Una propuesta basada en estimadores Bootstrap robustos para la evaluación de la precisión de un modelo con la técnica de regresión lineal

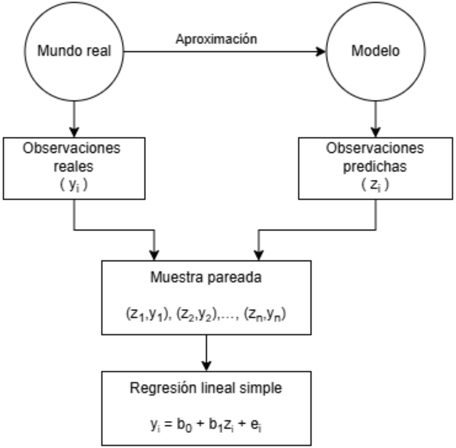
Br. Irving Cupul Uc

Examen Profesional:
Licenciado en Ingeniería de Software.

Asesores:
MC. Luis Colorado Martínez
MC. Salvador Medina Peralta

12 de Diciembre de 2024

Validación de un modelo con regresión lineal simple



Representación simplificada de un sistema en la necesidad de explicar la realidad.

Tipos de modelos:

- Conceptuales
- Físicos
- Matemáticos
- Probabilísticos
- Gráficos

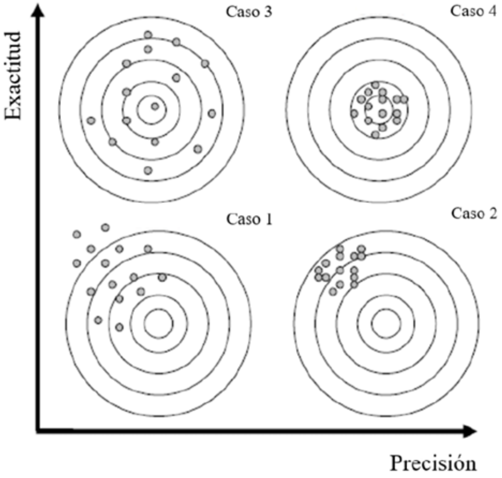
Aplicación de los modelos:

- Ingeniería
- Física
- Ciencias químicas
- Ciencias económicas
- Ciencias administrativas
- Ciencias biológicas
- Ciencias sociales

Validación de un modelo con regresión lineal simple

Exactitud y Precisión

- Exactitud: qué tan cerca están los valores reales de los valores predichos.
- Precisión: qué tan cerca están entre ellos los valores predichos por el modelo.



Validación de un modelo con regresión lineal simple

¿Cómo evalúa la Exactitud y Precisión?

Dada la muestra pareada $(z_1, y_1), (z_2, y_2), \ldots, (z_n, y_n)$ se considera el modelo de regresión lineal:

$$y_i = \beta_0 + \beta_1 z_i + \epsilon_i, \text{ donde } 1 \le i \le n.$$

El modelo evaluado es exacto si $\boldsymbol{\beta_0 = 0}$ y $\boldsymbol{\beta_1 = 1}$. Y es preciso si el coeficiente de determinación R^2 está cercano a 1.

Inconvenientes:

- Las pruebas estadísticas para evaluar la exactitud dependen del supuesto de normalidad y de varianza constante de los ϵ_i.
- La precisión se mide de manera determinística.

Trabajos realizados para evaluación de la exactitud y precisión

- Balam, R. (2012). "Evaluación de la exactitud y precisión de un modelo con regresión lineal”. Tesis de Maestría, Facultad de Matemáticas, Universidad Autónoma de Yucatán.

  Para los casos:

  a) NVC se propuso una región de confianza con la F conjunta para evaluar la exactitud y un ICB BCa con residuales balanceados para la precisión.

  b) NNVC se emplearon ICB Percentil con sesgo corregido y un ICB BCa para evaluar la exactitud y la precisión respectivamente, ambos con los residuales balanceados.

  c) NVD y NNVD se emplearon ICB Percentil con sesgo corregido y un ICB BCa para evaluar la exactitud y la precisión respectivamente, ambos con las muestras pareadas balanceadas.

