



Quaglino, Marta
Pagura, José Alberto
Dianda, Daniela
Hernandez, Lucia
Puigsubira, Cristina

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística

ESTUDIO DE CAPACIDAD DE SISTEMAS DE MEDIDA UTILIZANDO INTERVALOS DE CONFIANZA GENERALIZADOS.

1. Introducción

1.1. Fundamentos de los estudios de sistemas de medida.

Las metodologías estadísticas para la mejora continua de procesos dependen fuertemente de datos de mediciones para identificar oportunidades de mejora. De ello se desprende la importancia que reviste el estudio cuidadoso de las propiedades del sistema que proporciona tales mediciones, ya que una parte de la variabilidad observada en ellas será inherente a la característica de calidad del producto o servicio que está siendo medida, pero otra parte será producto de la forma en que se realice la medición. Por lo tanto, si logra determinarse con anterioridad que el sistema de medición que se utiliza es adecuado, en el sentido de no introducir variabilidad adicional en las mediciones, se habrá eliminado una posible fuente de variación en el resultado del proceso. Un reflejo de la relevancia que tiene el proceso de medición es la inclusión de una etapa específica en la conocida estrategia de calidad Seis Sigma, que se denomina "MEDIR".

Cuando un sistema de medición es considerado como un proceso en el que intervienen la característica de interés del producto, los operarios que realizan las mediciones, el instrumento de medición, el medio ambiente, los procedimientos establecidos, etc., resulta fácil ver que son muchas las fuentes de variación que pueden ocasionar mediciones inexactas o poco precisas.

Para evaluar si el sistema de medición es adecuado debe tomarse conocimiento de todos los posibles factores que pueden afectar el proceso de medición. Para ello, existen estudios específicos orientados a evaluar cada una de las propiedades de las que debe gozar un buen sistema de medición.

En particular, los estudios de capacidad de sistemas de medición, más conocidos como estudios de repetibilidad y reproducibilidad (o estudios R&R), fueron diseñados para estudiar la precisión del sistema de medición, esto es, identificar cuánta variación está asociada al mismo y compararla con la variación total del proceso. Esta operación permite



decidir si el sistema es o no capaz de producir mediciones que reflejen el verdadero comportamiento del proceso.

Para llevar a cabo estos estudios, se supone que cada valor de medición y obtenido puede escribirse como la suma del verdadero valor x de la medida en la unidad, más un error ε debido al proceso de medición, es decir, $y_i = x_i + \varepsilon_i$.

Si se asume que x y ε son variables aleatorias que se distribuyen normal e independientemente con medias μ_x y 0 , y variancias σ_p^2 , representando a la variabilidad de las unidades o partes; y σ_{Gauge}^2 , haciendo referencia a la variabilidad del sistema de medición, respectivamente, entonces, la variabilidad total observada en las mediciones puede escribirse como la suma de estas dos componentes de variancia: $\sigma_{Total}^2 = \sigma_p^2 + \sigma_{Gauge}^2$.

A su vez, dos componentes de la variabilidad del sistema de medición son la reproducibilidad y la repetibilidad, las cuales hacen referencia a la variabilidad inherente al instrumento de medición y a la variabilidad introducida por los distintos operarios que operan el instrumento, respectivamente, lo cual lleva a la siguiente descomposición:

$$\sigma_{Gauge}^2 = \sigma_{Repetibilidad}^2 + \sigma_{Reproducibilidad}^2$$

Es claro que para evaluar la capacidad del sistema de medición, debe poder disponerse de estimaciones de las distintas dispersiones que hacen a la variabilidad total, para lo cual se han propuesto diversos métodos dependiendo de si tales dispersiones se consideran en forma de rangos estadísticos o desvíos estándares. Dado el gran crecimiento en la disponibilidad de herramientas informáticas producido en los últimos tiempos, el método más ampliamente usado para llevar a cabo este tipo de estudios es el diseño de experimentos analizando los datos por medio del análisis de la variancia (ANOVA), lo cual conlleva además una ganancia tanto en cantidad como en precisión de los resultados obtenidos frente a los métodos más básicos. Dicha técnica permitirá analizar los datos recogidos en el experimento, investigando la contribución del conjunto de factores considerados a la variabilidad total de las mediciones, y estimar puntualmente las variancias asociadas a cada componente.

