

PRUEBAS DE HIPÓTESIS EN EL MODELO ESTRUCTURAL DE CALIBRACIÓN COMPARATIVA

Olga Lidia Solano Dávila¹

RESUMEN.- En el presente trabajo estudiamos inferencia estadística en el modelo de calibración comparativa que fue considerado inicialmente por Barnett [1], específicamente mostraremos la utilidad de las pruebas de hipótesis en el modelo estructural de calibración comparativa.

Utilizando el modelo presentado por Theobald y Mallison [14] proponemos una prueba asintótica usando la estadística de Wald, para probar hipótesis de interés, para lo cual, mostramos los estimadores de máxima verosimilitud y los resultados obtenidos sobre la matriz observada y esperada de Fisher.

Presentamos una ilustración de la metodología propuesta utilizando un conjunto de datos tomados de la literatura estadística.

PALABRAS CLAVE.- Modelo estructural de calibración comparativa. Matriz de información de Fisher. Estadística de Wald. Máxima verosimilitud.

HYPOTHESIS TESTING ON STRUCTURAL COMPARATIVE CALIBRATION MODEL

ABSTRACT.- In this work study statistic inference on comparative calibration model introduced for Barnett [1]. Consider hypothesis testing in structural comparative calibration model presented in Theobald and Mallison [4].

We study hypothesis testing using the model presented for Theobald and Mallison [14]. Estimates of maximum likelihood are considered and asymptotic test is proposed based in the Wald statistic. Fisher information matrix is presented.

An illustration of the methodology is presented with a data set taken from the literature

KEYWORDS.- Structural comparative calibration model. Fisher information matrix. Wald statistic. Maximum likelihood.

¹ Universidad Nacional Mayor de San Marcos. Facultad de Ciencias Matemáticas. e-mail: osolanod@unmsm.edu.pe.

1. Introducción

El objetivo de este trabajo es estudiar inferencia estadística en el modelo de calibración comparativa. El interés de este modelo es comparar métodos o instrumentos de medición de una cantidad desconocida x en un grupo de n unidades experimentales o muestrales, suponiendo que las mediciones de los instrumentos están distribuidos de acuerdo a una distribución normal multivariante.

En la sección 2 presentamos el modelo de calibración comparativa considerando p instrumentos de medición basado en el modelo propuesto por Barnett [1] y Williams [16], luego consideramos tres métodos de medición y mostramos el modelo estructural de regresión y el modelo estructural de calibración comparativa propuesto por Theobald y Mallison [14], utilizamos este modelo para obtener la matriz observada y esperada de Fisher, las cuales son presentadas en la sección 3 y 4 respectivamente. La estadística de Wald es propuesta para probar hipótesis de interés, esto es presentado en la sección 5. En la sección 6 mostramos una ilustración de la metodología propuesta con un conjunto de datos presentados en la literatura. La sección 7 está dedicada a nuestras conclusiones y algunas consideraciones para estudios posteriores.

2. Modelos de Calibración Comparativa

En esta sección presentamos el modelo de calibración comparativa considerando p instrumentos de medición, luego mostramos el modelo estructural de regresión de calibración, Barnett [1] y el modelo estructural de calibración comparativa propuesto por Theobald y Mallison [3] para tres instrumentos de medición.

2.1 El Modelo de Calibración Comparativa

Supongamos que tenemos p instrumentos o métodos de medición para medir una característica x en un grupo de n unidades experimentales o muestrales. Consideremos

x_j : el verdadero valor de la característica de interés en la unidad j .

y_{ij} : medida dada por el instrumento i para la característica de la unidad j -ésima,

$$i = 1, \dots, p, \quad j = 1, \dots, n.$$

El modelo de calibración comparativa frecuentemente utilizado en la literatura (ver Barnett [1] y Williams [16]) para este tipo de situación es definido por la relación lineal

$$\begin{cases} Y_{1j} = \alpha_1 + \beta_1 x_j + e_{1j} \\ \vdots \\ Y_{pj} = \alpha_p + \beta_p x_j + e_{pj}, \end{cases} \quad (1)$$

donde e_{ij} y x_j son independientes, $e_{ij} \sim N(0, \sigma_i^2)$ y $x_j \sim N(\mu_x, \sigma_x^2)$ $i = 1, \dots, p$, $j = 1, \dots, n$.

Los parámetros $\alpha_1, \dots, \alpha_p$ son asociados a los sesgos aditivos y β_1, \dots, β_p a los sesgos multiplicativos de las medidas obtenidas de los instrumentos.

Como es considerado en la literatura, vea por ejemplo, Barnett [1], vamos a suponer que $\alpha_1 = 0$ y $\beta_1 = 1$, esto es, el primer instrumento es considerado como el instrumento de referencia. Sobre el modelo estructural normal con $p = 3$, Barnett [1] consigue estimadores de máxima verosimilitud explícita para los parámetros del modelo.