Trabajos realizados para evaluación de la exactitud y precisión

- Zacarías, K. (2023). "Evaluación de la exactitud de un modelo cuando no se cumplen Los supuestos en la técnica de regresión lineal". Tesis de Maestría, Facultad de Matemáticas, Universidad Autónoma de Yucatán.

a) Se propone la construcción de una región de confianza para el vector (β_0, β_1) con diferentes esquemas Bootstrap robustos. Si el punto (0,1) esta dentro de la región, el modelo es exacto, de lo contrario es inexacto.

b) Se implementaron seis esquemas de remuestreo Bootstrap **para estimar la distribución de la estadística F conjunta** cuando no se cumplen los supuestos de normalidad y varianza constante:
   - i) El Bootstrap Robusto Simple.
   - ii) Los tres esquemas propuestos por Wu (1986).
   - iii) Y los dos esquemas propuestos por Lui (1988).

c) Se propuso un estimador para el cuantil de la distribución F conjunta estimada.

El coeficiente de determinación R^2

¿Cómo evaluar la precisión?

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \text{ y } 0 \leq R^2 \leq 1,$$

Inconvenientes:

- No se conoce su distribución.
- No se puede usar para hacer predicciones.
- En presencia de valores atípicos o agrupamientos (clustering) en los datos, R^2 puede ser engañoso, pues un valor grande podría ser el resultado de estos factores, en lugar de un ajuste verdadero.

¿Cómo evaluar la precisión?

En este trabajo se propone un método que permite evaluar la precisión de un modelo con la técnica de regresión lineal; y se basa en implementar diversos esquemas de remuestreo y estimar la precisión, a través de intervalos de confianza Bootstrap (ICB) para el coeficiente de determinación R^2, del modelo de regresión entre los valores reales y predichos del modelo que se desea evaluar.

Objetivo General

Determinar la precisión de un modelo con la técnica de regresión lineal por medio de intervalos de confianza basado en diferentes esquemas de remuestreo Bootstrap y medir sus eficiencias a través de un estudio de simulación.

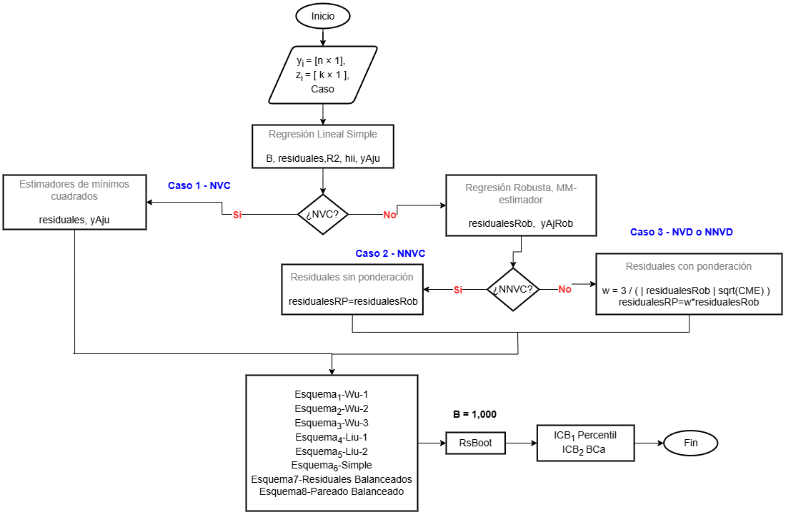
Objetivo Específicos

1. Desarrollar la metodología para medir la precisión de un modelo con la técnica de regresión lineal por medio de intervalos de confianza basado en diferentes esquemas de remuestreo Bootstrap.
2. Determinar la precisión de un modelo cuando se cumplan los supuestos de normalidad y varianza constante.
3. Determinar la precisión de un modelo cuando no se cumplan los supuestos de normalidad y/o varianza constante.

Objetivo Específicos

4. Diseñar e implementar un estudio de simulación para evaluar la eficiencia de la metodología propuesta.
5. Simular modelos exactos-precisos (EP) y modelos exactos-imprecisos (EI) mediante la propuesta de Febles (2014) y Zacarias (2023); cuando se cumplan o no los supuestos de normalidad e igualdad de varianzas.
6. Determinar la eficiencia de los esquemas Bootstrap propuestos para medir la precisión de un modelo.

La propuesta



La propuesta

2. **Esquemas de remuestreo Bootstrap para la estimación de R^2.**
   Para los esquemas propuestos en Rana, es necesario generar una muestra Bootstrap y_i^* dado por :

$$y_i^* = z_i\hat{\beta}_{MM} + \frac{t_i^*\hat{e}_i^{WMM}}{\sqrt{1-h_{ii}}},$$

   y ajustar la regresión simple:

$$y_i^* = \beta_0^* + \beta_1^* z_i + \epsilon_i^*,$$

   y obtener las nuevas R^{2*} para estimar su distribución.