1.2. Diseño de experimentos para estudios R&R

Un diseño experimental ampliamente utilizado para llevar a cabo los estudios de repetibilidad y reproducibilidad es aquel que considera como factores a los operarios que realizan las mediciones y a las piezas o unidades medidas, en todas las combinaciones posibles de sus niveles, considerando también la posible existencia de interacción entre ambos factores. Dado el contexto en que se realiza este tipo de estudios, lo habitual es que tanto las piezas como los operarios que participan del experimento constituyan una



muestra aleatoria de poblaciones más grandes de piezas y operarios, razón por la cual el modelo se asume de efectos aleatorios.

Sea entonces el siguiente modelo para representar las mediciones realizadas por o operarios midiendo r veces cada una de p piezas:

$$Y_{ijk} = \mu_Y + P_i + O_j + (PO)_{ij} + \varepsilon_{ijk}; \quad i = 1, \dots, p; \quad j = 1, \dots, o; \quad k = 1, \dots, r$$

Tanto los efectos aleatorios del modelo, $P_i, O_j, (PO)_{ij}$, como el término de error, ε_{ijk} , son variables aleatorias distribuidas normal e independientemente con media cero y variancias dadas por $V(P_i) = \sigma_P^2, V(O_j) = \sigma_O^2, V(PO_{ij}) = \sigma_{OP}^2$ y $V(\varepsilon_{ijk}) = \sigma^2$ respectivamente.

La tabla ANOVA para este diseño, incluyendo las esperanzas de los cuadrados medios, se muestra en la siguiente tabla:

Tabla1: Análisis de la variancia para los datos del ejemplo.

Fuente de Variación	Grados de libertad	Cuadrados medios (CM)	Esperanzas de los CM
Operarios	$o - 1$	$CM_o = \frac{pr \sum_{j=1}^o (\bar{y}_{.j} - \bar{y}_{...})^2}{o - 1}$	$\theta_o = \sigma^2 + r\sigma_{PO}^2 + pr\sigma_o^2$
Partes	$p - 1$	$CM_p = \frac{or \sum_{i=1}^p \sum_{j=1}^o (\bar{y}_{i..} - \bar{y}_{...})^2}{p - 1}$	$\theta_p = \sigma^2 + r\sigma_{PO}^2 + or\sigma_p^2$
Interacción	$(p - 1)(o - 1)$	$CM_{PO} = \frac{r \sum_{i=1}^p \sum_{j=1}^o (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2}{(p - 1)(o - 1)}$	$\theta_{PO} = \sigma^2 + r\sigma_{PO}^2$
Error (instrumento)	$op(r - 1)$	$CM_\varepsilon = \frac{\sum_{i=1}^p \sum_{j=1}^o \sum_{k=1}^r (y_{ijk} - \bar{y}_{ij.})^2}{op(r - 1)}$	$\theta_\varepsilon = \sigma^2$

A partir de las expresiones de las esperanzas de los cuadrados medios, es fácil obtener estimaciones puntuales para las componentes de variancia, igualando dichos valores esperados a los valores de los cuadrados medios obtenidos en el análisis. De esta manera se obtiene:

$$\hat{\sigma}^2 = CM_\varepsilon; \quad \hat{\sigma}_{PO}^2 = \frac{CM_{PO} - CM_\varepsilon}{n}; \quad \hat{\sigma}_o^2 = \frac{CM_o - CM_{PO}}{pn}; \quad \hat{\sigma}_p^2 = \frac{CM_p - CM_{PO}}{on}$$

Cuando se utiliza este método, es posible encontrar que la estimación de una componente de variancia es negativa, lo cual no resulta razonable, ya que por definición estas componentes son no negativas. Ante esta situación, una alternativa es asumir que



una estimación negativa está indicando que el efecto de la misma no es estadísticamente significativo y por lo tanto asignar el valor cero a esa componente de variancia manteniendo sin cambios las demás estimaciones no negativas.

