



Universidad
del Valle



años
1945 - 2020

Semana de
la Ingeniería
18 - 21 noviembre 2020

4 Revolución Industrial

NUESTRA GRAN OPORTUNIDAD

Retos y Perspectivas de la Ingeniería para el Desarrollo Social, Económico y Tecnológico.



Semana de
la Ingeniería
18 - 21 noviembre 2020

PROPUESTA DE UN MÉTODO DE IMPUTACIÓN
BASADO EN LA DISTRIBUCIÓN NORMAL
MULTIVARIADA PARA LOS REGISTROS DE $PM_{2.5}$
EN LA CIUDAD DE CALI

- ▶ Resumen
- ▶ Marco conceptual
 - PM 2.5
 - Datos faltantes
 - Normal multivariada y propiedades
 - Metodología de imputación
- ▶ Resultados
 - Descriptivas
 - Desempeño del método



Semana de
la Ingeniería
18 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD

Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Se propone un método de imputación de datos faltantes del contaminante $PM_{2.5}$, de la estación Univalle, 2018. Cada día de la semana se asume distribuido normal 24-variado y a partir de las propiedades de esta distribución, se imputan los datos hallando las distribuciones condicionales para las horas faltantes a partir de las horas disponibles. Se prueba el supuesto de datos faltantes completamente al azar (MCAR) mediante el test de Little (1988) y se prueba normalidad multi-variada usando el test de Mardia (1970). Se evalúa la propuesta vía simulación, obteniendo valores promedio de $R^2 = 0.47$ y $RMSE = 8.6$ respectivamente, lo que muestra que la propuesta es promisoría mejorando imputaciones anteriormente utilizadas.



Semana de
la Ingeniería
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

$PM_{2.5}$

Conjunto de partículas suspendidas en el aire que tienen tamaño menor o igual a $2.5 \mu m$ (micras). conformadas por sustancias químicas tales como Sulfato, Nitrato, Amoníaco, Carbón y Metales pesados. El $PM_{2.5}$ en estado de “acumulación” (entre 0.1 y $0.2 \mu m$) pueden durar suspendidas varios días en el aire.

Se afirma que el tamaño del material particulado está inversamente relacionado con el potencial para contraer problemas de salud cardiovascular (OMS, 2018), debido a que las partículas que ingresan por las vías respiratorias pueden alojarse en los pulmones, o pasar directamente a los vasos sanguíneos



Semana de
la Ingeniería
18 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

$PM_{2.5}$

Lugar	Tiempo de Exposición	Nivel Máx.
Minambiente	Anual	$25 \mu g/m^3$
	Diario	$50 \mu g/m^3$
ONU	Anual	$10 \mu g/m^3$
	Diario	$25 \mu g/m^3$

Table: Nivel máximo de exposición permisible del contaminante $PM_{2.5}$

Los niveles máximos permisibles internacionales son más estrictos y Colombia se piensa igualar su norma a este nivel para el año 2030

Datos faltantes

Sea $Y = (Y_1, \dots, Y_n)$ un vector aleatorio con función de probabilidad f_θ siendo θ el vector de parámetros de interés, y a $M = (M_1, \dots, M_n)$ como el indicador de falta asociado al vector aleatorio Y , que toma el valores:

$$m_i = \begin{cases} 0; & \text{si } y_i \text{ es NO observada} \\ 1; & \text{si } y_i \text{ es observada} \end{cases}$$

La probabilidad de que M tome el valor de $m = (m_1, \dots, m_n)$ dado que Y toma el valor de $y = y_1, \dots, y_n$ es:

$$g_\phi(m|y) = g_\phi(m|(y_{(0)}, y_{(1)}))$$



Semana de
la Ingeniería
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Rubin (1976) Plantea formalmente el fenómeno de datos faltantes, de los cuales se identifican tres principales situaciones en la literatura:

- ▶ Los datos faltantes son perdidos totalmente al azar (MCAR): Es el supuesto más general en el que se afirma que mi proceso generador es independiente de mi variable de interés Y :

$$g_{\phi}(m|y) = g_{\phi}(m)$$

- ▶ Los datos faltantes son perdidos al azar (MAR): Este supuesto afirma que el proceso generador de faltas depende solo de mis datos observados:

$$g_{\phi}(m|y) = g(m|y_{(1)})$$



Semana de
la Ingeniería
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Datos faltantes

- Los datos faltantes no son perdidos al azar (MNAR): Dicho supuesto radica en que la presencia de un dato faltante en una variable Y depende tanto de sus valores observados como sus valores perdidos, que en términos de probabilidad se denota como:

$$g_{\phi}(m|y) = g_{\phi}(m|y_{(1)}, y_{(0)})$$

Así pues, la probabilidad de que falten datos en Y depende tanto de los valores observados como de los valores perdidos.



NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

4 Revolución Industrial

Test para probar MCAR

si X_i es iid y la falta de datos es MCAR, los estimadores por máxima verosimilitud tanto de μ y Σ son $\tilde{\mu} = \bar{Y}$ y $\tilde{\Sigma} = n * \frac{\hat{\Sigma}}{n-1}$. Se procede a establecer el estadístico:

$$d_0^2 = \sum_{j=1}^J r_j (\bar{y}_{(1),j} - \mu_{(1),j}) \Sigma_{(1),j}^{-1} (\bar{y}_{(1),j} - \mu_{(1),j})^T \quad (1)$$

Con $\bar{y}_{(1),j} = r_j^{-1} \sum_{i \in S_j} y_{(1),i}$ y $\Sigma_{(1),j} = D_j^T \Sigma D_j$, medias y matrices de covarianzas de las variables observadas en el patrón j , que se distribuye χ^2 con $\sum_j p_j - p$ grados de libertad.



Semana de
la Ingeniería
18 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD

Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Test para probar MCAR

Little demuestra que d_0^2 es el estadístico de razón de verosimilitud para probar el modelo:

- ▶ $(y_{(1),i} | m_i) \sim N(\mu_{(1,j)}, \Sigma_{(1,j)})$ con $i \in S_j$, contra el modelo alternativo
- ▶ $(y_{(1),i} | m_i) \sim N(v_{(1,j)}, \Sigma_{(1,j)})$

Siendo $v_{(1),j}$ el vector de medias de las variables observadas (diferentes de $\mu_{(1,j)}$) que son distintos en cada patrón j .



Semana de
la Ingeniería
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Normal Multivariada

La función de densidad multivariada es una generalización de la densidad normal univariada. Así, dado un vector aleatorio $\mathbf{X}^T = [X_1, X_2, \dots, X_p]$ con distribución normal p-variada su función de densidad se denota a continuación:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})} \quad (2)$$

- ▶ $\boldsymbol{\mu}$: Representa el vector de medias del vector aleatorio.
- ▶ Σ : La matriz de varianzas y covarianzas entre cada una de las p variables.



Semana de
la Ingeniería
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Propiedades

- Todo subconjunto de X se distribuye normal multivariado. Si se particionan X , μ y Σ respectivamente como

$$\underset{(p \times 1)}{X} = \begin{bmatrix} \underset{(q \times 1)}{X_1} \\ \text{---} \\ \underset{((p-q) \times 1)}{X_2} \end{bmatrix} \quad \underset{(p \times 1)}{\mu} = \begin{bmatrix} \underset{(q \times 1)}{\mu_1} \\ \text{---} \\ \underset{((p-q) \times 1)}{\mu_2} \end{bmatrix} \quad (3)$$

$$\underset{(p \times p)}{\Sigma} = \begin{bmatrix} \underset{(q \times q)}{\Sigma_{11}} & \underset{(q \times (p-q))}{\Sigma_{12}} \\ \text{---} & \text{---} \\ \underset{((p-q) \times q)}{\Sigma_{21}} & \underset{((p-q) \times (p-q))}{\Sigma_{22}} \end{bmatrix} \quad (4)$$

Propiedades

- ▶ Teniendo el mismo escenario anterior, la distribución condicional de X_1 dado que $X_2 = x_2$ es normal con

$$\mu_{X_1|X_2=x_2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$

$$\Sigma_{X_1|X_2=x_2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Enunciado de otra forma, la distribución de $X_1|X_2 = x_2$ es $N_q(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$



Semana de
la Ingeniería
18 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Estimación

Para hallar los estimadores de μ y Σ se usa el método de estimación por máxima verosimilitud teniendo como resultado que:

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{j=1}^n X_j \quad Y \quad \hat{\Sigma} = S = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}) (X_j - \bar{X})^T$$

En este caso S pueden presentar problemas de condicionamiento debido a que $n \approx p$. Godino Gomez (2014)



