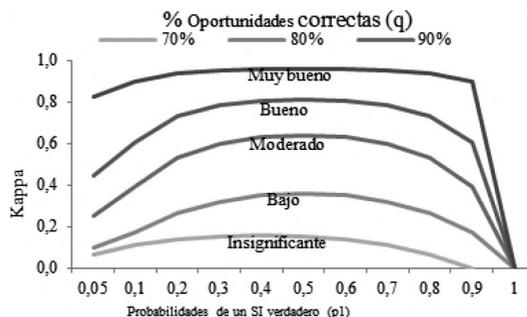


Figura 5. Índice Kappa vs p1, caracterizado por q.

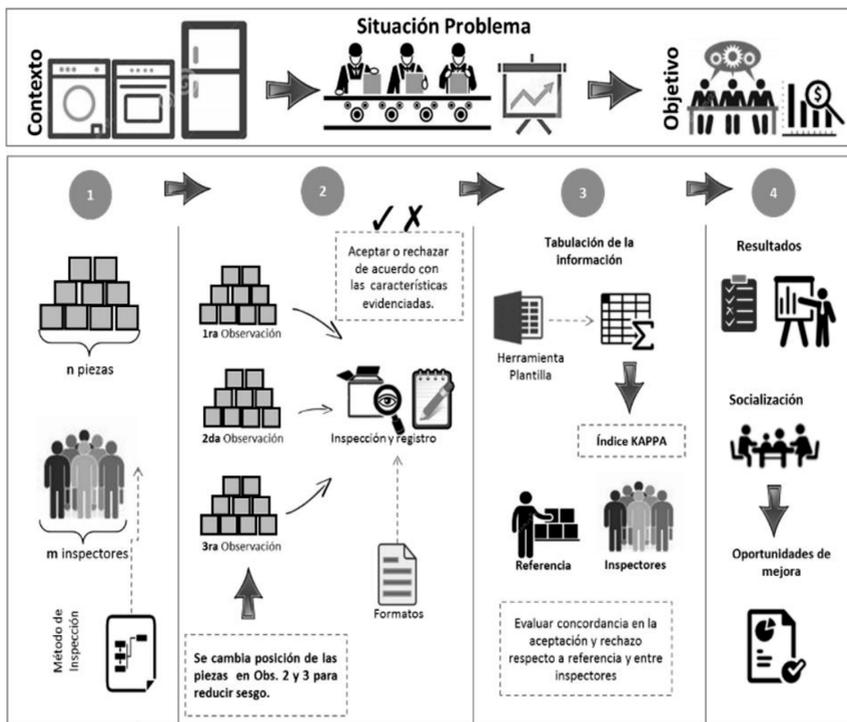


Figura 6. Esquema de aplicación de un estudio R&R en proceso.

Rechazar) en el formato suministrado. Finalizada esta observación, se cambia la posición de algunas piezas y se inicia con la segunda observación registrando la decisión tomada por cada pieza (Aceptar o Rechazar); se repite el mismo procedimiento para la tercera observación.

Finalmente, se calcula el índice Kappa y socializar resultados. Por lo tanto, es necesario tabular la información en la herramienta desarrollada para establecer el índice Kappa respecto al estándar. En la socialización de los resultados es fundamental recopilar las oportunidades de mejora, tanto para el proceso de inspección visual, como para el sistema de valoración propuesto.

Una de las dificultades que se ha identificado en la aplicación del estudio R&R por atributos, usando el método de análisis de riesgos, es la tabulación de datos y realización de los cálculos necesarios para hallar el índice Kappa. Por tanto, se elabora una herramienta que simplifica el ingreso de datos y arroja de manera automática, con ayuda de macros y tablas dinámicas, los resultados de la prueba R&R. A continuación, se describe el procedimiento

general que se debe llevar a cabo para obtener los resultados. La Figura 7 resume el procedimiento a seguir en la aplicación del método de análisis de riesgos haciendo uso de una herramienta propuesta.

La herramienta propuesta fue desarrollada a partir del Visual Basic. Una vez se tienen las tablas de contingencia con los datos recopilados, en la pestaña “evaluación de resultados” se podrá visualizar, para cada inspector, el índice Kappa. Usando la escala propuesta por Landis y Koch [18] se realiza la evaluación de los resultados a partir del resultado del índice Kappa.