Una vez obtenidas estas estimaciones puntuales, es posible cuantificar las componentes de variancia de interés en los estudios R&R, es decir, la variabilidad debida al sistema de medición y sus componentes de repetibilidad y reproducibilidad, teniendo en cuenta que la variancia debida a los operarios y a la interacción entre operarios y piezas representan la reproducibilidad del sistema, mientras que la repetibilidad está asociada a la variancia residual. Se tiene entonces:

$$\left. \begin{array}{l} \hat{\sigma}_{repetibilidad}^2 = \hat{\sigma}^2 \\ \hat{\sigma}_{reproducibilidad}^2 = \hat{\sigma}_O^2 + \hat{\sigma}_{OP}^2 \end{array} \right\} \Rightarrow \hat{\sigma}_{Gauge}^2 = \hat{\sigma}_{repetibilidad}^2 + \hat{\sigma}_{reproducibilidad}^2$$

1.3. Indicadores de bondad para los sistemas de medición.

Son numerosos los criterios que se han propuesto para determinar la capacidad de un sistema de medición. Todos ellos están basados en estimaciones de las componentes de variancia que ya se han mencionado, siendo el más usual el indicador de precisión a tolerancia (PTR, por Precisión to Tolerance Ratio):

$$PTR = \frac{6 \sigma_{Gauge}}{LSE - LIE}$$

donde LSE y LIE son los límites de especificación dados para el proceso. Como regla práctica, AIAG (2002) recomienda decidir según el siguiente criterio: si este porcentaje es menor al 10%, el sistema de medición es capaz, si es mayor al 30% el sistema de medición es incapaz, en el sentido de que está introduciendo mucha variabilidad al proceso. Si se obtiene un porcentaje entre el 10% y el 30%, se está en una zona de indecisión, y la decisión depende de factores tales como la capacidad del proceso o el costo de mala clasificación de las unidades.

Varios autores, entre ellos Montgomery y Runger (1993/1994) han señalado que el indicador PTR no necesariamente proporciona una medida de cuán bien se comporta el sistema de medición, ya que un proceso que es altamente capaz respecto de las especificaciones, puede tolerar sistemas de medición que produzcan variabilidades más grandes que procesos menos capaces. Por ello, se ha propuesto un indicador alternativo, construido comparando la variabilidad del sistema de medición con la variabilidad total del proceso, el cual recibe el nombre de %R&R:

$$\% R \& R = \frac{\sigma_{Gauge}}{\sigma_{Total}}$$



Con este indicador, la decisión se toma siguiendo los mismos criterios definidos para el indicador PTR.

Dado que en el contexto de estos estudios, la variabilidad asociada al sistema de medición está compuesta por una variabilidad debida a la repetibilidad del sistema y otra debida a la reproducibilidad, es común también evaluar cuál de los dos factores es el que

tiene mayor participación, mediante los cocientes: $\frac{\sigma_{repetibilidad}^2}{\sigma_{Gauge}^2}$ y $\frac{\sigma_{reproducibilidad}^2}{\sigma_{Gauge}^2}$.

Otros indicadores también ampliamente utilizados son: el cociente señal-ruido (SNR, por signal to noise ratio) que refleja la capacidad de discriminación del sistema, y los basados en la probabilidad de mala clasificación, de aplicación evidente en los estudios para atributos.

La cantidad SNR queda definida como $\sqrt{\frac{2 \sigma_p^2}{\sigma_{Gauge}^2}}$ e interpretado como el número de

categorías distintas que el sistema de medición es capaz de distinguir, recomendándose obtener valores iguales o mayores a cinco.

En cuanto a las probabilidades de mala clasificación δ (probabilidad de clasificar una unidad buena, como defectuosa) y β (probabilidad de considerar buena, una "parte" que en realidad es defectuosa), se establecen valores deseables y a partir de la comparación de ellos con los resultados proporcionados por un experimento diseñado se decide acerca de la capacidad del sistema de medida.

Independientemente del indicador que se utilice, en todos los casos será posible obtener una estimación puntual del mismo, ya que, como se mencionó, todos ellos están basados en las estimaciones puntuales de las componentes de variancia ya presentadas. No obstante, dado que se trata de una estimación, la información provista por el indicador podría verse ampliamente enriquecida si pudiera acompañarse de una medida de la incertidumbre asociada al proceso de estimación, es decir, de un intervalo de confianza.