Semana de
la Ingeniería
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD

Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Estimador de shrinkage

Se presenta una estimación diferente de las matrices de varianza y covarianza mediante el estimador de Shrinkage o encogimiento, que se denota de la siguiente manera:

$$\Sigma^* = \lambda T + (1 - \lambda)S \quad (5)$$

Donde T es llamada matriz objetivo y representa un modelo reducido de la matriz de covarianzas en el que se estiman menos parámetros y $\lambda \in [0, 1]$ representa la intensidad de contracción (Schäfer & Strimmer 2005).

$$T = \begin{cases} s_{ii} & si, i = j \\ 0 & si, i \neq j \end{cases} \quad (6)$$



Semana de
la Ingeniería
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

El método de imputación propuesto se encuentra basado en el siguiente proceso generador:

$$X_j \sim N_{24}(\bar{X}_j, S_j); \quad j = 1, 2, \dots, 7$$

Donde j se refiere a cada día de la semana comenzando por el lunes y terminando en el domingo. \bar{X}_j y S_j son los estimadores máximo verosímiles de los parámetros de la distribución normal multivariada de cada día.

Así el método de imputación propuesto se basa en utilizar la distribución condicional de $X_{j1}|X_{j2} = x_{j2}$ en un día k donde se tienen m_k datos faltantes, donde:

- ▶ X_{j1} Hace referencia a los m_k datos que se deben imputar en ese día.
- ▶ $X_{j2} = x_{j2}$ Serán los datos que se observaron en ese día.



Semana de
la Ingeniería
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Entonces si se quiere imputar M datos en total en K días, para cada día k se genera un vector z_k de tamaño m_k a partir de la distribución mencionada anteriormente y estos serán los valores que se van imputar.

$$\hat{y} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_K \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} \quad (7)$$



**Semana de
la Ingeniería**
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Nota :

- ▶ Un día podrá tener de 1 a 6 mediciones faltantes consecutivas como máximo ya que de acuerdo a la norma, debe contarse con al menos el 75% de información para realizar análisis y estimaciones de indicadores.
- ▶ La generación de los escenarios se realiza siguiendo la metodología propuesta por Otero & Presiga (2018), este método comienza haciendo un análisis de los datos faltantes en brechas horarias.
- ▶ **Definición:** Brecha de tamaño n : Una brecha de tamaño n , con $n \in (1, \dots, 6)$ cuenta con n horas seguidas en las que hay datos faltantes pero antes y después de estas no pueden haber horas que presentan datos faltantes.



Semana de
la Ingeniería
18 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Generación de escenarios de simulación

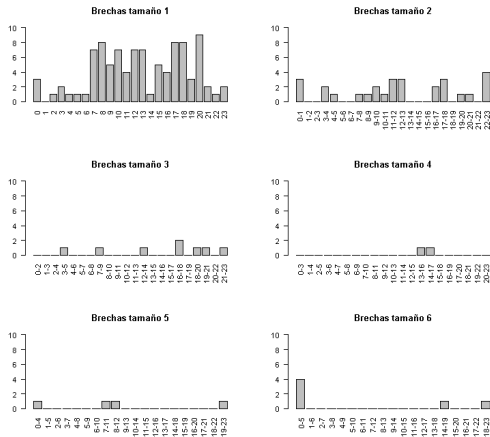


Figure: Frecuencia de las brechas según su posición y tamaño

Generación de escenarios de simulación

Tipo de brecha	Frec. Absoluta(días)	f_i	F_i
Brechas de tamaño 1	78	0,6783	0,6783
Brechas de tamaño 2	18	0,1565	0,8348
Brechas de tamaño 3	7	0,0609	0,8957
Brechas de tamaño 4	2	0,0174	0,9130
Brechas de tamaño 5	4	0,0348	0,9478
Brechas de tamaño 6	6	0,0522	1
Total	115	1	-

Table: Distribución de días de acuerdo al tipo de brecha de faltantes



**Semana de
la Ingeniería**
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Generación de escenarios de simulación

Cantidad de brechas	Frec. Absoluta(días)	f_i	F_i
Una brecha	63	0,8077	0,8077
Dos brechas	12	0,1538	0,9615
Tres brechas	2	0,0256	0,9872
Cuatro brechas	1	0,0128	1
Cinco brechas	0	0	1
Seis brechas	0	0	1
Total	78	1	-

Table: Distribución de brechas de tamaño 1



Semana de
la Ingeniería
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Generación de escenarios de simulación