RESULTADOS

Los resultados expuestos a continuación se obtuvieron a partir de simulación Montecarlo. La distribución se seleccionó en razón a que se deberían generar números aleatorios desde una variable aleatoria discreta, que modelaran el número acuerdos y desacuerdos con tres niveles diferentes de inspección y tres tasas diferentes de capacitación de los operarios.

El primer escenario de inspección establece un proceso de inspección normal, típico de procesos

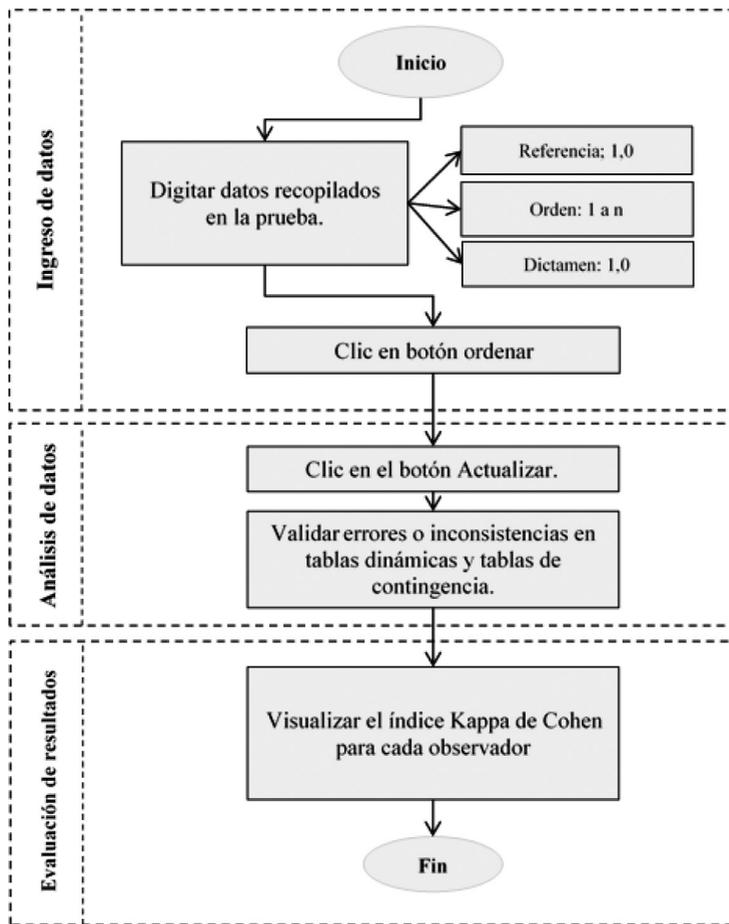


Figura 7. Aplicación del método en la herramienta diseñada.

donde las no-conformidades presentadas son habituales y existe caracterización de las mismas, pero los procesos de actualización son deficientes y los operarios no participan la unificación de criterios. El segundo escenario simula un proceso de inspección ineficiente o muy malo, donde las probabilidades de incurrir en errores tipo I son muy altas, en este caso no hay caracterización de las no-conformidades y no hay buenas condiciones para la identificación. Finalmente se describe un tercer escenario donde se simula un proceso de inspección muy bueno, el cual es propio de no-conformidades habituales, bien caracterizadas, definidas a partir de las necesidades del cliente y de la unificación de criterios con participación de empleados conocedores del proceso.

En cuanto a la capacitación, en el primer nivel simula procesos de inspección donde sólo algunos

de los inspectores participan en la caracterización de no-conformidades, sin embargo, todos reciben capacitación y el proceso de inspección cuenta con alta variabilidad en cuanto a la experiencia, pero, han pasado por todos los procesos de producción, aunque no hayan participado directamente en el proceso de inspección; este nivel de capacitación se llamará Medio. El segundo nivel simula procesos de inspección que reciben capacitación cada vez que se observa características de una no-conformidad documentada, además todos los inspectores participan en la unificación de criterios y tienen experiencia en los subprocesos de inspección realizados en los procesos de producción; este escenario corresponde a un nivel de capacitación "Alto". El tercer nivel de capacitación, que se denomina "Bajo", simula procesos de inspección en donde los inspectores no participan en la unificación de criterios y hay poca capacitación. Es decir, los inspectores desconocen

las características del producto relevantes para el cliente, no tienen una experiencia amplia en procesos de inspección, no tienen una clara unificación de criterios y no han recibido capacitación respecto a la adecuada aplicación de un método de inspección visual.