La propuesta

- **Wu 1**: $t_i^* \in N(0,1),\ i = 1,\dots,n.$
- **Wu 2**: Los t_i^* corresponden a una muestra con reemplazo de los residuos normalizados.
- **Wu 3**: El valor t_i^* se obtiene mediante un remuestreo con reemplazo al vector de residuales transformados,

$$R_{ai} = \frac{\hat{e}_i^{WMM} - Mediana(\hat{e}_i^{WMM})}{NMAD(\hat{e}_i^{WMM})},\ NMAD = \frac{1}{0.6745} Mediana\{|\hat{e}_i^{WMM} - Mediana(\hat{e}_i^{WMM})|\}.$$

- **Liu 1**: $t_i^* \in Gamma(4, 1/2),\ i = 1,\dots,n$
- **Liu 2**: $t_i^* = H_i D_i - E(H_i)E(D_i)$ donde
  $H_i \in N(0.5 \times \sqrt{17/6} + \sqrt{1/6}, 0.5) = N(1.2498, 0.5)$
  $D_i \in N(0.5 \times \sqrt{17/6} - \sqrt{1/6}, 0.5) = N(0.4334, 0.5),\ i = 1,\dots,n.$ Ambas independientes entre si.

La propuesta

Para los otros esquemas, para la muestra Bootstrap y_i^* se obtiene remuestreando los residuales:

$$y_i^* = \hat{y}_i + e_i^*,$$

y ajustar la regresión simple:

$$y_i^* = \beta_0^* + \beta_1^* z_i + \epsilon_i^*,$$

y obtener las nuevas R^{2*} para estimar su distribución.

- **Bootstrap Simple**:
  e_i^* se obtienen de forma aleatoria con reemplazo de los residuales.
- **Bootstrap de Residuales Balanceados**:
  e_i^* se obtiene de un conjunto aleatorio de residuales seleccionados de tal forma que cada uno aparece el mismo número de veces (nB).

La propuesta

- **Bootstrap Pareado Balanceado**:
  Para la estimación de la distribución de la R^{2*} se considera la muestra pareada (y_i, z_i), y se remuestrea aleatoriamente B veces con remplazo y de forma balanceada, para obtener las muestras $\mathbf{y}_1^*, \mathbf{y}_2^*, \ldots, \mathbf{y}_B^*$ y $\mathbf{z}_1^*, \mathbf{z}_2^*, \ldots, \mathbf{z}_B^*$ y ajustar una regresión simple entre los vectores $\mathbf{y}_i^*$ y $\mathbf{z}_i^*$, $b = 1, 2, \ldots, B$.

La propuesta

3. **Cálculo del los intervalos de confianza para R^2.**

Intervalos de confianza Bootstrap Percentil (**ICB Percentil**) por esquema:

1. Se obtienen las muestras, $\hat{R}_1^{2*}, \hat{R}_2^{2*}, \ldots, \hat{R}_B^{2*}$.
2. Las B muestras $\hat{R}_1^{2*}, \hat{R}_2^{2*}, \ldots, \hat{R}_B^{2*}$ se ordenan de manera ascendente, tal que $\hat{R}_1^{2*} \leq \hat{R}_2^{2*} \leq \cdots \leq \hat{R}_B^{2*}$.
3. Determinar los cuantiles LI y LS, para el nivel de confianza del $(1-\alpha)\%$ en la muestra Bootstrap ordenada, con $LI = \hat{R}^{2*}_{(\alpha/2)\times B}$ y $LS = \hat{R}^{2*}_{(1-\alpha/2)\times B}$.
4. El intervalo de confianza esta dado por: $[LI_{per}, LS_{per}]$.