El caso es que para la mayoría de los indicadores mencionados, no existen intervalos de confianza exactos, dado que las combinaciones de variables aleatorias chi-cuadrado a partir de las cuales se construyen las estimaciones puntuales, en general, no siguen una distribución de probabilidad conocida. En tales situaciones, existen dos enfoques ampliamente usados para obtener intervalos de confianza aproximados: por un lado, el método de muestras grandes modificado (MLS, por Modified Large Samples) inicialmente propuesto por Graybill y Wang (1980), y por otro lado, la construcción de intervalos de confianza generalizados, introducidos por Weerahandi (1993).



En este trabajo se presenta una reseña de la metodología de construcción de intervalos de confianza generalizados y una aplicación particular en el contexto de los estudios de sistema de medida para obtener intervalos para los indicadores de bondad utilizados.

2. Intervalos de confianza generalizados.

En su definición convencional, un intervalo de confianza para el parámetro θ se construye a partir de la siguiente expresión:

$$P(A(X) \leq \theta \leq B(X)) = 1 - \alpha \quad (1)$$

en la que $A(X)$ y $B(X)$ son dos estadísticas basadas en un vector de variables aleatorias X proveniente de una población cuya distribución depende del parámetro desconocido θ ; y α es una constante preestablecida entre 0 y 1. Si los valores observados de las dos estadísticas mencionadas son: $a = A(x)$ y $b = B(x)$, entonces el intervalo $[a, b]$ se dice intervalo de confianza para θ , con coeficiente de confianza $(1 - \alpha)$.

Los intervalos de confianza así definidos, gozan de las dos importantes propiedades que se detallan a continuación.

P1) Dada una situación particular en la que se construye una estimación por intervalo para el parámetro de interés θ con coeficiente de confianza $(1 - \alpha)$, si el mismo experimento se repite un gran número de veces para obtener nuevos conjuntos de observaciones, entonces el $(1 - \alpha)\%$ de los intervalos de confianza obtenidos a partir de cada uno de estos conjuntos de datos usando (1) incluirán el verdadero valor del parámetro.

P2) Si se considera un gran número de situaciones independientes en las que se construyen intervalos del $(1 - \alpha)\%$ de confianza para el parámetro de interés en cada caso, entonces el $(1 - \alpha)\%$ de tales intervalos cubrirán al verdadero valor del parámetro correspondiente.

Es claro que, en el contexto de aplicación práctica de este tipo de estimación, la importancia real recae sobre la propiedad P2). No obstante, el cumplimiento de P1) es importante en el sentido de que implica el cumplimiento de P2). Sin embargo son muchas las situaciones en las que la propiedad 1 no se cumple. Es por ello que la propuesta inicial de construcción de lo que luego se denominaría intervalo de confianza generalizado fue ideada de manera que pudiera garantizarse la verificación de esta segunda propiedad, independientemente del cumplimiento o no de la primera.

El enfoque tradicional de la estimación por intervalo se plantea en términos de una cantidad pivotal, es decir, una función del parámetro de interés desconocido, cuya distribución no depende de dicho parámetro, a partir de la cual se construye el intervalo



deseado. Con un enfoque similar, la construcción de intervalos de confianza generalizados se realiza a partir de lo que se llama **cantidad pivotal generalizada (CPG)**, definida por Weerahandi (1993) como sigue:

Sea $R = r(X, x, v)$ una función de un vector de variables aleatorias X , con función de distribución que depende de $v = (\theta, \delta)$, un vector de parámetros desconocidos entre los cuales θ es el parámetro de interés y δ es un vector de parámetros de ruido; y x el vector de valores observados de X . La función $R = r(X, x, v)$ se considera una CPG si:

- Para un valor dado de X , o sea un x fijo, la distribución de probabilidad de R no depende de los parámetros desconocidos, y
- El valor observado de R , llámese $r_{obs} = r(x, x, v)$, es igual al parámetro de interés θ .

El primer requisito se impone para asegurar que, para cada valor del coeficiente de confianza deseado, se pueda hallar un subconjunto de valores de R sin necesidad de conocer los parámetros desconocidos. El segundo requisito garantiza que los enunciados de probabilidad basados en CPGs llevan a regiones de confianza que dependen únicamente de los valores observados x .