Los valores de referencia fueron obtenidos a partir de una prueba en campo, en una empresa manufacturera que se caracterizaba por el primer escenario tanto en eficacia de inspección, como en capacitación. En la prueba de campo: 6 inspectores, evaluaron 20 piezas, en 3 ocasiones. Los resultados medios son observados en la Tabla 6.

Con base en los resultados de la prueba de campo se obtuvieron las probabilidades de acertar en una inspección normal y un nivel de capacitación Medio (ver Tabla 7).

Usando la Tabla 7 se generaron 1000 observaciones a través de simulación de Montecarlo cuando se tienen piezas no-conformidades y sin no-conformidades. Para todos los escenarios se tuvo en cuenta que tanto las condiciones de la inspección como la capacitación de los inspectores afecta el número de acuerdos [15, 16], para este estudio fue considerado que cada uno de los niveles, su riesgo es afectado en un 10%, según sea el caso (ver Tabla 8).

Escenario 1: proceso de inspección normal

Este escenario corresponde a procesos de inspección donde las condiciones de diagnóstico son adecuadas, existe un puesto de trabajo que cuenta con la iluminación adecuada, pero no se cuenta con herramientas y técnicas que faciliten el proceso de inspección. Usando simulación Montecarlo y las consideraciones expuestas anteriormente.

A partir de lo anterior, se calcula el índice Kappa (ver Tabla 9) y se analiza según la clasificación establecida por Landis y Koch [18]. Es posible establecer que

la condición de diagnóstico genera un escenario no homogéneo, por las diferencias marcadas entre el índice Kappa de cada una de las condiciones de capacitación simuladas. Especialmente en lo referente a la condición de capacitación baja, por lo tanto, se evidencia que un proceso de inspección visual por atributos es muy sensible dado que es un proceso que está altamente influenciado por las características asociadas al inspector como lo es la capacitación [18].

Escenario 2: proceso de inspección ineficiente

Este escenario corresponde a procesos de donde no se tienen métodos ni procedimientos de diagnóstico definidos. Este escenario también se corre considerando los tres tipos de capacitación de los inspectores. Entonces en las condiciones de este escenario, los inspectores tienen un nivel de concordancia bajo e insignificante respecto a la referencia. Esto indica que el escenario no es idóneo para llevar a cabo un proceso de inspección visual por atributos ya que la probabilidad de errores tipo I y tipo II se incrementan notablemente. Además, independiente del tipo de capacitación que presenten los inspectores, las piezas no serán clasificadas adecuadamente (ver Tabla 10).

Escenario 3: proceso de inspección muy bueno

Este escenario corresponde a procesos de inspección donde se tiene un método de diagnóstico definido, con criterios previamente establecidos. En este

Tabla 6. Información de referencia.

Tabla de Contingencia A*Referencia			
A	Referencia		Total
	0	1	
0	15	4	19
1	6	35	41
Total	21	39	60

Tabla 7. Relación probabilística entre el estado real y diagnóstico.

		Estado Real	
		No-Conformidad	Conformidad
Diagnóstico	No-Conformidad	0,7142	0,8974
	Sin No-Conformidad	0,2858	0,1026

Tabla 8. Resumen aciertos por escenarios vs. Nivel de capacitación.

Escenario de inspección	Estado de la pieza	Nivel de Capacitación					
		Media		Alta		Baja	
		Diagnóstico No-conforme (0)	Diagnóstico conforme (1)	Diagnóstico No-conforme (0)	Diagnóstico conforme (1)	Diagnóstico No-conforme (0)	Diagnóstico conforme (1)
1	No-conforme (0)	16	5	17	2	13	11
	Conforme (1)	5	35	4	37	8	28
2	No-conforme (0)	13	12	13	10	10	16
	Conforme (1)	8	27	8	29	11	23
3	No-conforme (0)	17	2	18	0	14	22
	Conforme (1)	4	37	3	39	7	31

contexto los niveles de capacitación alta y media presentan un nivel de acuerdo muy bueno, y el nivel de capacitación baja presenta un nivel de acuerdo moderado, respecto a la referencia, lo cual indica que son idóneos para llevar a cabo procesos de inspección visual por atributos (ver Tabla 11), sin embargo, si se tiene un nivel de capacitación baja es necesario tener controles adicionales al proceso, porque el sistema de medición puede salirse de control.

