

UNIVERSIDAD
NACIONAL
DE COLOMBIA

Universidad Nacional de Colombia

Disparidad en Tiempos de Reparación: Un Análisis entre Verizon
(ILEC) y Competidores (CLEC)

Autor:

Juan Daniel Diaz Alvarez
juadiazal@unal.edu.co

Juan Sosa
Ph.D Estadística

Noviembre del 2023

Índice

1. Introducción	2
2. Análisis Bayesiano	3
2.1. Distribución posterior de η	3
2.2. Análisis de sensibilidad	4
2.3. Bondad de ajuste	6
3. Análisis Frecuentista	7
4. Simulación	8

1. Introducción

Verizon es la principal empresa de telecomunicaciones en una amplia zona del este de Estados Unidos, conocida como ILEC (Incumbent Local Exchange Carrier). Como ILEC, tiene la responsabilidad de ofrecer servicios de reparación a los clientes de otras compañías telefónicas que compiten en esta región, denominadas CLEC (Competing Local Exchange Carrier).

Verizon está sujeto a sanciones si los tiempos de reparación para los clientes de CLEC son considerablemente peores que los tiempos de reparación para los clientes de Verizon. Para evaluar esta situación, se ha recopilado un conjunto de datos de Verizon que incluye muestras de los tiempos de reparación para dos grupos: uno con 1664 clientes de Verizon (ILEC) y otro con 23 clientes de la competencia (CLEC).

El objetivo es determinar si existe una diferencia significativa entre los tiempos promedio de reparación de estos dos grupos, lo que podría considerarse como evidencia para tomar medidas, como imponer sanciones a Verizon. Para cumplir con el objetivo de estudio, se propone el análisis en tres momentos claves:

1. **Análisis Bayesiano.** En este momento consideraremos lo siguiente:

$$y_{k,i} \stackrel{\text{iid}}{\sim} \text{Exp}(\lambda_k) \quad y_{k,i} > 0, \quad \lambda_k > 0$$

$$\lambda_k \sim \text{GI}(a_k, b_k)$$

Donde $i = 1, \dots, n_k$ y $k = 1, 2$. Los valores de k corresponden a ILEC y CLEC, respectivamente. Por lo tanto, la distribución posterior:

$$\lambda_k | \mathbf{y}_k \sim \text{GI}(a_k + n_k, b_k + s_k)$$

Donde a_k y b_k son hiperparámetros conocidos para cada grupo y $s_k = \sum_{i=1}^{n_k} y_{k,i}$. En consecuencia, la inferencia sobre la diferencia entre los tiempos promedio de reparación sería alrededor de:

$$\eta = \lambda_1 - \lambda_2$$

2. **Análisis frecuentista.** En este momento, vamos a considerar *Bootstrap* paramétrico, *Bootstrap* no paramétrico y la teoría asintótica de MLE. Para esta última tengase en cuenta lo siguiente:

$$\hat{\lambda}_{k,\text{MLE}} = \bar{y}_k, \quad \bar{y}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} y_{k,i}$$

$$\hat{\eta}_{\text{MLE}} = \hat{\lambda}_{1,\text{MLE}} - \hat{\lambda}_{2,\text{MLE}}$$

Dado la teoría asintótica de la estimación por MLE, $\lambda_k \overset{A}{\sim} N(\hat{\lambda}_k, I_k^{-1})$. Por lo tanto, la distribución asintótica de $\hat{\eta}_{MLE}$ es:

$$\hat{\eta}_{k,MLE} \overset{A}{\sim} N\left(\hat{\lambda}_1 - \hat{\lambda}_2, \frac{\bar{y}_1^2}{n_1} + \frac{\bar{y}_2^2}{n_2}\right)$$

3. Simulación de poblaciones exponenciales, en diferentes escenarios aplicando el método Bayesiano, teoría asintótica de MLE, *Bootstrap* paramétrico y no paramétrico.