2.2. El Modelo Estructural de Regresión de Calibración

El modelo definido en (1) por

$$\mathbf{Y}_j = \mathbf{a} + \mathbf{b}x_j + \mathbf{e}_j = \mathbf{a} + \mathbf{B}r_j, \quad j = 1, \dots, n, \quad (2)$$

donde

$$\mathbf{Y}_j = (Y_{1j}, Y_{2j}, Y_{3j})^T, \mathbf{e}_j = (e_{1j}, e_{2j}, e_{3j})^T, r_j = (x_j, \mathbf{e}_j^T)^T, \mathbf{a} = (\alpha_1, \alpha_2, \alpha_3)^T, \\ \mathbf{b} = (\beta_1, \beta_2, \beta_3)^T, \mathbf{B} = [\mathbf{b}, \mathbf{I}_3] \text{ e } \mathbf{I}_3 \text{ es la matriz identidad de dimensión } 3 \times 1, \quad j = 1, \dots, n.$$

El modelo estructural normal de calibración comparativa es obtenido suponiendo que

$$r_j \sim N_4(\boldsymbol{\eta}_x, \boldsymbol{\Omega}), \quad (3)$$

donde $\boldsymbol{\eta}_x = (\mu_x, 0, 0, 0)^T$ y $\boldsymbol{\Omega} = \text{Diag}(\sigma_x^2, \sigma_1^2, \sigma_2^2, \sigma_3^2)$. En este caso,

$$E(r_j) = \boldsymbol{\eta}_x \quad \text{y} \quad \text{Var}(r_j) = \boldsymbol{\Omega} = \begin{pmatrix} \sigma_x^2 & \mathbf{0}_3^T \\ \mathbf{0}_3 & D(\boldsymbol{\varphi}) \end{pmatrix},$$

donde $D(\boldsymbol{\varphi}) = \text{Diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)$, $\boldsymbol{\varphi} = (\sigma_1^2, \sigma_2^2, \sigma_3^2)^T$ y $\mathbf{0}_3$ es un vector nulo de dimensión 3×1 , $j = 1, \dots, n$.

De (2), (3) y por las propiedades de la distribución normal tenemos que

$$\mathbf{Y}_j \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

donde $E(\mathbf{Y}_j) = \mathbf{a} + \mathbf{b}\mu_x = \boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta})$ y $\text{Var}(\mathbf{Y}_j) = D(\boldsymbol{\varphi}) + \sigma_x^2 \mathbf{b}\mathbf{b}^T = \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$, con

$$\boldsymbol{\theta} = (\mu_x, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \sigma_x^2, \sigma_1^2, \sigma_2^2, \sigma_3^2)^T.$$

Bajo las condiciones del modelo de calibración definido en (2), $\boldsymbol{\theta}$ no es identificable (ver Bolfarine y Galea [2]), esto es, existen $\boldsymbol{\theta}_*$ y $\boldsymbol{\theta}_{**}$ tal que $F(y; \boldsymbol{\theta}_*) = F(y; \boldsymbol{\theta}_{**})$, para todo y , donde F es la distribución de \mathbf{Y}_j , $j = 1, \dots, n$. Por ejemplo, si

$$\boldsymbol{\theta}_* = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)^T = \mathbf{1}_{11}^T,$$

y

$$\boldsymbol{\theta}_{**} = (2, -2, -2, -2, 2, 2, 2, 0.25, 1, 1, 1)^T,$$

entonces

$$E_{\boldsymbol{\theta}_*}(\mathbf{Y}_j) = E_{\boldsymbol{\theta}_{**}}(\mathbf{Y}_j) = (2, 2, 2)^T$$

y

$$\text{Var}_{\theta_x}(\mathbf{Y}_j) = \text{Var}_{\theta_x}(\mathbf{Y}_i) = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

La falta de identificabilidad implica la ausencia de estimadores consistentes, lo que tiene serias implicaciones en la teoría asintótica de estimación y pruebas de hipótesis. Sin embargo, el hecho de que θ sea identificable no implica que existe un estimador consistente para θ . Un ejemplo interesante puede ser visto en Grabielsen [7].

Para solucionar el problema de identificabilidad es necesario colocar restricciones sobre el parámetro θ . En este caso, Barnett [1] asume que existe un método de medición, que él llama instrumento de referencia, que mide x sin sesgo. Específicamente, suponiendo que el instrumento de referencia sea el primero, Barnett [1] asume que

$$\alpha_1 = 0 \quad \text{y} \quad \beta_1 = 1. \quad (4)$$

Así el modelo definido en (2) puede ser escrito por

$$\mathbf{Y}_j = \mathbf{a} + \mathbf{b}x_j + \mathbf{e}_j = \mathbf{a} + \mathbf{B}r_j, \quad (5)$$

donde $\mathbf{a} = (0, \alpha^T)^T$, $\mathbf{b} = (1, \beta^T)^T$, con $\alpha = (\alpha_2, \alpha_3)^T$ y $\beta = (\beta_2, \beta_3)^T$.