Los resultados asociados a este escenario generalmente están influenciados por mejoras realizadas al proceso, tales como estandarización de procesos de diagnóstico, validación de criterios de acuerdo con los requerimientos del cliente, incorporación de tecnología, así como puestos de diagnóstico adecuados tanto en métodos como en herramientas para realizar la inspección.

Tabla 9. Índice Kappa escenario 1.

	Medio	Alto	Bajo
Índice Kappa	0,637	0,775	0,3262

Tabla 10. Índice Kappa escenario 2.

	Medio	Alta	Baja
Índice Kappa	0,2982	0,3548	0,0625

Tabla 11. Kappa escenario 3.

	Medio	Alta	Baja
Índice Kappa	0,7752	0,8863	0,4565

A partir de la simulación de las diferentes condiciones de inspección, así como del nivel de capacitación, se evidencia que los niveles de acuerdo más bajo, se deben al nivel de capacitación, la cual se relaciona con la experiencia, participación en la socialización y unificación de los criterios de inspección y diagnóstico (ver Figura 8). Se verifica que cuando se tiene unas condiciones de inspección deficientes, el nivel de capacitación alta y media son muy cercanos, lo cual es evidencia de que la capacitación también tiene su punto máximo. A partir de ese momento, el nivel de acuerdo deja de depender del inspector y sus condiciones.

Evaluando los resultados de la Figura 8, respecto a la Figura 5, se puede establecer que si se mantiene una capacitación adecuada donde todos reciben capacitación y el proceso de inspección cuenta con alta variabilidad en cuanto a la experiencia, sin embargo, existe alguna participación en la caracterización, documentación y establecimiento de criterios en la definición de no-conformidades, inspectores participan en la unificación de criterios y tienen experiencia en los subprocesos de inspección realizados en los procesos de producción para el escenario de inspección normal (escenario 1), se tiene un nivel de calidad del proceso (q) entre 80% y 95%, para el escenario de inspección ineficiente (escenario 2) el nivel de calidad del proceso (q) disminuye en al menos un 20% respecto al escenario de inspección muy bueno (escenario 3), donde el nivel de calidad se ubica entre 95% y 99%.

El comportamiento del índice Kappa de Cohen en cada escenario, para cada nivel de capacitación es

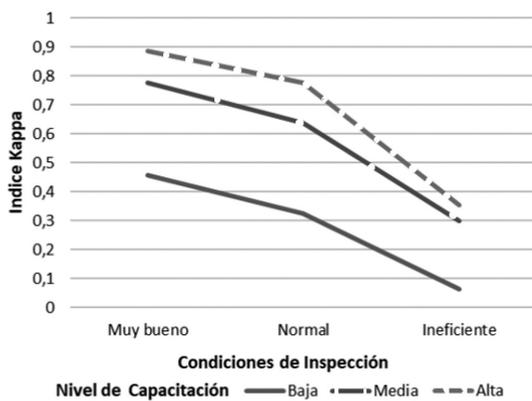


Figura 8. Escenarios de Inspección vs Nivel de Capacitación.

consistente con el número de acuerdos simulados, como es el caso del nivel alto, que presentó el mayor nivel de concordancia (índice Kappa), respecto a los demás.

En un proceso de producción se pueden obtener índices de concordancia como los expuestos en el escenario 3, generalmente después de aplicar mejoras en el proceso de inspección, estandarización de métodos y capacitación. Es importante resaltar que se debe tener en cuenta características físicas como la capacidad visual al momento de seleccionar los inspectores y diseñar procesos de capacitación para este tipo de procesos.

