“Propuesta de Método de Contraste de Hipótesis y su Implementación Computacional: Una Aplicación de los Wavelets"

Henry Aquino Mosquera

Maestría en Matemática Aplicada con mención en Matemática Computacional
Universidad Nacional San Marcos

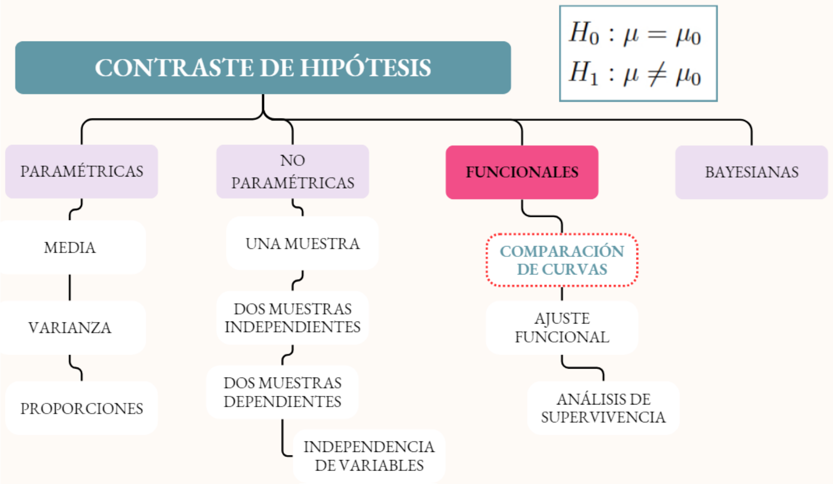
1 de julio de 2025

UNMSM

1. **Conceptos preliminares**
   - Contraste de hipótesis
   - Procesos estocásticos
2. Fundamentación del estudio
3. Análisis Multiresolución
4. Estimación
5. Contraste de hipótesis
6. Aplicación

1. Conceptos preliminares
   - Contraste de hipótesis
   - Procesos estocásticos
2. Fundamentación del estudio
3. Análisis Multiresolución
4. Estimación
5. Contraste de hipótesis
6. Aplicación

Contraste de hipótesis

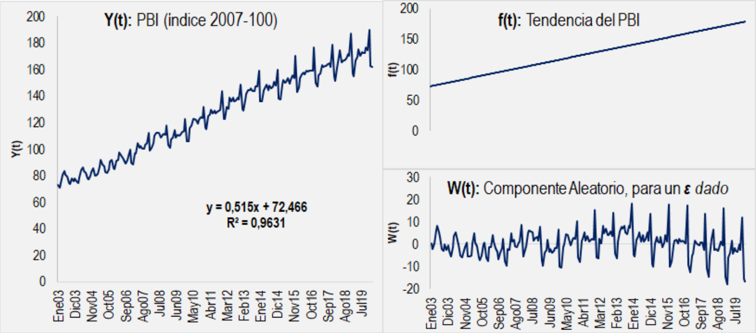


1. Conceptos preliminares
   - Contraste de hipótesis
   - Procesos estocásticos
2. Fundamentación del estudio
3. Análisis Multiresolución
4. Estimación
5. Contraste de hipótesis
6. Aplicación

Procesos estocásticos

- **Proceso estocásito:** *Variable Aleatoria (VA)* depende del "tiempo".
- **Ecuación general**: $Y(t) = f(t) + \varepsilon W(t)$, donde $f(t)$ un componente sistemático, ε la intensidad de la VA y $W(t)$ la VA.