La propuesta

Intervalos de confianza Bootstrap BCa (**ICB BCa**) por esquema:

1. Obtener una estimación de la $\hat{R}^2$ a partir de los datos originales.
2. Se obtienen las muestras, $\hat{R}_1^{2*}, \hat{R}_2^{2*}, \ldots, \hat{R}_B^{2*}$.
3. Se determina la proporción p de las $\hat{R}_i^{2*}$ que son mayores o iguales a $\hat{R}^2$.
4. Determinar $Z_0 = Z_p$ donde Z_p es el cuantil en la distribución normal estándar tal que $P(Z > Z_p) = p$.
5. Obtener la constante de aceleración a dada por:

$$a = \frac{\sum_{i=1}^{n} (\hat{R}^2_{-iprom} - \hat{R}^2_{-i})^3}{6\,(\sum_{i=1}^{n} (\hat{R}^2_{-iprom} - \hat{R}^2_{-i})^2\,)^{3/2}}\,,$$

donde: $\hat{R}^2_{-i}$ es la estimación con los datos originales quitando la i-ésima observación y $\hat{R}^2_{-iprom} = \frac{1}{n}\sum_{i=1}^{n} \hat{R}^2_{-i}$.

La propuesta

Intervalos de confianza Bootstrap BCa (**ICB BCa**) por esquema (...continuación):

6. Obtener $Z_L = \frac{Z_0 - Z_{\alpha/2}}{1 - a(Z_0 - Z_{\alpha/2})} + Z_0$ y $Z_U = \frac{Z_0 + Z_{\alpha/2}}{1 - a(Z_0 + Z_{\alpha/2})} + Z_0$ donde: $Z_{\alpha/2}$ es el cuantil en la distribución normal estándar tal que $P(Z > Z_{\alpha/2}) = \frac{\alpha}{2}$.
7. Encontrar $LI = INVCDF(\Phi(Z_L))$ y $LS = INVCDF(\Phi(Z_U))$ donde $INVCDF$ es el cuantil en la muestra Bootstrap con probabilidad $\Phi(Z_L)$ y $\Phi(Z_U)$ respectivamente y Φ es la distribución acumulada de la normal estándar, siendo $P(\hat{R}^{2*} < LI) = \Phi(Z_L)$ y$P(\hat{R}^{2*} < LS) = \Phi(Z_U)$.
8. El intervalo de confianza esta dado por: $[LI_{bca}, LS_{bca}]$.

Diseño del estudio de simulación

Para la simulación de los modelos se utilizaron los simuladores **ModNVC**(), **ModNNVC**(), **ModNNVD**() y **ModNVD**() (Febles, 2014; Zacarías, 2023). Se construyó la función **SimMod**() para la utilización de los simuladores.

Exacto-Preciso (EP): 60,000 modelos	**Exacto-Impreciso (EI): 60,000 modelos**
• $\beta_0 = 0$ y $\beta_1 = 1$ • R^2: aleatorio entre 0.8 y 0.99 • $n = 10, 15, 20, 25, 30$ y 35 • Supuestos: NVC, NVD, NNVC, NNVD • 500 modelos • 5 replicas	• $\beta_0 = 0$ y $\beta_1 = 1$ • R^2: aleatorio entre 0.1 y 0.33 • $n = 10, 15, 20, 25, 30$ y 35 • Supuestos: NVC, NVD, NNVC, NNVD • 500 modelos • 5 replicas

En total se respaldaron 48 matrices que contienen los 120,000 modelos y 48 matrices que contienen las R^2 de origen correspondiente a cada modelo simulado.

Eficiencias para los ICB

	A	B	C	D	E	F	G	H	I	J
1	Replica	Esquema	NumMod	NumModEfic	FrecEficIB1	FrecEficIB2	FrecEficIB1Unico	FrecEficIB2Unico	FrecEficIB1Emp2	FrecEficIB2Emp2
2	1	1	500	500	454	465	11	22	414	29
3	1	2	500	500	412	426	11	25	378	23
4	1	3	500	493	415	388	46	19	183	186
5	1	4	500	500	290	293	152	155	132	6
6	1	5	500	500	487	488	8	9	453	26
7	1	6	500	500	346	348	36	38	254	56
8	1	7	500	500	347	348	42	43	251	54
9	1	8	500	500	460	472	11	23	324	125
10	2	1	500	500	457	467	6	16	435	16
11	2	2	500	500	416	432	7	23	385	24
12	2	3	500	496	410	405	32	27	187	191

En total se construyeron 96 tablas para las eficiencias, de las cuales 48 fueron para modelos EP y 48 para modelos EI. Para resumir los resultados de las tablas se construyeron 8 nuevas tablas con las eficiencias promedio.

Análisis estadísticos

Para cada supuesto (NVC, NNVC, NVD, NNVD) se utilizó ANOVA en un arreglo factorial de tres factores, donde:

- y = Eficiencia del ICB (Percentil o BCa)
- β = Efecto del tamaño de la muestra (6 niveles).
- τ = Efecto del esquema (8 niveles).
- γ = Efecto del tipo de modelo (2 niveles).