CONCLUSIONES

La metodología desarrollada para la aplicación del método de análisis de riesgos en este trabajo valora cuantitativamente el desempeño de un sistema de inspección visual permitiendo medir el impacto de diferentes escenarios con variaciones en el nivel de capacitación, experiencia, unificación de criterios, características físicas, métodos y estándares disponibles. En este orden de ideas, para el caso de estudio se determinó que, de los tres niveles de capacitación evaluados.

Bajo las condiciones evaluadas en el escenario de condiciones de inspección ineficiente, para los tres niveles de capacitación, no se poseen las competencias necesarias para desempeñarse adecuadamente en un proceso de inspección visual por atributos, dado que los niveles de concordancia respecto a la referencia son bajos e incluso insignificantes de acuerdo.

A partir de las simulaciones, para los diferentes inspectores en los tres escenarios planteados, se puede concluir que es posible obtener niveles de calidad del proceso entre 95% y 99%, generalmente después de aplicar mejoras en procesos de inspección, estandarización de métodos, unificación de criterios y capacitación del personal; teniendo en cuenta, además, características físicas como la capacidad visual al momento de seleccionar los operadores para este tipo de procesos.

El éxito del sistema propuesto para el estudio R&R por atributos radicará principalmente en el interés de la organización en la unificación de criterios sobre no-conformidades de los diferentes procesos. Sin olvidar que es una tarea ardua que requiere tanto una valoración cuantitativa del nivel de concordancia, como herramientas simples y claras que se transformen en una guía durante la ejecución de sus tareas.

Por otra parte, y en relación con el estudio de la bibliografía, se puede concluir que: los análisis R&R están enfocadas principalmente a situaciones donde las variables de interés son medibles (continuas); pero su aplicación en sistemas cuyas variables son discretas, como la inspección visual por atributos, no está difundida, por la complejidad que representa no disponer de un instrumento que elimine la subjetividad en el dictamen de los inspectores.

De igual forma, en la bibliografía consultada, son pocos los estudios R&R por atributos comparados con los documentos encontrados para variables continuas. Aún con relación al estudio de la bibliografía, es posible concluir que, se debe prestar especial atención a los métodos de inspección a aplicar y a los elementos de los que disponen para tomar decisiones, tales como, clasificación de defectos, zonificación y definición de etapas macro del método de inspección. Todo esto para que se conviertan en las herramientas principales para enfrentar la subjetividad de su tarea.

Los métodos aplicados en los estudios R&R generalmente se limitan al cálculo del porcentaje de repetibilidad y del porcentaje de reproducibilidad, conocidos como “método corto”, una versión menos exhaustiva que la propuesta en el método de análisis de riesgo.

Con relación al análisis de riesgo, se puede concluir que es un método simple y confiable, respecto a los propuestos en la literatura evaluada que, aunque supone un nivel de operatividad alto, usa como estadístico el índice Kappa de Cohen. Este estadístico refleja la realidad del proceso, dado que retira el nivel de acuerdo debido al azar. En este estudio se simularon todas las observaciones como variable aleatoria y se evidenció que el Índice kappa depende de la probabilidad un acuerdo entre la referencia y el diagnóstico y por lo tanto está directamente capacidad del proceso de medición q .

Finalmente, al validar manualmente los cálculos para definir el nivel de concordancia versus la información suministrada por la herramienta diseñada, se concluye que la herramienta simplifica en gran medida el procedimiento de tabulación de los datos. Dicho esto, la metodología diseñada para el estudio R&R por atributos, puede adaptarse sin mayores reformas a diferentes semielaborados y aplicarse en diferentes tipos de industria, donde se busque evaluar el nivel de concordancia en sus sistemas de inspección visual por atributos. Además, la metodología por planteada podría extenderse a una simulación con estructura de diseño de experimentos para poder analizar las interacciones entre los factores.