Una vez hallada la CPG correspondiente, y establecido el nivel de confianza $(1 - \alpha)$ deseado, la región de confianza generalizada para el parámetro de interés θ se obtiene definiendo, en principio, una región $C_{1-\alpha}$ en el espacio muestral de R de modo que:

$$P(R \in C_{1-\alpha}) = 1 - \alpha \quad (2)$$

y a partir de ella, encontrando el subconjunto Θ_c del espacio paramétrico Θ de θ como:

$$\Theta_c(r) = \{\theta \in \Theta / r_{obs} \in C_{1-\alpha}\} \quad (3)$$

Para las regiones así definidas es posible demostrar que, a largo plazo, el $(1 - \alpha)\%$ de ellas construidas en situaciones independientes, cubrirán el verdadero valor del parámetro (ver Weerahandi (1993)). Esto significa que la construcción de regiones de confianza basadas en CPGs concordantes con la definición dada, asegura el cumplimiento de la propiedad número 2 enunciada para el enfoque tradicional de la estimación por intervalo. Por lo tanto, una estimación por intervalo obtenida de esta manera a través de una CPG es lo que se denomina un **intervalo de confianza generalizado (ICG)**.

Vale destacar antes de continuar, que el resultado mencionado no será válido a menos que los espacios muestrales en cada una de las situaciones en las que se construyen los intervalos sean independientes.

2.1. Obtención de las cantidades pivotaes generalizadas.



En vista de que, con este método, la obtención de los intervalos de confianza depende de la distribución de la cantidad pivotal R , son muy raras las ocasiones en que se llega a expresiones cerradas para el cálculo de los mismos; como regla general, los límites del intervalo serán obtenidos en forma aproximada mediante métodos de simulación o integración numérica.

Es claro entonces que el punto crucial en la construcción de un intervalo de confianza generalizado, es la derivación de la CPG. Tal derivación, según expresa Weerahandi (1993) en su trabajo, no es una tarea trivial y requiere algo de intuición. En los últimos años, varios son los autores que han utilizado los resultados de Weerahandi para desarrollar procedimientos inferenciales en problemas no convencionales, tales son los casos, entre otros, de Chiang (2001) quien propuso un enfoque para construir intervalos de confianza para funciones de componentes de variancia en modelos lineales mixtos balanceados; Krishnamoorthy y Mathew (2004) quienes aplicaron ICGs para el desarrollo de límites de tolerancia unilaterales en modelos aleatorios balanceados y desbalanceados y Burdick, Borror y Montgomery (2005) quienes utilizaron esta metodología para proponer ICGs para varios indicadores de bondad de los sistemas de medición en el contexto del control de procesos. Sin embargo, y a pesar de que cada autor ha presentado numerosos ejemplos de derivación de CPGs para la construcción de ICGs en situaciones particulares, no se ha provisto aún un método general para construir estas cantidades, de modo que cada nuevo problema requiere algo de innovación y creatividad para hallar una cantidad pivotal apropiada.

En un trabajo posterior, Iyer y Patterson (2002) proponen el siguiente procedimiento general para construir CPGs para una función de varios parámetros.

Manteniendo la notación y terminología ya presentada, considérese un vector de variables aleatorias X , cuya distribución es función de un vector de parámetros de interés $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$, sea $\gamma = h(\theta_1, \theta_2, \dots, \theta_k)$ una función escalar de los parámetros θ_i para la cual se requiere un intervalo de confianza, y sea x el vector de valores observados de X .

Si para cada uno de los parámetros θ_i $i = \overline{1, k}$ existe una CPG, es decir, existen funciones $T_i = t_i(X, x, \underline{\theta})$ con $i = 1, \dots, k$ tales que $T = (T_1, T_2, \dots, T_k)$ tiene distribución conjunta libre del vector de parámetros $\underline{\theta}$; y para cada vector x fijo, el valor observado de cada una de estas funciones T_i , $t_{i_{obs}} = t_i(x, x, v)$, coincide con el parámetro θ_i correspondiente, para todo $i = \overline{1, k}$. Entonces, la función:

$$R = h(T_1(X, x, \underline{\theta}), T_2(X, x, \underline{\theta}), \dots, T_k(X, x, \underline{\theta}))$$

es una cantidad pivotal generalizada para el parámetro γ .