Sobre la condición de identificabilidad dada en (4), el vector de parámetros tiene la forma de

$$\theta = (\mu_x, \alpha^T, \beta^T, \sigma_x^2, \varphi^T)^T.$$

2.2.1. Confiabilidad

El problema de comparar instrumentos o métodos de medición puede ser estudiado haciendo inferencias sobre los sesgos $\alpha_2, \alpha_3, \beta_2$ y β_3 de los instrumentos 2 y 3 comparándolos con el instrumento de referencia.

Definición 2.1: La confiabilidad del i -ésimo instrumento, denotado por ρ_i , se define por:

$$\rho_i = \frac{\beta_i^2 \sigma_x^2}{\beta_i^2 \sigma_x^2 + \sigma_i^2}, \quad i = 1, 2, 3. \quad (6)$$

El coeficiente ρ_i es una medida similar al coeficiente de determinación usado en regresión lineal y puede ser interpretado como la proporción de la variación observada «explicada» por el i -ésimo método de medición y $0 \leq \rho_i \leq 1$, $i = 1, 2, 3$.

Una hipótesis inicial de interés, usando el modelo (5) podría ser:

$$\mathbf{H}_{01} : \alpha_2 = \alpha_3 = 0, \quad \beta_2 = \beta_3 = 1, \quad \rho_1 = \rho_2 = \rho_3.$$

Esto es, estamos interesados en probar que los instrumentos miden la característica x sin sesgo y con igual confiabilidad.

Otra hipótesis que puede ser probada es:

$$H_{02} : \alpha_2 = \alpha_3 = 0, \beta_2 = \beta_3 = 1.$$

Aquí queremos probar si los instrumentos miden x sin sesgo.

Para probar si los instrumentos son igualmente confiables, la hipótesis es

$$H_{03} : \rho_1 = \rho_2 = \rho_3.$$

2.3. EL MODELO ESTRUCTURAL DE CALIBRACIÓN COMPARATIVA

Theobald y Mallison [14] usan una reparametrización del modelo (5), en términos de análisis de factores. Específicamente sea

$$x_j = \mu_x + \sigma_x F_j \quad j = 1, \dots, n \quad (7)$$

donde $F_j \sim N(0,1)$. Reemplazando (7) en (5) tenemos que

$$\begin{aligned} Y_j &= \mathbf{a} + \mathbf{b}(\mu_x + \sigma_x F_j) + e_j \\ &= \underbrace{\mathbf{a} + \mathbf{b}\mu_x}_{\boldsymbol{\mu}} + \underbrace{\sigma_x \mathbf{b} F_j}_{\boldsymbol{\lambda}} + e_j \\ &= \boldsymbol{\mu} + \boldsymbol{\lambda} F_j + e_j \\ &= \boldsymbol{\mu} + \mathbf{L} v_j \end{aligned} \quad (8)$$

donde $\mathbf{Y}_j = (Y_{1j}, \dots, Y_{3j})^T$, $\mathbf{e}_j = (e_{1j}, \dots, e_{3j})$, $\mathbf{v}_j = (F_j, e_j^T)^T$, $\boldsymbol{\mu} = \mathbf{a} + \mathbf{b}\mu_x$, $\boldsymbol{\lambda} = \sigma_x \mathbf{b}$, $\mathbf{L} = [\mathbf{b}, \mathbf{I}_3]$ con $\mathbf{a} = (0, \alpha_2, \alpha_3)$ y $\mathbf{b} = (1, \beta_2, \beta_3)^T$, $j = 1, \dots, n$.

Asumiendo normalidad $Y_j \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, donde $\boldsymbol{\mu} = \mathbf{a} + \mathbf{b}\mu_x$ y $\boldsymbol{\Sigma} = \boldsymbol{\lambda}\boldsymbol{\lambda}^T + D(\boldsymbol{\varphi})$, con $D(\boldsymbol{\varphi}) = \text{Diag}(\boldsymbol{\varphi})$, $\boldsymbol{\varphi} = (\sigma_1^2, \sigma_2^2, \sigma_3^2)^T$.

Sea el parámetro de interés $\boldsymbol{\theta} = (\boldsymbol{\mu}^T, \boldsymbol{\lambda}^T, \boldsymbol{\varphi}^T)^T$, entonces los estimadores de máxima verosimilitud (EMV) de $\boldsymbol{\mu}$, $\boldsymbol{\lambda}$ y $\boldsymbol{\varphi}$ son dados por $\hat{\boldsymbol{\mu}} = \bar{\mathbf{Y}}$, $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3)^T$, $\hat{\boldsymbol{\varphi}} = (\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_3^2)^T$ con

$$\hat{\lambda}_1 = \sqrt{\frac{S_{12} S_{13}}{S_{23}}}, \quad \hat{\lambda}_2 = \hat{\lambda}_1 \frac{S_{23}}{S_{13}}, \quad \hat{\lambda}_3 = \hat{\lambda}_1 \frac{S_{23}}{S_{12}}, \quad \hat{\sigma}_i = S_{ii} - \hat{\lambda}_i^2, \quad i = 1, 2, 3.$$

Donde $\bar{\mathbf{Y}}$ es la media muestral de los Y_j y S_{ij} son los elementos de la matriz de covarianza muestral. Note que el estimador de σ_i^2 puede asumir valores negativos, la solución de este problema se realiza utilizando el estimador basado en la verosimilitud irrestricta propuesto por Carter [3].