2. Análisis Bayesiano

2.1. Distribución posterior de η

Ajustaremos un modelo Gamma-Inversa-Exponencial, presentado en la Introducción, para esto se propuso como hiperparámetros $a_k = 3$ y $b_k = 17$, para todo valor de k . Por lo tanto, ajustado las distribuciones para λ_k , se obtiene la distribución de η , así como estimaciones puntuales, como la media, el coeficiente de variación - CV y el intervalo de credibilidad - IC al 95%. Por lo tanto, se obtienen los siguientes resultados:

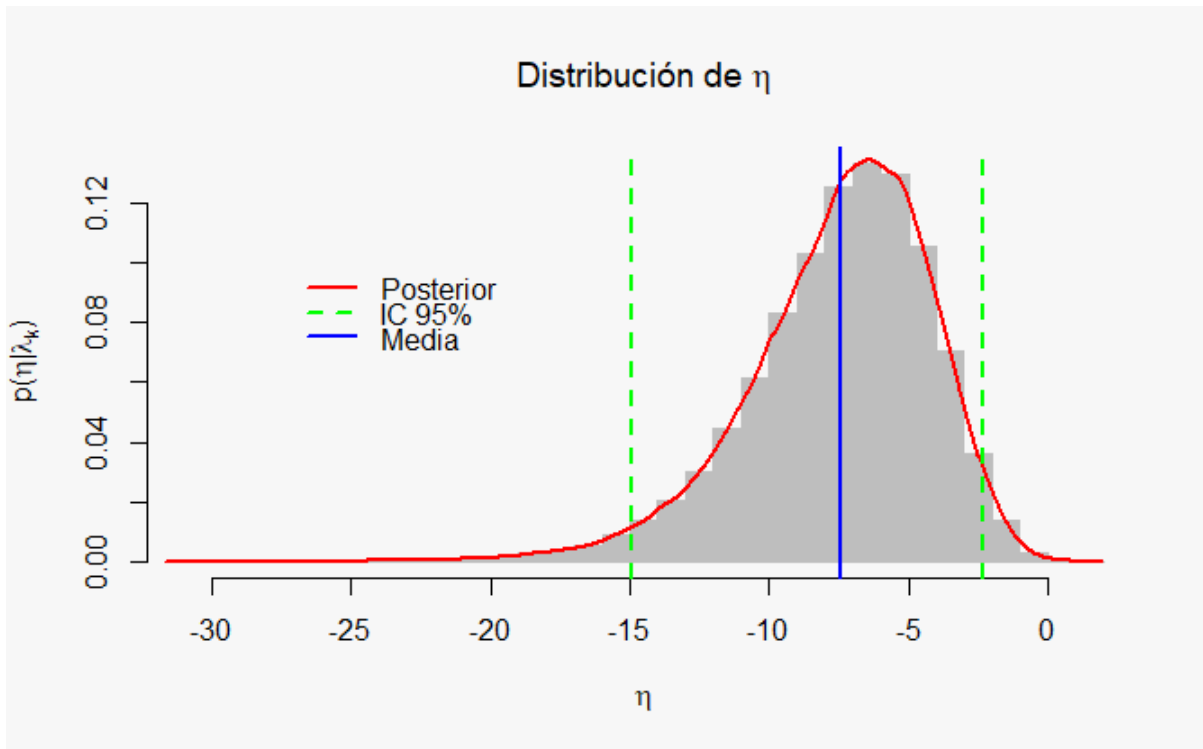


Figura 1: Distribución posterior η

Media	CV %	IC Inf.	IC Sup.
-7.462	43.5	-14.960	-2.316

Tabla 1: Estadísticas de η

Basándonos en la Figura 1 y la Tabla 1, se concluye que el tiempo promedio de reparación de los clientes CLEC es significativamente mayor que el de Verizon. La diferencia media es de -7.46, lo que implica que CLEC tiene un tiempo de reparación 7.46 unidades mayor que Verizon. Además, se observa una alta variabilidad, con una variación relativa del 43.5 %. Esto se refleja en el Intervalo de Credibilidad, donde, con un 95 % de confianza, el valor real de η cae en el rango (-14.960, -2.316). Por lo tanto, se debería multar a Verizon dado que los tiempos medios de reparación son peores para los clientes de alguna CLEC.

2.2. Análisis de sensibilidad

Se realizará un análisis de sensibilidad considerando, los siguientes estados información externa al conjunto de datos. Para cada estado de información calcularemos la media y el coeficiente de variación - CV, de manera previa y posterior; además el Intervalo de Credibilidad - IC para la posterior.