Esta proposición es fácilmente demostrada notando que la distribución de R queda completamente determinada por la distribución conjunta de T y el vector de valores observados x , con lo cual es libre de los parámetros desconocidos θ_i , y además el valor observado de R reproduce exactamente el parámetro de interés γ :

$$R_{obs} = h(t_1(x, x, \underline{\theta}), t_2(x, x, \underline{\theta}), \dots, t_k(x, x, \underline{\theta})) = h(\theta_1, \theta_2, \dots, \theta_k) = \gamma$$

Las definiciones introducidas por Weerahandi sobre CPGs y sus respectivos ICGs, así como la proposición debida a Iyer y Patterson serán utilizadas en este trabajo con el objetivo de desarrollar ICGs para los indicadores de bondad usualmente utilizados en los estudios de evaluación de sistemas de medición.

2.2. Cantidad pivotal generalizada para el indicador PTR.

Como ya se ha definido, el indicador PTR se obtiene como:

$$PTR = \frac{6 \sigma_{Gauge}}{LSE - LIE}$$

Llámesse $\gamma = PTR$ el parámetro de interés, el cual depende de σ_{Gauge}^2 que, a su vez, está compuesto por la suma de las componentes de variancia asociadas al error, a los operarios y a la interacción entre partes y operarios. También se ha visto que tales componentes de variancia están relacionadas funcionalmente con las esperanzas de los cuadrados medios del análisis de la variancia, de modo que en virtud de tales relaciones puede escribirse:

$$\sigma_{Gauge}^2 = \sigma^2 + \sigma_O^2 + \sigma_{OP}^2 = \frac{p(r-1)\theta_\varepsilon + (p-1)\theta_{PO} + \theta_O}{pr}$$

es decir, $\gamma = f(\theta_\varepsilon, \theta_{PO}, \theta_O)$.

Las propiedades distribucionales que se derivan de los supuestos asumidos para el modelo planteado en el análisis de la variancia, serán la base para construir las CPGs necesarias para la obtención del intervalo de confianza deseado.

A partir de estos supuestos, se deduce que $\frac{(o-1)S_O^2}{\theta_O}$, $\frac{(p-1)(o-1)S_{OP}^2}{\theta_{OP}}$ y $\frac{po(r-1)S_\varepsilon^2}{\theta_\varepsilon}$

son variables aleatorias independientes con distribución chi-cuadrado con $o-1$, $(p-1)(o-1)$ y $po(r-1)$ grados de libertad respectivamente. Por lo tanto, puede escribirse:



$$\frac{(o-1)S_o^2}{\theta_o} = W_1; \quad \frac{(p-1)(o-1)S_{OP}^2}{\theta_{OP}} = W_2; \quad \frac{po(r-1)S_\varepsilon^2}{\theta_\varepsilon} = W_3 \quad (4)$$

donde W_1 , W_2 y W_3 son variables aleatorias independientes chi-cuadrado con los grados de libertad mencionados anteriormente. Invertiendo estas expresiones convenientemente, se tiene:

$$\theta_o = \frac{(o-1)S_o^2}{W_1}, \quad \theta_{OP} = \frac{(p-1)(o-1)S_{OP}^2}{W_2}, \quad \theta_\varepsilon = \frac{po(r-1)S_\varepsilon^2}{W_3}$$

Luego, reemplazando las variables aleatorias S_i^2 por sus valores muestrales CM_i , se obtienen las cantidades:

$$R_1 = \frac{(o-1)CM_o}{W_1}, \quad R_2 = \frac{(p-1)(o-1)CM_{OP}}{W_2}, \quad R_3 = \frac{po(r-1)CM_\varepsilon}{W_3}$$

Es claro ver que estas cantidades R_i son variables aleatorias cuyas distribuciones no dependen de los parámetros desconocidos θ_o , θ_{OP} y θ_ε ; y que sus valores observados en la muestra dan como resultado los parámetros θ_o , θ_{OP} y θ_ε respectivamente. Esto significa que las funciones R_1 , R_2 y R_3 constituyen CPGs para los parámetros θ_o , θ_{OP} y θ_ε respectivamente.

A partir de estas CPGs, aplicando la proposición debida a Iyer y Patterson, puede encontrarse una CPG para el parámetro de interés γ .