- **Ejemplo**:



Ejemplo de variables con diferentes categorías/grupos

Variación de ingresos laborales mensuales de Lima y Callao (var %):

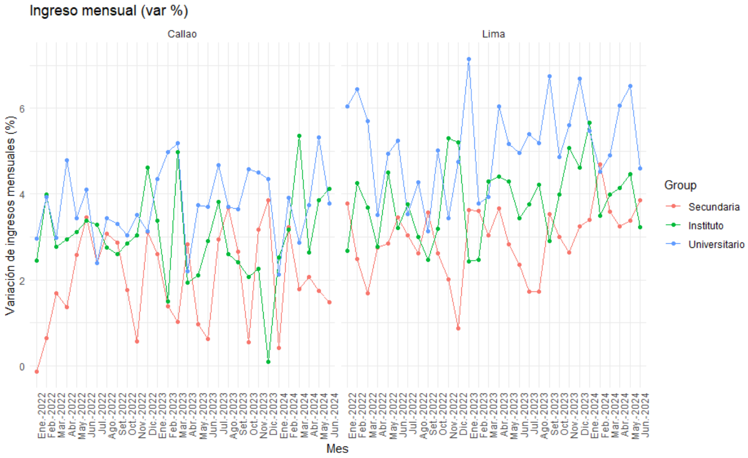


Figura 1: Comparación de ingresos según niveles de educación 2022-2024

Objetivos de la investigación

Análisis de varianza y multivariado de varianza (ANOVA y MANOVA):

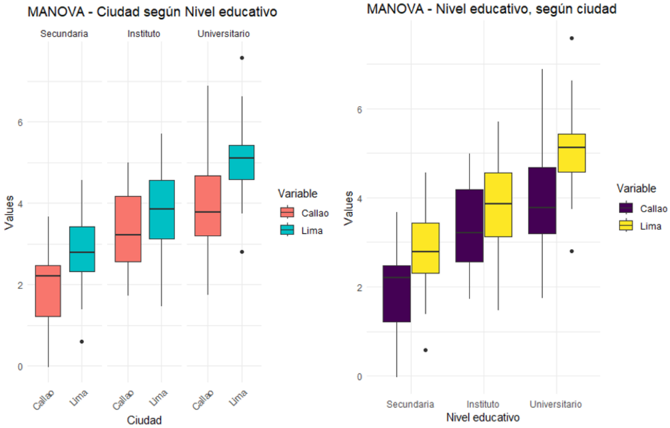


Figura 2: Resultados ilustrativos del análisis MANOVA

Ecuación ANOVA y MANOVA

- **ANOVA:** Evalúa la hipótesis nula, que asume igualdad entre las medias. Formula ANOVA:

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$

- **MANOVA:** El MANOVA extiende el ANOVA para comparar las medias de varios grupos en dos o más variables dependientes simultáneamente. Formula MANOVA:

$$F = \frac{\det(\mathbf{W})}{\det(\mathbf{B}) \cdot (N - g)}$$

Los parámetros no son controlables, ya que dependen exclusivamente de la información disponible. Esto expone el análisis a evaluaciones inadecuadas, especialmente por la presencia habitual de datos atípicos en la evidencia empírica, lo que puede llevar a conclusiones sesgadas.

Objetivos de la investigación

Objetivo general de la investigación: Propuesta de método matemático (funcional) de contraste de hipótesis para comparar curvas: test con propiedades estadísticas óptimas para un contraste de prueba de hipótesis funcional:

$$H_0 : \text{“Dos o más grupos de curvas son iguales”.}$$

$$H_1 : \text{“Otro caso”.}$$

Objetivos específicos

Objetivos específicos

- Estimar funciones que se aproximen a datos reales mediante un estimador basado en la expansión de wavelets.
- Desarrollar y validar un procedimiento de prueba utilizando coeficientes de wavelets para verificar si no existe una tendencia común entre las muestras.
- Desarrollar y validar un procedimiento de prueba utilizando coeficientes de wavelets para verificar si existe o no diferencias entre curvas.

Justificación

Justificación:

- Los parámetros utilizados en las ecuaciones incluyen valores promedios, y parámetros estadísticos determinísticos (valores fijos - no controlables), y otros datos puntuales, como el número de datos o grupos.
- La implementación de las wavelets mejorará el contraste de hipótesis:
  - Mediante el control de datos anómalos en la información analizada.
  - A través de la selección del nivel de aproximación de los datos procesados.