REFERENCIAS

- [1] L.F. Romero e D.B. Romero. “Estudio R y R por atributos de un proceso de inspección en el sector automotriz”. *Produção em Foco*. Vol. 1 N° 1, pp. 171-194. 2011. DOI:10.14521/P2237-5163.2011.0001.0010.
- [2] R. Hoyer and B. Hoyer. “What is Quality”. *Quality Progress*. Vol. 34 N° 7, pp. 53-62. 2001. ISSN: 0033524X.
- [3] AIAG Automotive Industry Action Group. “Measurement System Analysis”. 2006 Third IEEE International Workshop on Electronic Design, Test and Applications (DELTA), pp. 248-252 Massachusetts, USA. 2006. ISSN: 978-0-7695-2500-6. DOI: 10.1109/DELTA.2006.62.
- [4] M. Chen and J. Lyu. “A novel evaluation model for measurement system analysis”. *Production Planning and Control*. Vol. 20 N° 5, pp. 420-430. 2009. ISSN: 13665871. DOI: 10.1080/09537280902897777.
- [5] D. Dalalah and A. Diabat. “Repeatability and reproducibility in med labs: a procedure to measurement system analysis”. *IET Science, Measurement & Technology*. Vol. 9 N° 7, pp. 826-835. ISSN: 1751-8822. DOI: 10.1049/iet-smt.2014.0363.
- [6] M.G. Otero Fernández, A.A. Mazorra Lopetey y M. Ulloa Enríquez. “Evaluación del sistema de medición en la fabricación de gel de hidróxido de aluminio”. *Revista Cubana de Farmacia*. Vol. 48 N° 4, pp. 533-541. 2014. ISSN: 0034-7515.
- [7] J.O. Villalvazo, C. Camargo, W. Enrique, J. De Vega, O.R. López, B. Jesús, E.O. Tiznado, J. Andres y L. Barreras. “Validación de un sistema de medición aplicado a un estudio de termografía sensorial para la detección de desórdenes de traumas acumulados”. *Revista Ingeniería Industrial*. Vol. 10 N° 1, pp. 69-84. 2011. ISSN: 0717-9103.
- [8] B. Kocak, E.S. Durmaz, E. Ates and M.B. Uluhan. “Radiogenomics in clear cell renal cell carcinoma: Machine learning-based high-dimensional quantitative CT texture analysis in predicting PBRM1 mutation status”. *American Journal of Roentgenology*. Vol. 212 N° 3, pp. 55-63. 2019. ISSN: 0361803X. DOI: 10.2214/AJR.18.20443.
- [9] M. Bisele, M. Bencsik, M. Lewis and C. Barnett. “Optimisation of a machine learning algorithm in human locomotion using principal component and discriminant function analyses”. *PLoS One*. Vol. 12 N° 9, pp. 1-19. 2017. ISSN: 19326203. DOI: 10.1371/journal.pone.0183990.
- [10] S. Unnikrishnan, J. Donovan, R. Macpherson and D. Tormey. “Machine Learning for Automated Quality Evaluation in Pharmaceutical Manufacturing of Emulsions”. *Journal of Pharmaceutical Innovation*. Vol. 15 N° 3, pp. 392-403. 2020. ISSN: 18725120. DOI: 10.1007/s12247-019-09390-8.
- [11] C. Heng, H. Shen, F. Wang and Y. Chen. “Calibrated color measurement of cashmere using a novel computer vision system”. *Measurement*. Vol. 185, pp. 1-11. 2021. ISSN: 0263-2241.
- [12] M. Narwaria, W. Lin, I.V. McLoughlin, S. Emmanuel and L. Chia. “Fourier transform-based scalable image quality measure”. 2012. *IEEE Transactions on Image Processing*.