Siendo $\gamma = f(\theta_o, \theta_{PO}, \theta_\varepsilon) = \frac{6}{LSE - LIE} \sqrt{\frac{p(r-1)\theta_\varepsilon + (p-1)\theta_{PO} + \theta_o}{pr}}$, entonces la CPG

correspondiente puede obtenerse aplicando la misma función f a las CPGs halladas para los parámetros θ_o , θ_{PO} y θ_ε . De esta manera se obtiene la siguiente cantidad pivotal generalizada que permitirá obtener el intervalo de confianza deseado:

$$CPG = f(R_1, R_2, R_3) = \frac{6}{LSE - LIE} \sqrt{\frac{po(r-1)^2 CM_\varepsilon}{r W_3} + \frac{(p-1)^2 (o-1) CM_{PO}}{pr W_2} + \frac{(o-1) CM_o}{pr W_1}} \quad (5)$$

El desarrollo de esta cantidad pivotal resuelve casi en su totalidad el problema de hallar el intervalo de confianza deseado, ya que una vez obtenida esta expresión, los límites del intervalo pueden obtenerse mediante un procedimiento de simulación.



Para ello, se simulará una cantidad N de valores independientes de las variables W_1 , W_2 y W_3 a partir de la distribución Chi-cuadrado. Estos valores simulados, junto con los valores observados de los cuadrados medios obtenidos en el ANOVA permitirán construir N valores de la CPG. Luego, los percentiles $\alpha/2$ y $1 - \alpha/2$ de la distribución de valores simulados de CPG serán los límites inferior y superior del intervalo de confianza deseado.

3. Un ejemplo de aplicación.

Con el objetivo de presentar un ejemplo de aplicación del procedimiento dado para la construcción del intervalo de confianza del indicador PTR, se consideró el caso extraído del Measurement System Analysis Reference Manual (2002), página 181. Allí se encuentran los resultados de un experimento realizado con el objetivo de analizar un sistema de medida para determinadas piezas, que involucra a 3 operarios, midiendo 3 veces cada una de 10 piezas, obteniendo un total de 90 ensayos aleatorizados.

El análisis de estos datos se realizó mediante un ANOVA, aplicado a un modelo de dos factores cruzados con interacción para representar las mediciones.

Las estimaciones de las componentes de variancia obtenidas fueron:

$$\hat{\sigma}_O^2 = 0,052; \quad \hat{\sigma}_P^2 = 1,089; \quad \hat{\sigma}_{OP}^2 = 0; \quad \hat{\sigma}_E^2 = 0,046$$

A partir de estas estimaciones, se obtienen las correspondientes a las componentes de repetibilidad y reproducibilidad, y con ellas la estimación de la componente de variancia debida al sistema de medición. Los valores obtenidos fueron:

$$\hat{\sigma}_{Repetibilidad}^2 = 0,046 \quad \text{y} \quad \hat{\sigma}_{Reproducibilidad}^2 = 0,052, \quad \text{con lo cual, } \hat{\sigma}_{Gauge}^2 = 0,098.$$

Con estas estimaciones puntuales, se procedió a realizar la estimación por intervalo con el recurso de intervalos de confianza generalizados. Dado que la obtención de los límites de confianza se realiza por medio de un estudio de simulación y, como ya se mencionara, su aplicación reiterada podría conducir a diferentes límites de confianza, debe optarse por un número adecuado de simulaciones que garanticen que dicha divergencia sea muy pequeña. En este caso, se decidió la realización de 100.000 simulaciones, valor recomendado en la literatura para esta clase de problemas. Esto significa que el procedimiento simulará 100.000 valores de cada una de las tres variables aleatorias chi-cuadrado especificadas en la fórmula (4), cuya combinación con los valores de los cuadrados medios obtenidos en el ANOVA, dará origen a 100.000 valores de la CPG dada por (5).

Cabe destacar, que para el ejemplo particular que se ha tomado aquí, no se dispone de valores de los límites de especificación. Por este motivo, el intervalo de confianza se construirá únicamente para la componente de variancia asociada al sistema de medición.



El correspondiente intervalo para el indicador completo, se obtendría sin más que multiplicar los extremos encontrados para σ_{Gauge} por el factor $\frac{6}{(LSE - LIE)}$.