Para las comparaciones múltiples se utilizó Tukey al 5%.

En total se corrieron ocho ANOVAs, en cada comparación múltiple de Tukey se compararon 96 medias dando un total de 4,560 comparaciones de pares de medias.

Comparación de la eficiencia del ICB Percentil para NVC

Cuando se tiene NVC y se utilizó el ICB Percentil para evaluar la precisión, se obtuvo interacción triple significativa ($TipoMod \times TM \times Esq : F = 11.97, P < 0.0001$). Con base en una eficiencia promedio de al menos 95%, el mejor esquema resultó Liu2 sin importar el tamaño de muestra y tipo de modelo.

Condición	Eficiencia Promedio	Grupos Homogéneos
EINVC10Liu2	0.951182	defghijk
EPNVC25Wu1	0.9516	defghijk
EPNVC35Wu1	0.9548	efghijkl
EPNVC10Liu2	0.9576	fghijklm
EINVC15Liu2	0.959163	fghijklm
EINVC30Liu2	0.964	ghijklm
EINVC25Liu2	0.9644	hijklm
EINVC20Liu2	0.9684	ijklm
EINVC35Liu2	0.9688	ijklm
EPNVC15Liu2	0.9752	jklm
EPNVC20Liu2	0.9852	klm
EPNVC30Liu2	0.99	lm
EPNVC25Liu2	0.9916	m
EPNVC35Liu2	0.992	m

Promedios con igual letra no difieren (P>0.05), prueba de Tukey.

Comparación de la eficiencia del ICB BCa para NVC

Cuando se tiene NVC y se utilizó el ICB BCa para evaluar la precisión, se obtuvo interacción triple significativa ($TipoMod \times TM \times Esq : F = 3.60, P < 0.0001$). Con base en una eficiencia promedio de al menos 95%, el mejor esquema resultó Liu2 sin importar el tamaño de la muestra, sin embargo, sólo identifica al tipo de modelo EP *a priori* simulado.

Condición	Eficiencia Promedio	Grupos Homogéneos
EPNVC30Wu1	0.9532	opqrstu
EPNVC25Wu1	0.9544	pqrstu
EPNVC35Wu1	0.9616	qrstu
EPNVC10Liu2	0.9644	rstu
EPNVC15Liu2	0.9736	stu
EPNVC20Liu2	0.9804	tu
EPNVC35Liu2	0.9892	u
EPNVC25Liu2	0.9896	u
EPNVC30Liu2	0.9896	u

Promedios con igual letra no difieren (P>0.05), prueba de Tukey.

Comparación de las eficiencias de los ICB para NNVC

Comparación con ICB Percentil:

Con base en una eficiencia promedio de al menos 95%, el mejor esquema resultó Liu2 sin importar el tamaño de muestra y tipo de modelo; con excepción del caso EINNVC10Liu2, sin embargo, su eficiencia promedio es 94.52%.

Comparación con ICB BCa:

Con base en una eficiencia promedio de al menos 95%, se obtuvo dos mejores esquemas Liu2 y Wu1 sin importar el tamaño de la muestra, sin embargo, ambos sólo identifican al tipo de modelo EP *a priori* simulado.

Comparación de las eficiencias de los ICB para NVD

Comparación con ICB Percentil:

Con base en una eficiencia promedio de al menos 95%, el mejor esquema resultó Liu2 sin importar el tamaño de muestra y tipo de modelo.

Comparación con ICB BCa:

Con base en una eficiencia promedio de al menos 95%, el mejor esquema resultó Liu2 sin importar el tamaño de la muestra, sin embargo, sólo identifica al tipo de modelo EP *a priori* simulado.