Cantidad de brechas	Frec. Absoluta(días)	f_i	F_i
Una brecha	12	0,6667	0,6667
Dos brechas	4	0,2222	0,8889
Tres brechas	2	0,1111	1
Cuatro brechas	0	0	1
Total	18	1	-

Table: Distribución de brechas de tamaño 2

Cantidad de brechas	Frec. Absoluta(días)	f_i	F_i
Una brecha	6	0,8571	0,8571
Dos brechas	1	0,1428	1
Tres brechas	0	0	1
Total	7	1	-

Table: Distribución de brechas de tamaño 3



Semana de
la Ingeniería
18 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Generación de escenarios de simulación

Paso 1: A partir de la matriz de días completos se escoge al azar un porcentaje T de estos. $T = \{20\%, 40\%, 60\%, 80\%, 100\%\}$.

Paso 2: Se toma el primer día elegido en el paso anterior y se genera un número de la distribución $U \sim (0, 1)$ y según el valor obtenido este día tendrá cierto tipo de brecha.

Paso 3: Teniendo el tipo de brecha, se procede a elegir la cantidad de brechas que se van a generar dentro del día, de nuevo haciendo uso de la distribución $U \sim (0, 1)$.

Paso 4: Después de tener el tipo de brecha y su cantidad a generar en un día se eligen aleatoriamente las posiciones en las que estas se ubicaran.

Paso 5: Se repiten los pasos 2, 3 y 4 para todos los días seleccionados en el paso 1.

Paso 6: Se realiza el proceso de imputación propuesto a la base de datos contaminada.

Paso 7: Teniendo la base de datos imputada, se procede a evaluar el método mediante el RMSE y el R^2 .



Semana de
la Ingeniería
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^M (P_i - O_i)^2} \quad (8)$$

$$R^2 = \left(\frac{1}{M} \frac{\sum_{i=1}^M (P_i - \bar{P})(O_i - \bar{O})}{\sigma_P \sigma_O} \right)^2 \quad (9)$$



**Semana de
la Ingeniería**
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD

Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Resultados: Descriptivas

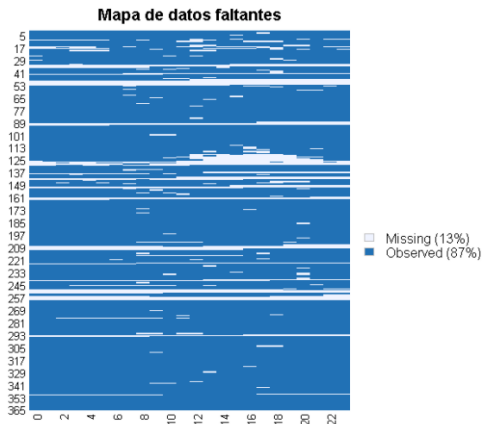


Figure: Mapa de datos faltantes

Resultados: Descriptivas

Cant. Fal	Categorización	Frec.	Frec. Rel	Frec. Acum
0	Completo	220	0,603	0,603
1-6	Parcial/Completo	84	0,230	0,833
7-23	Parcial/Incompleto	42	0,115	0,948
24	Incompleto	19	0,052	1
Total		365	1	-

Table: Clasificación de los días según la cantidad datos faltantes



**Semana de
la Ingeniería**
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Resultados: Descriptivas

	Día de la semana	Días Completos	Frecuencia
1	Lunes	26	0.5
2	Martes	34	0.65
3	Miércoles	33	0.63
4	Jueves	28	0.54
5	Viernes	32	0.62
6	Sábado	35	0.67
7	Domingo	32	0.62

Table: Datos completos por día de la semana



Semana de
la Ingeniería
18 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD

Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Resultados: Descriptivas

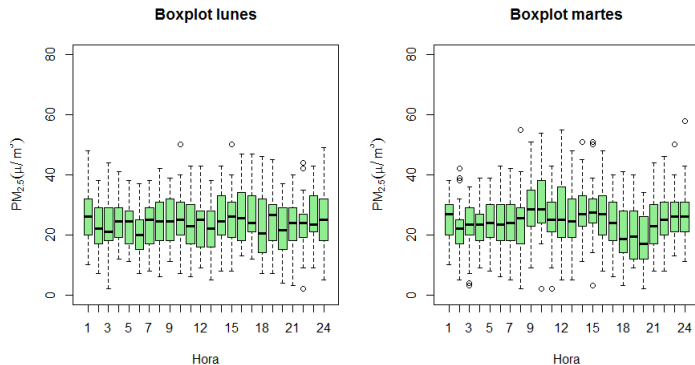


Figure: Gráficos de caja por hora y día de la semana



**Semana de
la Ingeniería**
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Resultados: Descriptivas

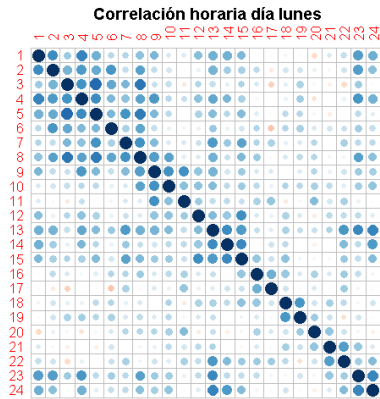


Figure: Correlaciones horarias día lunes

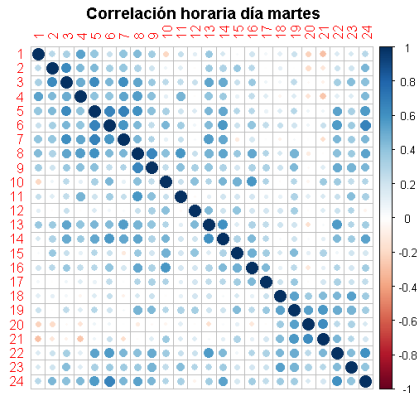


Figure: Correlaciones horarias día martes



**Semana de
la Ingeniería**
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Resultados: Desempeño del método

% Días falt.	% Datos falt.
20%	0.016 (0.011, 0.022)
40%	0.033 (0.024, 0.041)
60%	0.049 (0.041, 0.061)
80%	0.066 (0.051, 0.078)
100%	0.083 (0.071, 0.097)

Table: Porcentaje de datos faltantes dado el porcentaje de días faltantes.



**Semana de
la Ingeniería**
18 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Resultados: Desempeño del método

% Días falt.	Matriz de covarianza	RMSE	R2
20%	$\hat{\Sigma}$	5.682 (3.91, 8.194)	0.698 (0.352, 0.875)
	$\hat{\Sigma}^*$	8.391 (6.224, 10.281)	0.434 (0.164, 0.635)
40%	$\hat{\Sigma}$	5.698 (4.374, 7.487)	0.698 (0.464, 0.829)
	$\hat{\Sigma}^*$	8.565 (7.201, 10.075)	0.411 (0.217, 0.596)
60%	$\hat{\Sigma}$	5.732 (4.479, 7.086)	0.699 (0.554, 0.808)
	$\hat{\Sigma}^*$	8.597 (7.407, 9.985)	0.41 (0.282, 0.568)
80%	$\hat{\Sigma}$	5.707 (4.618, 6.799)	0.7 (0.578, 0.804)
	$\hat{\Sigma}^*$	8.536 (7.51, 9.32)	0.417 (0.325, 0.526)
100%	$\hat{\Sigma}$	5.718 (4.761, 6.673)	0.7 (0.581, 0.795)
	$\hat{\Sigma}^*$	8.556 (7.523, 9.996)	0.411 (0.281, 0.507)

Table: Resultados del desempeño del método



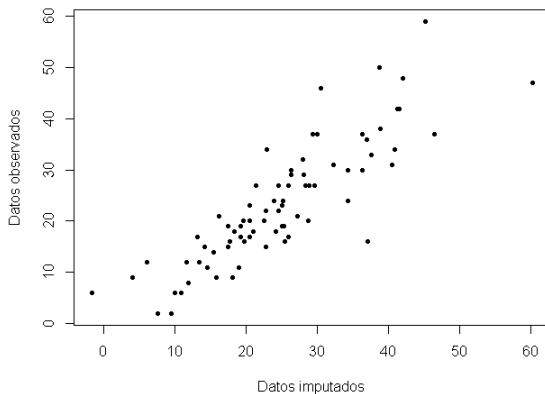
Semana de
la Ingeniería
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Resultados: Desempeño del método

Dispersión entre datos imputados y observados



**Semana de
la Ingeniería**
18 - 21 noviembre 2020

Figura: Gráfico de dispersión

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Resultados: Aplicación de la imputación