Idea intuitiva del AMR

El requisito fundamental del Análisis de Multiresolución (AMR) es que los espacios generados por $\mathcal{V}_j$ estén anidados de la siguiente manera:

$$\ldots \subset \mathcal{V}_{-2} \subset \mathcal{V}_{-1} \subset \mathcal{V}_0 \subset \mathcal{V}_1 \subset \mathcal{V}_2 \subset \ldots \subset L^2 \tag{1}$$

A partir de (1), podemos comenzar en cualquier $\mathcal{V}_j$, por ejemplo, en $j = 0$, y escribir

$$\mathcal{V}_{j_0} \subset \mathcal{V}_1 \subset \mathcal{V}_2 \subset \cdots \subset L^2$$

Visto de otra manera, se puede generar la siguiente expresión

$$\mathcal{V}_1 = \mathcal{V}_{j_0} \oplus \mathcal{W}_0$$

Que se extiende a

$$\mathcal{V}_2 = \mathcal{V}_{j_0} \oplus \mathcal{W}_0 \oplus \mathcal{W}_1$$

En general esto da

$$L^2 = \mathcal{V}_{j_0} \oplus \mathcal{W}_0 \oplus \mathcal{W}_1 \oplus \ldots \tag{2}$$

Sistema intuitivo

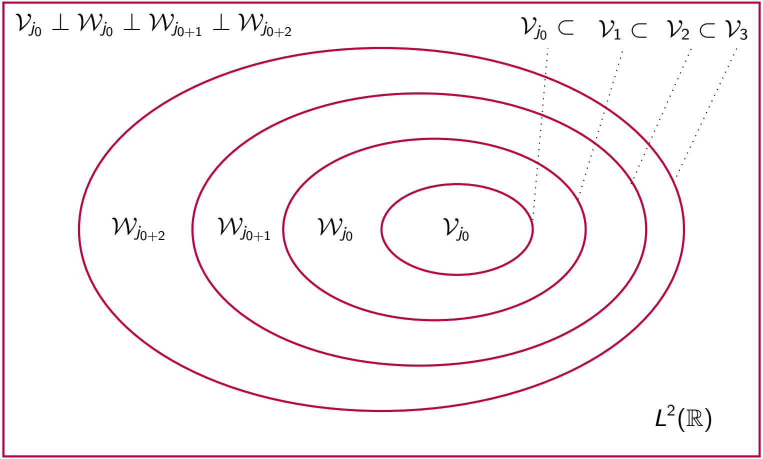


Figura 3: Función de escala y Espacio de Vectores Wavelet

1. Conceptos preliminares
2. Fundamentación del estudio
3. Análisis Multiresolución
   - Sistema intuitivo
   - Función de escala
   - Función de wavelet
4. Estimación
5. Contraste de hipótesis
6. Aplicación

Función de escala

Se genera una familia de parametrización de funciones a partir de la función de escala básica $\varphi(t) \in L^2(\mathbb{R})$, cuyo escalado y traslación genera

$$\varphi_{j,k}(t) = 2^{j/2}\varphi(2^j t - k); j, k \in \mathbb{Z}. \tag{3}$$

Para un j fijo, el espacio generado $\mathcal{V}_j$ sobre k es

$$\mathcal{V}_j = \overline{\text{Span}}\left\{\varphi_{j,k}(t) : k \in \mathbb{Z}\right\}$$

Función de wavelet

En base a la ecuación (2) se define la función wavelet

$$\psi(t) = \sum_{k} h_k \varphi(2^j t - k); \tag{4}$$

Puedo generar otro tipo de funciones

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k); \tag{5}$$

$$\mathcal{W}_j = \overline{\text{Span}}\{\psi_{jk}(t) : k \in \mathbb{Z}\}$$

para el caso de combinaciones finitas o infinitas, lo cual nos lleva

$$L^2(\mathbb{R}) = \mathcal{V}_{j_0} \oplus \mathcal{W}_{j_0} \oplus \mathcal{W}_{j_{0+1}} \oplus \dots \tag{6}$$

de ahí que

$$f(t) = \sum_{k=1}^{2^{j_0}-1} c_{j_0,k}\varphi_{j_0,k}(t) + \sum_{j \geq j_0}^{+\infty} \sum_{k=0}^{2^{j_0}-1} d_{j,k}\psi_{j,k}(t) \tag{7}$$

donde $f(t) \in L^2(\mathbb{R})$, y $c_{j_0}, d_{j,k} \in \mathbb{R}$

Secuencia de funciones

Consideremos una sucesión desconocida de funciones, que se expresa como:

$$f_1, f_2, \ldots, f_r; \quad f_i \in L^2(\mathbb{R}), \quad i = 1, 2, \ldots, r$$