Observación

Sobre el modelo de calibración comparativa y el modelo presentado por Theobald y Mallison [15], las definiciones de confiabilidad del i -ésimo instrumento son dadas por

$$\rho_i = \frac{\lambda_i^2}{\lambda_i^2 + \sigma_i^2},$$

Además las hipótesis de interés H_{01} , H_{02} y H_{03} definidas en la sección (2) son equivalentes a

$$\begin{aligned} H_{01} : \mu_1 = \mu_2 = \mu_3, \quad \lambda_1 = \lambda_2 = \lambda_3, \quad \rho_1 = \rho_2 = \rho_3, \\ H_{02} : \mu_1 = \mu_2 = \mu_3, \quad \lambda_1 = \lambda_2 = \lambda_3 \quad \text{y} \\ H_{03} : \rho_1 = \rho_2 = \rho_3, \end{aligned}$$

respectivamente.

Sobre el modelo presentado por Theobald y Mallison [14], los estimadores de máxima verosimilitud de ρ_1 , ρ_2 y ρ_3 son dados por

$$\hat{\rho}_i = \frac{\hat{\lambda}_i^2}{\hat{\lambda}_i^2 + \hat{\sigma}_i^2}, \quad i = 1, 2, 3.$$

3. La Matriz de Información Observada de Fisher

En esta sección vamos a presentar la matriz de información observada de Fischer en el modelo de calibración propuesto por Theobald y Mallison [14], que es una reparametrización del modelo propuesto por Barnett [1].

La matriz de información observada nos servirá para obtener la matriz de información esperada.

Sea el vector de parámetro $\theta = (\mu^T, \lambda^T, \varphi^T)$ para el modelo estructural. De (8) la función de log-verosimilitud para el modelo es de la forma

$$L(\theta) = \sum_{j=1}^n \ell_j(\theta) \quad (9)$$

donde

$$\ell_j(\theta) = -\frac{3}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma| - \frac{1}{2} \|T_j\|^2,$$

con $\|T_j\|^2 = (Y_j - \mu)^T \Sigma^{-1} (Y_j - \mu)$, $|\Sigma| = c|D(\varphi)|$, $\Sigma^{-1} = D^{-1}(\varphi) - c^{-1}M$ y

$$c = 1 + \lambda^T D^{-1}(\varphi) \lambda, \quad M = D^{-1}(\varphi) \lambda \lambda^T D^{-1}(\varphi), \quad j = 1, \dots, n.$$

La matriz de derivadas de segundo orden son respecto a θ es dada en forma triangular por

$$I_F(\theta) = -L = -\frac{\partial L(\theta)}{\partial \theta \partial \theta^T} \Big|_{\theta=\hat{\theta}} = - \begin{bmatrix} L_{\mu\mu} & L_{\mu\lambda} & L_{\mu\phi} \\ & L_{\lambda\lambda} & L_{\lambda\phi} \\ \text{Simetría} & & L_{\phi\phi} \end{bmatrix} \quad (10)$$

donde $\hat{\theta}$ es el estimador de máxima verosimilitud de θ . Los elementos de la matriz son presentados en Solano [13].

4. La Matriz de Información Esperada de Fischer

En esta sección presentamos la matriz de información esperada de Fisher, que es muy importante para tratar hipótesis de interés, por ejemplo, si estamos interesados en probar si los tres instrumentos miden la características x sin sesgo y con igual confiabilidad, o probar si los instrumentos miden x sin sesgo o si los instrumentos son igualmente confiables.

La matriz de información esperada que denotamos por $J_F(\theta)$, es definido por

$$J_F(\theta) = E\left(-\frac{\partial^2 L}{\partial^2 \theta \theta^T}\right) = \sum_{j=1}^n E\left(\frac{\partial^2 l_j}{\partial \theta \theta^T}\right).$$

Teorema 4.1: Sobre el modelo de calibración estructural definido en (8). La matriz de información esperada es dada por, (ver Solano [13]).