Comparación de las eficiencias de los ICB para NNVD

Comparación con ICB Percentil:

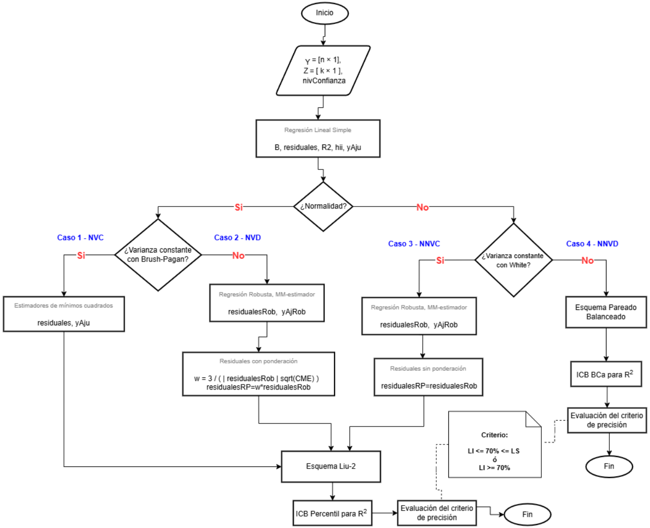
Con al menos el 93.96%, los dos mejores esquemas fueron Liu2 y ParBal para todos los TM con excepción de $n = 35$ para Liu2 (92.72%) y $n = 10$ para ParBal (91.28%). Ambos esquemas sólo identifican a los EI.

Comparación con ICB BCa:

Con al menos el 88.80%, con el esquema ParBal se obtuvo la mayor eficiencia promedio en todos los TM, también bajo el esquema Liu2 con excepción de $n = 35$ (88.52%) cuando el tipo de modelo es EP. El esquema ParBal identifica a los EI para $n = 25, 30, 35$ y las eficiencias no difieren estadísticamente con al menos 90.4%.

En resumen: Se determinó con al menos el 88.8% que para el supuesto NNVD se utilice el ICB BCa con el esquema de remuestreo pareado balanceado; con la limitación de que para modelos EI con tamaños de muestra "pequeño" $n = 10, 15, 20$, no se obtuvo un buen desempeño.

Propuesta final



Aplicación de la propuesta final

Caso NVC:

Para este caso se consideraron los datos experimentales de la ganancia diaria de peso (GDP) en ovinos; datos experimentales vs. modelo de simulación para estimar la ganancia diaria de peso (GDP) Osorio (2011). Estos datos se encuentran en el apéndice B de Balam (2012).

Aplicación de la propuesta final: Caso NVC

```
PRUEBA DE NORMALIDAD
    Estadistica   ValorCal    pValor
W Shapiro-Wilk 0.98158509 0.8865120
D    Liliefort 0.07458593 0.9558153

Conclusión: Se cumple el supuesto de normalidad con Shapiro y Lilliefort al 5 %.

PRUEBA T-STUDENT PARA MEDIA CERO EN LOS RESIDUALES
  Estadistica      ValorCal pValor
t   T-Student -1.816261e-16      1

Conclusión: Se cumple el supuesto de media cero en los residuales al  5 %.

PRUEBA DE IGUALDAD DE VARIANZAS CON BREUSH-PAGAN
    Estadistica ValorCal    pValor
BP Breush-Pagan 1.067363 0.3015417

Conclusión: Se cumple el supuesto de varianza constante con el estadístico de Breush-Pagan al 5 %.

PRUEBA DE DURBIN-WATSON PARA INDEPENDENCIA
          Estadistica ValorCal    pValor
DW Durbin-Watson test 1.857982 0.2962045

Conclusión: Se cumple el supuesto de independencia con Durbin-Watson al 5 %.

 ****** CASO: NORMALIDAD - HOMOCEDASTICIDAD (NVC) ******

PRECISION (R2) CON EL ICB PERCENTIL-ESQUEMA BOOTSTRAP LIU2
       Atributos   Valores
1             R2 0.7273763
2    R2BootMedia 0.7168087
3 DesvEstR2Boots 0.1317316
4           LIR2 0.3842338
5           LSR2 0.8992233

Conclusión: El modelo es preciso con el ICB Percentil al  95 %.

 Criterio: Si el  0.7  está contenido en el ICB o es mayor igual al limite inferior del ICB.
```

Aplicación de la propuesta final

Caso NNVD:

Se aplicó a los datos experimentales del volumen de una parcela en metros cúbicos a un diámetro superior ($n = 63$) de 10cm y los simulados con el modelo PTAEDA Chung (1987), el cual es un modelo estocástico. Cada simulación con el modelo PTAEDA corresponde a la media de 10 corridas del modelo para cada parcela. En cada parcela se mide la edad, el indice de sitio y el número de arboles por hectárea. Estos datos se encuentran en el apéndice B de Balam (2012).