En el marco del análisis de multiresolución utilizando wavelets, podemos definir un proceso que se representa mediante la siguiente ecuación:

$$Y_i(t) = \int_0^t f_i(s)\, ds + \varepsilon W_i(t), \quad i = 1, 2, \ldots, r \tag{8}$$

Donde $\varepsilon > 0$ y $W_i(t)$ denota un proceso estocástico que introduce variabilidad aleatoria, modelando así la incertidumbre inherente a la función $f_i(t)$.
Al derivar la ecuación anterior, se obtiene el (*modelo funcional - Ecuación Diferencial Estocástica*):

$$dY_i(t) = f_i(t)dt + \epsilon W_i(t)dt \tag{9}$$

Donde cada función f_i será representada de la siguiente forma (*propiedad de separación de términos deterministas y estocásticos*):

$$f_i(t) = m_0 + \mu(t) + a_i + g_i(t), i = 1, 2, \ldots, r \tag{10}$$

donde $\mu(t)$ es la función media y $g_i(t)$ la función que distingue a cada función.

Estructura funcional

Cuando los datos son observables se tiene una secuencia:

$$t_1, t_2, \ldots, t_n$$

donde

$$Y_i(t_1), Y_i(t_2), \ldots, Y_i(t_n); i = 1, 2, \ldots, r$$

Además, el tamaño de la serie para el caso discreto que permita elegir un nivel de resolución será (técnica de espejamiento):

$$J : n = 2^J$$

Aproximación de la series

Propuesta de estimador para $f(t)$

$$\widehat{f}(t) = \sum_{k} c_{j_0,k}\varphi_{j_0,k}(t) + \sum_{j \geq j_0}^{j_1} \sum \widehat{d}_{j,k}\psi_{j,k}(t) \tag{11}$$

donde

$$0 < j_0 < j_1 < J \tag{12}$$

Tener en cuenta

Metodología de estimación **dada**

- Criterio de elección de j_0, j_1
- Estimación de $\widehat{d}_{j,k}$

Por lo tanto,

- Mientras que en el universo funcional uno busca estimar los $y_i(t_n), (i \neq n)$ de cada función,
- en el universo de wavelets, se busca estimar coeficientes $d_i(j,k)$, siendo que $d_{j,k} = \langle f(t), \psi_{j,k}(t) \rangle = \int_{\mathbb{R}} f(t)\psi_{j,k}(t)dt$.

1. Conceptos preliminares
2. Fundamentación del estudio
3. Análisis Multiresolución
4. Estimación
5. **Contraste de hipótesis**
6. Aplicación

Tener en cuenta

Modelo básico del proceso estocástico

$$\frac{dY_i(t)}{dt} = \frac{df_i(t)}{dt} + \epsilon \frac{dW_i(t)}{dt}$$

Estimación de $f(t)$

$$\widehat{f}(t) = \sum_k c_{j_0,k}\varphi_{j_0,k}(t) + \sum_{j \geq j_0}^{j_1} \sum \widehat{d}_{j,k}\psi_{j,k}(t)$$

Los coeficientes son:

$$d_{j,k} = \langle f(t), \psi_{j,k}(t) \rangle = \int_{\mathbb{R}} f(t)\psi_{j,k}(t)dt.$$

donde $d_{j,k}$ es un parámetro - **transformada de wavelet discreta de $f(t)$**.