$$J_F(\theta) = \begin{pmatrix} J_{\mu\mu} & 0 & 0 \\ 0 & J_{\lambda\lambda} & J_{\lambda\phi} \\ 0 & J_{\phi\lambda} & J_{\phi\phi} \end{pmatrix},$$

donde

$$J_{\mu\mu} = -\sum_{j=1}^n \Sigma^{-1},$$

$$J_{\lambda\lambda} = -\sum_{j=1}^n \left\{ (1 - c^{-1}) D^{-1}(\varphi) + c^{-1} (2c^{-1} - 1) M \right\},$$

$$J_{\lambda\phi} = -\sum_{j=1}^n \left\{ c^{-1} D(\lambda) D^{-2}(\varphi) - c^{-2} M D^{-1}(\varphi) D(\lambda) \right\},$$

$$J_{\phi\lambda} = (J_{\lambda\phi})^T \quad \text{y}$$

$$J_{\phi\phi} = -\sum_{j=1}^n \left\{ \frac{D^{-2}(\varphi)}{2} - c^{-1} D^2(\lambda) D^{-3}(\varphi) + \frac{c^{-2}}{2} D(\lambda) D^{-1}(\varphi) M D^{-1}(\varphi) D(\lambda) \right\}.$$

5. Prueba de Hipótesis

Utilizaremos el estadístico de Wald para probar las hipótesis H_{01} , H_{02} y H_{03} , en la versión del modelo estructural propuesto por Theobald y Mallison [14]. Para algunas propiedades de esta prueba se puede ver Cox y Hinkley [5].

Sea $\theta_* = (\mu^T, \lambda^T, \rho^T)^T$, donde $\rho = (\rho_1, \rho_2, \rho_3)^T$ (que es función de λ y φ). Entonces por el principio de invarianza del EMV, el EMV de θ_* es dado por $\hat{\theta}_* = (\hat{\mu}^T, \hat{\lambda}^T, \hat{\rho}^T)^T$, donde $\hat{\mu}$, $\hat{\lambda}$ y $\hat{\rho}$ son definidos en la sección (2).

Teorema 5.1: Sobre el modelo de calibración definido en (8), tenemos que

- a) $\hat{\theta}_*$ es un estimador consistente de θ_* y
 b) $\sqrt{n}(\hat{\theta}_* - \theta_*) \xrightarrow{D} N_9(0, \Omega(\theta_*))$,

donde

$$\Omega(\theta_*) = GJ_F^{-1}(\theta) G^T \quad (11)$$

con $G = \begin{pmatrix} I_3 & 0 & 0 \\ 0 & I_3 & 0 \\ 0 & G_1 & G_2 \end{pmatrix}$ y $G_1 = 2D^{-1}(\lambda)D(\rho)[I_3 - D(\rho)]$ y $G_2 = -D^{-2}(\lambda)D^2(\rho)$. La

prueba de este teorema puede verse en Solano [13].

Corolario 5.1: Bajo el modelo de calibración definido en (8)

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{D} N_3(0, QJ_F^{-1}(\theta) Q^T),$$

donde $Q = (0, G_1, G_2)$, con G_1 y G_2 definidos en el Teorema 5.1.

Para probar si los instrumentos miden la característica sin sesgo y con la misma confiabilidad, la hipótesis podría ser

$$H_{01} : \mu_1 = \mu_2 = \mu_3, \quad \lambda_1 = \lambda_2 = \lambda_3, \quad \rho_1 = \rho_2 = \rho_3.$$

Note que H_{01} puede ser escrita como $H_{01} : C\theta_* = q_0$,

donde

$$C = \begin{pmatrix} C_1 & 0 & 0 \\ 0 & C_1 & 0 \\ 0 & 0 & C_1 \end{pmatrix}$$

es una matriz de 6×9 de rango 6, con

$$C_1 = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \quad (12)$$

con $q_0 = 0$ vector nulo de dimensión 6.

Luego el estadístico de Wald, (ver Sen [11]), para probar H_{01} , es dado por

$$W_{01} = n(C\hat{\theta}_* - q_0)^T (C\hat{\Omega}(\theta_*)C^T)^{-1} (C\hat{\theta}_* - q_0),$$

que converge en distribución a $\chi_{(6)}^2$ bajo H_{01} , $\hat{\Omega}(\theta_*)$ es una matriz dada en (11) evaluada en el EMV de θ_* , $\hat{\theta}_*$.

Por lo tanto H_{01} es rechazada a un nivel de significancia de α si

$$W_{01} > \chi_{\alpha}^2(6),$$

donde $\chi_{\alpha}^2(6)$ es el punto crítico de la distribución $\chi_{(6)}^2$ correspondiente a un nivel de significación de α .

Para probar si los instrumentos miden sin sesgo, tenemos que probar la hipótesis

$H_{02} : \mu_1 = \mu_2 = \mu_3, \quad \lambda_1 = \lambda_2 = \lambda_3$. También, en este caso H_{02} puede ser escrita como

$H_{02} : C\theta_* = q_0$, donde, ahora,

$$C = \begin{pmatrix} C_1 & 0 & 0 \\ 0 & C_1 & 0 \end{pmatrix}$$

es una matriz 4×9 de rango 4 con C_1 como en (12) y $q_0 = 0$ es un vector nulo de dimensión 4.