Aplicación de la propuesta final: Caso NNVD

```
PRUEBA DE NORMALIDAD
   Estadistica  ValorCal       pValor
W Shapiro-Wilk 0.9090287 0.0002003722
D    Liliefort 0.1395817 0.0038165826

Conclusión: No se cumple el supuesto de normalidad con Shapiro y Lilliefort al 5 %.

PRUEBA T-STUDENT PARA MEDIA CERO EN LOS RESIDUALES
  Estadistica       ValorCal pValor
t   T-Student -7.099852e-16      1

Conclusión: Se cumple el supuesto de media cero en los residuales al  5 %.

PRUEBA DE IGUALDAD DE VARIANZAS CON WHITE
   Estadistica ValorCal       pValor
BP       White 23.60013 1.185793e-06

Conclusión: No se cumple el supuesto de varianza constante con el estadístico de White al 5 %.

PRUEBA DE DURBIN-WATSON PARA INDEPENDENCIA
          Estadistica ValorCal    pValor
DW Durbin-Watson test 1.740505 0.1258351

Conclusión: Se cumple el supuesto de independencia con Durbin-Watson al 5 %.

 ****** CASO: NO NORMALIDAD - HETEROCIDASTICIDAD (NNVD) ******

PRECISION (R2) CON EL ICB BCa-ESQUEMA BOOTSTRAP PAREADO BALANCEADO
       Atributos     Valores
1             R2 0.83505020
2    R2BootMedia 0.84055856
3 DesvEstR2Boots 0.03684941
4           LIR2 0.74039251
5           LSR2 0.89089504

Conclusión: El modelo es preciso con el ICB BCa al 95 %.

 Criterio: Si el  0.7  está contenido en el ICB o es mayor igual al limite inferior del ICB.
```

1. Se propuso un método para evaluar la precisión de un modelo con la técnica de regresión lineal ; basado en ocho esquemas de remuestreo, dos tipos de modelo y seis tamaños de muestra; a través de dos ICB para el coeficiente de determinación R^2 y se consideraron los cuatro escenarios posibles (NVC, NNVC, NVD y NNVD) ante el cumplimiento o no normalidad y/o varianza constante.
2. Se realizó un estudio de simulación para comparar las eficiencias de los intervalos de confianza para cada tipo de supuesto con respecto a los diferentes esquemas Bootstrap, tamaños de muestra y tipo de modelo.
3. Análisis estadísticos: se determinó con al menos un 95% que para los supuestos NVC, NNVC o NVD se utilice el ICB Percentil con el esquema de remuestreo Liu2 sin importar el tamaño de muestra. Y se determinó con al menos el 88.8% que para el supuesto NNVD se utilice el ICB BCa con el esquema de remuestreo pareado balanceado.

4. Propuesta final: para los supuestos NVC, NNVC o NVD se utilice el ICB Percentil con el esquema de remuestreo Liu2 (NVC: residuales de regresión lineal simple, NNVC: residuales robustos sin ponderar y NVD: residuales robustos ponderados) y para el supuesto NNVD se utilice el ICB BCa con el esquema de remuestreo Pareado Balanceado
5. Se aplicó la propuesta final, para la ganancia diaria de peso en ovinos, el modelo resultó ser de tipo NVC y preciso, coincidiendo con Balam (2012), cabe señalar que usó ICB BCa con residuales balanceados y en este trabajo se usó ICB Percentil con el esquema de remuestreo Liu2. Para el volumen por parcela, el modelo resultó de tipo NNVD y preciso, coincidiendo con Balam (2012), tanto en la decisión como en el esquema e ICB que utilizó.

Trabajos a futuro:

1. Desarrollar una librería en el lenguaje R que contenga la propuesta de este trabajo para la evaluación de la precisión, junto con la propuesta desarrollada por Zacarías (2023) para la evaluación de la exactitud y de esta manera tener una herramienta integral para la evaluación de un modelo con la técnica de regresión lineal.
2. Integrar la propuesta al Sistema de Validación de Modelos (Mazún, 2014).
3. Evaluar otros ICB que requieren cómputos más exhaustivos pero utilizando la programación en paralelo.

Referencias

Mazún, E. (2014). *Diseño e implementación en el svmrl del método de freese con sus extensiones para validad modelos en web* [Tesis de Licenciatura]. UADY, Facultad de Matemáticas.

Zacarías, K. (2023). *Evaluación de la exactitud de un modelo cuando no se cumplen los supuestos en la técnica de regresión lineal* [Tesis de Maestría]. UADY, Facultad de Matemáticas.

Muchas Gracias