Transformada aleatoria

A partir del proceso estocástico $Y(t)$ tenemos que

$$D_{j,k} = \langle Y(t), \psi_{j,k}(t)\rangle = \int \psi_{j,k}(t) dY(t) = \int \psi_{j,k}(t) f(t) dt + \varepsilon \int \psi_{j,k}(t) dW(t)$$

Donde $\int \psi_{j,k}(t) dW(t)$ es la integral estocástia de ITO.

$$D_{j,k} = d_{j,k} + \varepsilon \int \psi_{j,k}(t) dW(t); \int \psi_{j,k}(t) dW(t) \sim \mathcal{N}(0,1)$$

Metodología

- Para $H_0 : \mu(t) = 0$, el modelo $\bar{Y}(t) \longrightarrow$ ruido blanco Gausseano.
- Para $H_0 : g_i(t) = 0$, el modelo $Y_i - Y(t) \longrightarrow$ ruido blanco Gausseano.

Metodología

CASO 1: $H_0 : \mu(t) = 0$

$$\bar{Y}(t) = \frac{1}{r}\sum_{i=1}^{r} Y(t) = m_0 + \mu(t) + \varepsilon \bar{W}(t)$$

$$d\bar{Y}(t) = (m_0 + \mu(t))dt + \varepsilon d\bar{W}(t) \qquad (13)$$

Si $\mu(t) = 0$ (constante):

$\bar{D}_{j,k} = \int \psi_{j,k}(t)\mu(t)dt + \varepsilon \int \psi_{j,k}(t)d\bar{W}(t)$

$\bar{D}_{j,k} = \bar{d}_{j,k} + \varepsilon\bar{\xi}_{j,k};$

$\bar{d}_{j,k} = 0$, si $\mu(t) = 0$

Metodología

CASO 2: $H_0 : g_i(t) = 0$

Para $H_0 : g_i(t)$

$$d(Y_i - \bar{Y}(t) = (a_i + g_i(t))dt + \varepsilon(W_i - \bar{W})(t)$$

$$\widetilde{D}_{j,k} = \int \psi_{j,k}(t) g_i(t)dt + \varepsilon\bar{\xi}_{j,k}$$

$$\widetilde{D}_{j,k} = \widetilde{d}_{j,k} + \varepsilon\bar{\xi}_{j,k};$$

$$g_i(t) = 0;\ \forall_i$$

$$\widetilde{d}_{j,k} = 0;\ \forall_{j,k}$$

Tests

Test para $H_0 : \mu(t) = 0$

$$X_{j_0}, \ldots, X_{j_1}$$

$$X_j = (D_1(j,1), \ldots, D_1(j, 2^j - 1), \ldots, D_r(j,1), \ldots, D_r(j, 2^j - 1))$$

Por tanto, $E(X_j) = \mu_j$:

$$H_0 : \mu_{j0} = \cdots = \mu_{j1} = 0$$

$$H_1 : \exists_j,\ \mu_j \neq 0$$

Test

Para $H_0 : g_i(t) = 0$, construimos las matrices

$$\mathbf{X}_{j_0}, \ldots, X_{j_1} \mathbf{X}_{j_0}, \ldots, X_{j_1} \mathbf{x}_{j_0}, \ldots, x_{j_1} \mathbf{x}_{j_0}, \ldots, x_{j_1}$$

X_j cuya columna

$$\mathbf{X}_{ji} \ \mathbf{X}_{ji} \ \mathbf{x}_{ji} \ \mathbf{x}_{ji} = (D_i(j,k), \ldots, D_i(j, 2^j - 1)^T) \mathbf{para} i = 1, \ldots, r$$

Sea

1. Conceptos preliminares
2. Fundamentación del estudio
3. Análisis Multiresolución
4. Estimación
5. Contraste de hipótesis
6. Aplicación
   - Motivación