Así, el estadístico de Wald para probar H_{02} , denotada por W_{02} , es dado por

$$W_{02} = n(C\hat{\theta}_*)^T (C\hat{\Omega}(\theta_*)C^T)^{-1} (C\hat{\theta}_*),$$

que converge en distribución a $\chi_{(4)}^2$ bajo H_{02} .

Por lo tanto, H_{02} es rechazada a un nivel de significancia α si

$$W_{02} > \chi_{\alpha}^2(4),$$

donde $\chi_{\alpha}^2(4)$ es el punto crítico de la distribución $\chi_{(4)}^2$ correspondiente a un nivel de significación α .

Finalmente, para probar si los instrumentos son igualmente confiables, la hipótesis de interés es

$H_{03} : \rho_1 = \rho_2 = \rho_3$ o $H_{03} : C_1\rho = 0$, con C_1 como en (12). Así el estadístico de Wald, denotado

por W_3 , es obtenido del corolario 5.1 y es dado por

$$W_3 = n(C_1\rho)^T (C_1QJ_F^{-1}(\hat{\theta})Q^TC_1^T)^{-1} (C_1\rho),$$

que converge en distribución a $\chi_{(2)}^2$ bajo H_{03} .

Por lo tanto, se rechaza H_{03} a un nivel de significancia α si

$$W_{03} > \chi_{\alpha}^2(2),$$

donde $\chi_{\alpha}^2(2)$ es el punto crítico de la distribución $\chi_{(2)}^2$ correspondiente a un nivel de significación α .

6. Aplicación

El conjunto de datos fue obtenido por investigadores de la Escuela Paulista de Medicina y se refiere a un estudio para evaluar la calidad de tres métodos de medición usados para estimar el volumen del testículo humano. Los investigadores obtuvieron una muestra de 42 adolescentes entre 11 y 17 años del mismo hospital. Los tres métodos fueron operados sobre la mismas condiciones, o sea, todos los pacientes en la posición vertical, el mismo operador para los tres métodos y la medición del testículo derecho.

Estos métodos fueron los siguientes:

Método 1: método ya existente en el mercado y de alta precisión, que consiste de un sistema de ultra-sonido para obtener las dimensiones del testículo: largo (L_1), ancho (C_1) y profundidad (P) y a partir de ahí calcular el volumen (V_1) a través de la fórmula $V_1 = \pi/6 * L_1 * C_1 * P$.

Método 2: método gráfico el cual consiste de 6 anillos elípticos que representan volúmenes de 2 ml, 5 ml, 10 ml y 20 ml y a través de esos anillos se estima en forma visual el volumen del testículo. En el caso de estar entre dos anillos, por ejemplo entre 10 ml y 15 ml, el volumen estimado pasa a ser la medida aritmética entre estos dos, o sea 12,5 ml.

Método 3: Este método obtiene el ancho (C_3) y el largo (L_3) del testículo a través de un calibrador específico y el volumen es calculado por la fórmula $V_3 = \pi/6 * L_3 * C_3^2$.

Sea Y_{ij} la medida del volumen del testículo en ml dada por el método i para el paciente j , con $i = 1, 2, 3$ y $j = 1, \dots, 42$. El conjunto de observaciones $\{Y_{ij}\}$ es presentado en Solano [13]. Por ser un método conocido y comúnmente utilizado, adoptaremos el método uno, como el de referencia. Así de acuerdo como se explicó en la sección 2, asumimos que $\alpha_1 = 0$ y $\beta_1 = 1$.

De acuerdo con la sección 2, calculamos los estimadores de máxima verosimilitud para el conjunto de datos,

$$\begin{aligned} \mu_1 &= 10,081, & \mu_2 &= 9,012 & \mu_3 &= 9,562 \\ \lambda_1 &= 4,737, & \lambda_2 &= 4,154 & \lambda_3 &= 4,495 \\ \phi_1 &= 1,149, & \phi_2 &= 1,149 & \phi_3 &= 1,526 \\ \rho_1 &= 0,951, & \rho_2 &= 0,923 & \rho_3 &= 0,930. \end{aligned}$$

La obtención de la matriz de Fischer es de mucha utilidad para obtener la distribución asintótica de los estimadores. Por lo tanto de acuerdo con el Teorema 4.1, la matriz de información esperada de Fisher es dada por

$$J_F(\hat{\theta}) = \begin{pmatrix} \hat{J}_{\mu\mu} & 0 & 0 \\ 0 & \hat{J}_{\lambda\lambda} & \hat{J}_{\lambda\phi} \\ 0 & \hat{J}_{\phi\lambda} & \hat{J}_{\phi\phi} \end{pmatrix},$$