- Vol. 21 N° 8, pp. 3364-3377. 2012. ISSN: 10577149. DOI: 10.1109/TIP.2012.2197010.
- [13] N. Van Eck and L. Waltman. "VOSviewer: A computer program for bibliometric mapping". *Scientometrics*. Vol. 84 N° 2, pp. 886-897. 2010. ISSN: 1588-2861.
- [14] S. Windsor. "Attribute Gage R&R". *Six Sigma Forum Magazine*. Vol. 2 N° 4, pp. 23-28. 2003. ISSN: 1539-4069.
- [15] W.N. Van Wieringen. "Statistical Models for the Precision of Categorical Measurement". *Measurement*. Vol. 20, pp. 37-46. 2003. ISSN: 0263-2241.
- [16] J. De Mast and W.N. Van Wieringen. "Measurement system analysis for categorical measurements: agreement and kappa-type indices". *Journal of Quality Technology*. Vol. 39 N° 3, pp. 191-202. ISSN: 0022-4065.
- [17] C. Marques, N. Lopes, G. Santos, I. Delgado and P. Delgado. "Improving operator evaluation skills for defect classification using training strategy supported by attribute agreement analysis". *Measurement*. Vol. 119, pp. 129-141. 2018. ISSN: 0263-2241. DOI: 10.1016/j.measurement.2018.01.034.
- [18] J. Lyu and M. Chen. "Gauge capability studies for attribute data". *Quality and Reliability Engineering International*. Vol. 24, pp. 71-82. 2008. ISSN: 07488017.
- [19] G. Altobelli, P. Pignatelli and P. Casey. "Application of the hypothesis analysis method using cohen's kappa to measure the agreement between sorters". *Journal of the Society of Leather Technologists and Chemists*. Vol. 94 N° 4, pp. 144-148. 2010. ISSN: 0144-0322.
- [20] J. de Mast. "Agreement and Kappa-Type Indices". *The American Statistician*. Vol. 61 N° 2, pp. 148-153. 2007. DOI: 10.1198/000313007X192392.
- [21] J. de Mast, T. Akkerhuis and T. Erdmann. "The statistical evaluation of categorical measurements: Simple scales, but treacherous complexity underneath". *Quality Engineering*. Vol. 26 N° 1, pp. 16-322. 2014. ISSN: 08982112. DOI:10.1080/08982112.2013.846062.
- [22] E. Cortés-Reyes, J.A. Rubio-Romero y H. Gaitán-Duarte. "Métodos estadísticos de evaluación de la concordancia y la reproducibilidad de pruebas diagnósticas". *Revista Colombiana de Obstetricia y Ginecología*. Vol. 61 N° 3, pp. 247-255. 2010. ISSN: 0034-7434.
- [23] L. de F. Lazarini, L. Ohler, J. Schirmer and B. Roza. "Validation of the american quality assessment model and performance improvement to the brazilian transplant". *Revista Latino-Americana de Enfermagem*. Vol. 28. 2020. DOI:10.1590/1518-8345.3249.3252.
- [24] M. Jalayer and H. Zhou. "Evaluating the safety risk of roadside features for rural two-lane roads using reliability analysis". *Accident Analysis and Prevention*. Vol. 93, pp. 101-112. 2016. ISSN: 00014575. DOI: 10.1016/j.aap.2016.04.021.
- [25] W. Canestaro, D. Pritchard, L. Garrison, R. Dubois and D. Veenstra. "Improving the efficiency and quality of the value assessment process for companion diagnostic tests: The companion test assessment tool (CAT)". *Journal of Managed Care Pharmacy*. Vol. 21, N° 8, pp. 700-712. 2015. ISSN: 10834087. DOI: 10.18553/jmcp.2015.21.8.700.
- [26] N. Baudet, J. Marie and M. Pillet. "The Visual Inspection of Product Surface". *Food Quality and Preference*. Vol. 27 N° 2, pp. 153-160. 2012. ISSN: 0950-3293.
- [27] SAMIUC. "Medidas de concordancia". Sociedad Andaluza de Medicina Intensiva, Urgencias y Coronarias. Fecha de consulta: 20 de Agosto 2020. URL: <https://www.samiuc.es/estadisticas-variables-binarias/medidas-de-concordancia/kappa-de-cohen/>
- [28] J.A. Cohen. "A Coefficient of Agreement for Nominal Scales". *Educational and Psychological Measurement*. Vol. 20 N° 1, pp. 37-46. 1960. DOI: 10.1177/001316446002000104.
- [29] J.L. Fleiss. "Measuring Nominal Scale Agreement Among Many Raters". *Psychological Bulletin*. Vol. 76 N° 5, pp. 378-382. 1971. DOI: 10.1037/h0031619.
- [30] J.M. Bland. "Cohen's Kappa". University of York Department of Health Science. Inglaterra. 2008. URL: <https://www-users.york.ac.uk/~mb55/msc/clinimet/week4/kappash2.pdf>
- [31] J.R. Landis and G.G. Koch. "The Measurement of Observer Agreement for Categorical Data". *Biometrics*. Vol. 33 N° 1, pp. 159-174. 1977. ISSN: 1541-0420.