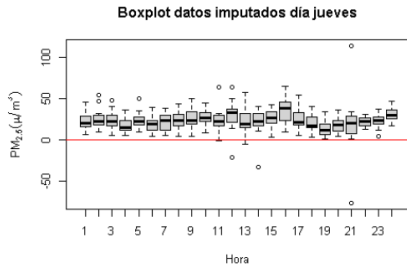
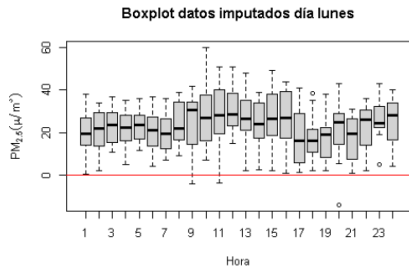


Figure: Diagramas de caja datos imputados

Figure: Diagramas de caja datos imputados



**Semana de
la Ingeniería**
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Resultados: Aplicación de la imputación

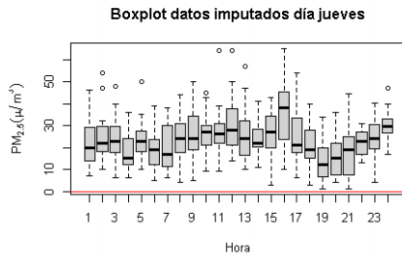
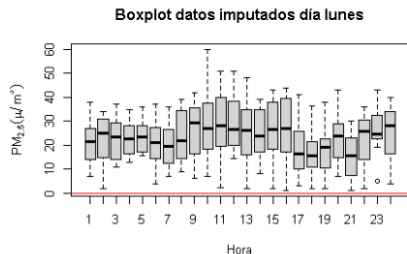


Figure: Diagramas de caja datos imputados

Figure: Diagramas de caja datos imputados



**Semana de
la Ingeniería**
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Resultados: Prueba MCAR

La tabla muestra el contraste del test de little, donde la hipótesis nula se afirma que el patrón de datos faltantes es MCAR.

Día	GL	Estadístico	Valor-P
Lunes	323	357.26	0.091
Martes	219	249.77	0.075
Miércoles	198	231.7	0.051
Jueves	284	302.14	0.219
Viernes	263	324.59	0.001
Sábado	153	194.31	0.013
Domingo	179	182.38	0.416

Table: Resultados prueba de Little



Semana de
la Ingeniería
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

Resultados: Supuesto normalidad

Dia/Test	Simetría		Curtosis		Cumplimiento
	Est.	P-valor	Est.	P-valor	
Lun	2331.20	0.99	-3.335	0.00	SI/NO
Mar	2559.89	0.70	-1.69	0.089	SI/SI
Mie	2477.60	0.95	-2.32	0.02	SI/NO
Jue	2329.22	0.99	-3.24	0.001	SI/NO
Vie	2464.42	0.97	-2.32	0.019	SI/NO
Sab	2616.78	0.40	-1.27	0.20	SI/SI
Dom	2466.60	0.96	-2.43	0.014	SI/NO

Table: Resultados del test de Mardia



Semana de
la Ingeniería
18 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

- ▶ A. Otero and M. Presiga. 'evaluación de un método de imputación basado en el análisis de datos funcionales para los registros de pm2.5 en la ciudad de cali'. Trabajo de grado es Estadística, Universidad del Valle, Colombia, 2019.
- ▶ R. Johnson and D. Wichern. Applied Multivariate Statistical Analysis. Applied Multivariate Statistical Analysis. Pearson Prentice Hall, 2007. ISBN 9780131877153. URL <https://books.google.com.co/books?id=gFWcQgAACAAJ>.
- ▶ C. K. Enders. Applied Missing Data Analysis. University of Pennsylvania, New York, USA, 2010



**Semana de
la Ingeniería**
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**

- ▶ D. B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.
- ▶ R. J. A. Little. A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83, 12 1988.
- ▶ K. V. Mardia. Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3):519–530, 1970.
- ▶ J. Cao, J. Chow, J. Watson, and S. Lee. A brief history of pm2.5 and its adverse effects. *AAQR*, pages accepted–, 01 2013.
- ▶ Godino Gomez, F. (2014), 'Estimación de matrices de covarianzas: Nuevas perspectivas'.



Semana de
la Ingeniería
15 - 21 noviembre 2020

NUESTRA GRAN OPORTUNIDAD
Retos y Perspectivas de la Ingeniería para el Desarrollo
Social, Económico y Tecnológico.

**4 Revolución
Industrial**