$$\begin{aligned} \hat{J}_{\mu\mu} &= \begin{pmatrix} 0,498 & -0,258 & 0,266 \\ -0,258 & -0,509 & -0,185 \\ -0,266 & -0,185 & 0,465 \end{pmatrix}, & \hat{J}_{\lambda\lambda} &= \begin{pmatrix} 0,495 & -0,247 & -0,254 \\ -0,247 & 0,502 & -0,177 \\ -0,254 & -0,177 & 0,459 \end{pmatrix}, \\ \hat{J}_{\lambda\phi} &= \begin{pmatrix} 0,045 & -0,016 & -0,017 \\ -0,023 & 0,032 & -0,012 \\ -0,024 & -0,012 & 0,030 \end{pmatrix}, & \hat{J}_{\phi\lambda} &= \begin{pmatrix} 0,045 & -0,023 & -0,024 \\ -0,016 & 0,032 & -0,012 \\ -0,017 & -0,012 & 0,030 \end{pmatrix}, \\ \hat{J}_{\phi\phi} &= \begin{pmatrix} 0,124 & 0,033 & 0,035 \\ 0,033 & 0,130 & 0,017 \\ 0,035 & 0,017 & 0,108 \end{pmatrix}. \end{aligned}$$

La inversa de la matriz de Información Esperada de Fisher es

$$J_F^{-1} = \begin{pmatrix} 23,588 & 19,678 & 21,293 & 0 & 0 & 0 & 0 & 0 & 0 \\ 19,678 & 18,708 & 18,672 & 0 & 0 & 0 & 0 & 0 & 0 \\ 21,293 & 18,672 & 21,731 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 12,450 & 9,810 & 10,610 & -0,772 & 0,242 & 0,325 \\ 0 & 0 & 0 & 9,810 & 10,143 & 9,328 & 0,276 & -0,521 & 0,076 \\ 0 & 0 & 0 & 10,610 & 9,328 & 11,702 & 0,343 & 0,071 & -0,660 \\ 0 & 0 & 0 & -0,772 & 0,276 & 0,343 & 9,957 & -2,289 & -3,082 \\ 0 & 0 & 0 & 0,242 & -0,521 & 0,071 & -2,289 & 8,543 & -0,635 \\ 0 & 0 & 0 & 0,325 & 0,660 & -0,660 & -3,082 & -0,635 & 10,589 \end{pmatrix},$$

Para probar si los tres instrumentos miden la característica sin sesgo y con la misma confiabilidad, la hipótesis es:

$$H_{01} : \mu_1 = \mu_2 = \mu_3, \quad \lambda_1 = \lambda_2 = \lambda_3, \quad \rho_1 = \rho_2 = \rho_3.$$

Para realizar esa prueba de hipótesis, fue calculada la matriz $\hat{\Omega}(\theta_*)$ que es la matriz dada en (11) y evaluada en el EMV de θ_* , $\hat{\theta}_*$.

$$\hat{\Omega}(\hat{\theta}_*) = \begin{pmatrix} 23,588 & 19,678 & 21,293 & 0 & 0 & 0 & 0 & 0 & 0 \\ 19,678 & 18,708 & 18,672 & 0 & 0 & 0 & 0 & 0 & 0 \\ 21,293 & 18,672 & 21,731 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 12,450 & 9,810 & 10,610 & 0,276 & 0,324 & 0,293 \\ 0 & 0 & 0 & 9,810 & 10,143 & 9,328 & 0,182 & 0,373 & 0,267 \\ 0 & 0 & 0 & 10,610 & 9,328 & 11,702 & 0,195 & 0,316 & 0,367 \\ 0 & 0 & 0 & 0,276 & 0,182 & 0,195 & 0,022 & 0,001 & 0,000 \\ 0 & 0 & 0 & 0,324 & 0,373 & 0,316 & 0,001 & 0,035 & 0,008 \\ 0 & 0 & 0 & 0,293 & 0,267 & 0,367 & 0,000 & 0,008 & 0,031 \end{pmatrix},$$

$$\hat{\theta}_* = (10,081 \ 9,012 \ 9,562 \ 4,737 \ 4,154 \ 4,495 \ 0,951 \ 0,923 \ 0,930)^T,$$

Usando estos resultados y la estadística de Wald, podemos probar la hipótesis que acabamos de mencionar. Para $n = 42$, tenemos que $W_{01} = 30,997$, comparando con $\chi^2_{(6)} = 12,59$, se rechaza la hipótesis nula, con un nivel de significancia del 5%, y concluimos que los instrumentos miden con sesgo el volumen del testículo derecho y/o no son igualmente confiables.

Para obtener información respecto al rechazo de la hipótesis H_{01} , vamos a probar en forma separada si los instrumentos miden con sesgo y si tienen la misma confiabilidad.

Para probar si los tres instrumentos miden la característica sin sesgo la hipótesis es,

$$H_{02} : \mu_1 = \mu_2 = \mu_3, \quad \lambda_1 = \lambda_2 = \lambda_3.$$

Usando los resultados de $\hat{\Omega}(\hat{\theta}_*)$ y la estadística de Wald, podemos probar la hipótesis. Luego para $n = 42$, tenemos que $W_{02} = 31,215$, comparando con $\chi^2_{(4)} = 9,49$, se rechaza la hipótesis nula, a un nivel de significancia del 5%, esto es, los métodos miden con sesgo el volumen del testículo derecho. No podemos decir que el método i nos da la medida exacta del volumen del testículo derecho.