A partir de los 100.000 valores de la CPG generados mediante la simulación, se obtienen los percentiles del 2,5% y del 97,5%, de la distribución ordenada. El intervalo de confianza resulta entonces:

$$IC_{95\%} = (0,0543 ; 1,0578)$$

Si bien aquí concluiría la búsqueda del intervalo de confianza para la componente de variancia de interés, a los autores les pareció útil agregar un estudio elemental que refleje el efecto del número de simulaciones sobre la variabilidad que se produce en los límites de confianza obtenidos al reiterar su cálculo.

El ensayo consiste entonces en ejecutar tres estudios de simulación con 10.000, 50.000 y 100.000 simulaciones, repitiendo cada uno, 30.000 veces. Cada uno de los 30.000 estudios dará como resultado un intervalo de confianza generalizado, con lo cual, al final del procedimiento se tendrán 3 grupos de 30.000 intervalos de confianza, uno para cada una de las 3 cantidades de simulaciones fijadas.

Sobre estos resultados, se calculó la media, desvío estándar, mínimo y máximo de los límites obtenidos, resultados que se presentan en la tabla 2.

Como se esperaba, se observa una disminución en la variabilidad de los límites de confianza encontrados, a medida que aumenta el número de simulaciones. Puede destacarse que la variabilidad es mucho mayor en el límite superior que en el inferior en las tres situaciones.

Se menciona también que parece muy adecuada la elección de 100000 simulaciones, ya que la diferencia entre el máximo y el mínimo encontrados para el límite inferior se percibe recién en el cuarto dígito a la derecha de la coma decimal lo que significa que la diferencia entre ambos es inferior a 10000^{-1} . La divergencia en entre mínimo y máximo del límite superior de confianza es algo más pronunciada.

Tabla 2: Medidas resumen para los límites de confianza según el número de determinaciones.

	Medidas resumen	Número de determinaciones		
		10.000	50.000	100.000
Límite inferior	Media	0.0545725	0.0545715	0.0545723
	Desvío Estándar	1.0174 E-7	2.0068 E-8	1.0338 E-8



	Mínimo	0.053379	0.0540418	0.0542403
	Máximo	0.0558522	0.0551226	0.0548822
Límite superior	Media	1.0673735	1.0678731	1.0675585
	Desvío estándar	0.0021255	0.0004149	0.0002141
	Mínimo	0.9108099	0.9874008	1.0178823
	Máximo	1.2971066	1.1541273	1.1132967

Si bien estos resultados corresponden a un ensayo realizado sobre un ejemplo, los mismos constituyen un punto de partida para continuar investigando en esta línea de trabajo que parece ser muy promisorio. Tanto la identificación de la cantidad óptima de simulaciones necesarias para lograr resultados confiables, como en la influencia de muchos otros factores sobre el comportamiento de estos intervalos de confianza, como puede ser, entre otros, la capacidad real del sistema de medición que se analiza son aspectos que merecen una consideración más detallada. Constituye también una importante línea de investigación, la aplicación de esta metodología en experimentos con más factores y con modelos más complejos.

4. Bibliografía.

AUTOMOTIVE INDUSTRY ACTION GROUP, Measurement System Analysis, 3rd edition, AIAG, Detroit, 2002.

R. K. BURDICK, C. M. BORROR y D. C. MONTGOMERY, Design and Analysis of Gauge R&R Studies, ASA-SIAM, Philadelphia, 2005.



A. K. CHIANG, A simple method for constructing confidence intervals for functions of variance components, *Technometrics*, 43 (2001), pp. 356-367.

F. A. GRAYBILL y C. M. WANG, Confidence Intervals on nonnegative linear combinations of variances, *Journal of the American Statistical Association*, 75 (1980), pp. 869-873.

H. k. IYER y P. PATTERSON, A recipe for Constructing Generalized Pivotal Quantities and Generalized Confidence Intervals. Technical report 2002-10, Department of Statistics, Colorado state University, Fort Collins, 2002. También disponible en: http://www.stat.colostate.edu/research/2002_10.pdf.

D. C. MONTGOMERY y G. C. RUNGER, Gauge Capability and Designed Experiments. Part I: Basic Methods. *Quality Engineering*, 6 (1993/1994), pp. 115-135.

S. WEERAHANDI, Generalized Confidence Intervals, *Journal of the American Statistical Association*, 88 (1993), pp. 899-905. (Corrections: 89 (1994), p.726)