Análisis y procesamiento de información

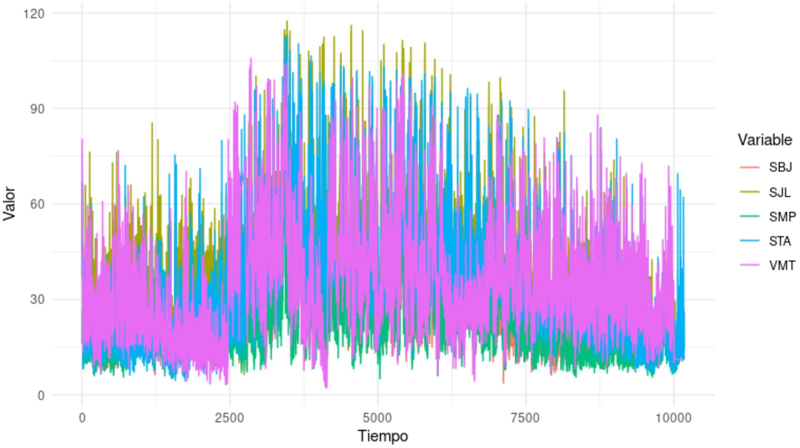


Figura 4: Series de MP2.5 por estación de control

Análisis y procesamiento de información

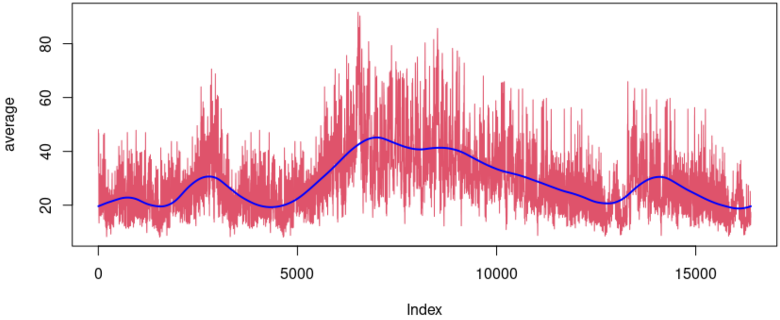


Figura 5: Curva media y estimación Wavelet

Análisis y procesamiento de información

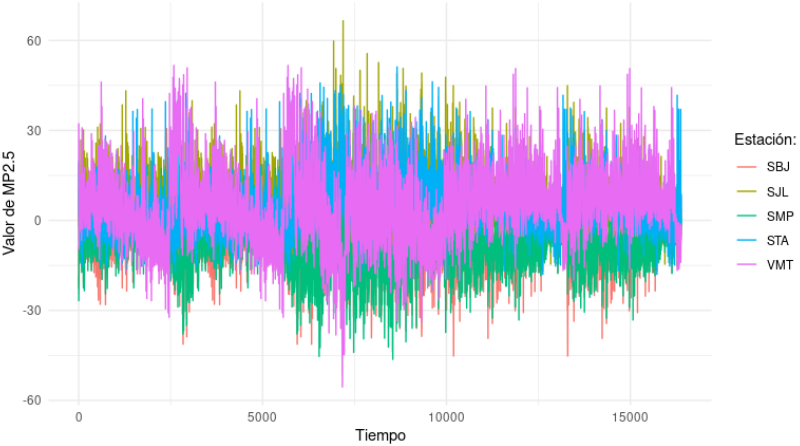


Figura 6: Series de MP2.5 centradas a la media global

Análisis y procesamiento de información

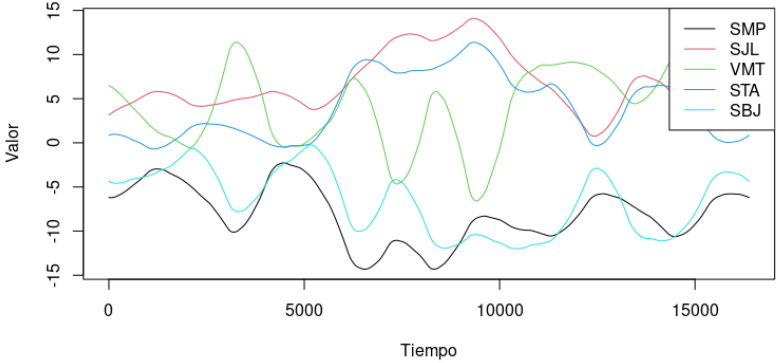


Figura 7: Estimaciones Wavelet de las series de MP2.5

Gracias!