Para probar si los tres instrumentos son igualmente confiables, la hipótesis es,

$$H_{03} : \rho_1 = \rho_2 = \rho_3.$$

Usando estos resultados y la estadística de Wald, tenemos que $W_{03} = 0,655$, comparando con $\chi^2_{(2)} = 5,99$, no se rechaza la hipótesis nula, a un nivel de significancia del 5%, esto es, los métodos son igualmente confiables.

De los resultados presentados anteriormente, concluimos que los dos métodos propuestos son tan buenos como el método de ultra-sonido, o sea son igualmente confiables para la medición del volumen del testículo derecho. El uso de diferentes tamaños de los anillos para el método gráfico permite una rápida comparación visual para los testículos. El tiempo de manipulación es reducido comparado con los otros métodos de medición, decreciendo la natural turbación de los adolescentes.

7. Discusión y Comentarios

En el presente trabajo estudiamos pruebas de hipótesis estadística sobre el modelo estructural de calibración comparativa, propuesto por Theobald y Mallison [14] considerando tres métodos o instrumentos de medición. Los estimadores de máxima verosimilitud son mostrados, asimismo los resultados de la matriz de información observada y esperada de Fisher y una prueba asintótica es propuesta utilizando el estadístico de Wald.

De los resultados de la aplicación concluimos que los dos instrumentos de medición propuestos (método 1 y 2) son tan buenos como el método de ultra-sonido, esto es, son igualmente confiables para la medición del volumen del testículo derecho, a un nivel de significancia del 5%.

Cuando consideramos tres métodos de medición EMV pueden ser obtenidos en forma explícita. Para más de tres instrumentos no existen expresiones explícitas para los estimadores de máxima verosimilitud de los parámetros, por lo tanto es necesario utilizar algún método numérico para obtenerlos. Bolfarine y Galea-Rojas [2] utilizaron el algoritmo EM para encontrar estos estimadores.

REFERENCIAS BIBLIOGRÁFICAS

- [1] BARNETT, V. D. *Simultaneous pairwise linear structural relationships*. Biometrics, Vol 25 pag. 129-142 (1969).
- [2] BOLFARINE, H. Y GALEA-ROJAS, M. *Maximum likelihood estimation of simultaneous pairwise linear structural relationships*. Biometrical Journal Vol 37 pag. 673-689. (1995)
- [3] CARTER, R. *Restricted maximum likelihood estimation of several bias and reliability in the comparison measuring methods*. Biometrics Vol 37, pag. 733-741 (1981).
- [4] CHIPKEVITCH, E., NISHIMURA, R. T., TU, D.G.S. Y GALEA-ROJAS, M. *Clinical measurement of testicular volume in adolescents: Comparison of the reliability of 5 methods*. Journal of Urology Vol 156, 6, pag. 2050-2053 (1996).
- [5] COX, D.R. Y HINKLEY, D. V. *Theoretical Statistics*: Chapman and Hall. London. (1974).
- [6] DUNN G. *Design and analysis of reliability: The statistical evaluation of measurement errors*. Edward Arnold, New York, (1992)
- [7] GRABIELSEN, A. Consistency and identificability. *Journal of econometrics* 8, 261-263. (1978).
- [8] GRAYBILL, F.A. *Matrices with Applications in Statistics*. Second Edition. Wadsworth. California (1983).
- [9] KELLY, G. *The influence function in the errors in variables problem*. Annals of Statistics Vol 12 pag. 87-100 (1984).
- [10] KELLY, G. *Use of the structural equations model in assessing the realibility of a new measurement technique*. Applied Statistics Vol 34 pag. 258-263. (1985)
- [11] SEN, P.K. Y SINGER, J. M. *Large Sample Methods in Statistics. An Introduction with Applications*. Chapman y Hall. New York. (1993)
- [12] SHYR, J.Y. Y GLESER, L.J. *Inference about comparative precision in linear structural relationships*. Journal of the Statistical Planning and Inference Vol 14 pag. 339-358 (1986).
- [13] SOLANO, O. *Influência Local em Modelos de Calibração Comparativa*. Tese de Mestrado. IMEC-UNICAMP. Brasil (2000).
- [14] THEOBALD, C.M. Y MALLISON, J.R. *Comparative calibration, linear structural relationship and congeneric measurements*. Biometrics Vol 34 pag. 39-45 (1978)
- [15] WEISBERG, S. *Applied Linear Regresión*. Second Edition. John Wiley & Sons, Inc. New York (1985).
- [16] WILLIAMS, E. J. *Regression methods in calibration problems*. *Bulletin of the International Statistical Institute* Vol 43, 17-28. (1969).