- Estado de información - Previa 1: $a_k = 3$ y $b_k = 17$, para $k = 1, 2$
- Estado de información - Previa 2: $a_k = 2$ y $b_k = 8.5$, para $k = 1, 2$
- Estado de información - Previa 3: $a_k = 3$ y $b_1 = 16.8$ y $b_2 = 33$, para $k = 1, 2$
- Estado de información - Previa 4: $a_k = 2$ y $b_1 = 8.4$ y $b_2 = 16.5$, para $k = 1, 2$

Las estadísticas para la información previa está dada por la Tabla a continuación.

Estadísticas	Previa 1	Previa 2	Previa 3	Previa 4
Media	0	0	-8.1	-8.1
CV %	Ind.	Ind.	228.58	Ind.

Tabla 2: Media y CV a priori

Tras ajustar los hiperparámetros en el análisis Bayesiano de tiempos de reparación entre Verizon (ILEC) y competidores (CLEC), se observa que para la previa 1 y 2 la diferencia del tiempo medio es 0, es decir, que ILEC y CLEC tienen el mismo tiempo medio de reparación, lo cual contrasta con la información que proporciona la previa 3 y 4, las cuáles indican que los tiempos de reparación de

CLEC son mayores que las de ILEC. Las indeterminaciones (Ind.) en los coeficientes de variación a priori de Previa 1, 2 y 4 indican una incertidumbre considerable en estas estimaciones, al igual que el coeficiente de variación de la Previa 3, que si bien se conoce, la incertumbre sobre la estimación es sustancial.

Los resultados de las estimaciones de η para cada estado de información es:

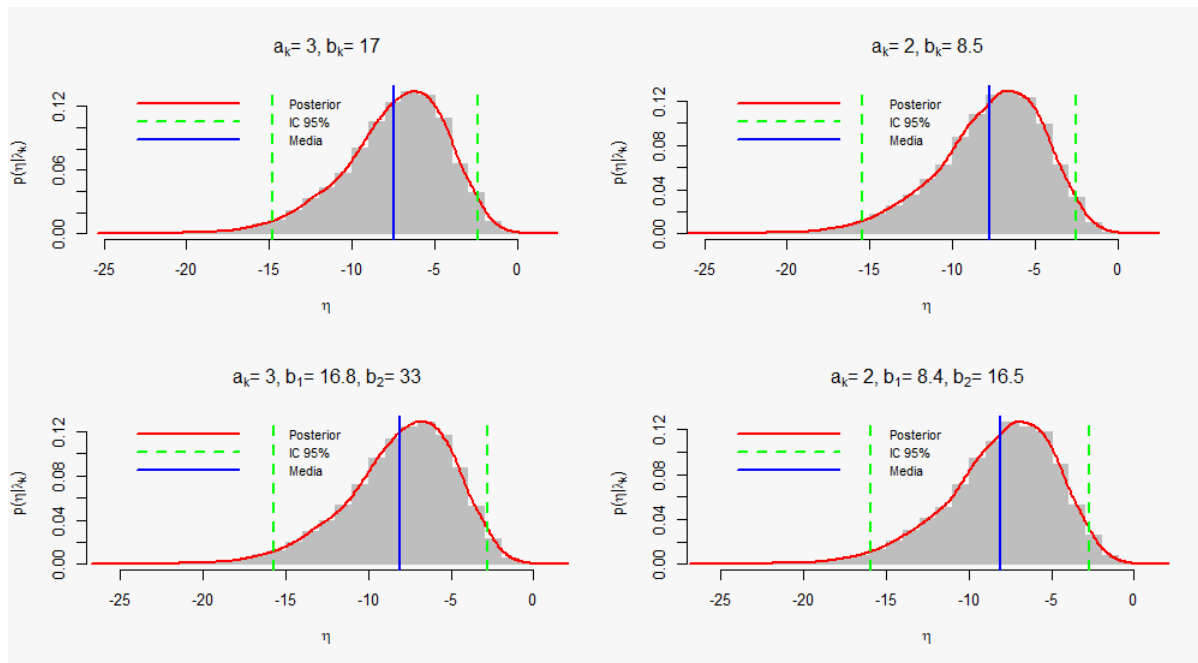


Figura 2: Distribuciones Posteriores

Previas	Medias	CV%	IC Inf	IC Sup
Previa 1	-7.449	43.3	-14.795	-2.345
Previa 2	-7.759	43.3	-15.475	-2.488
Previa 3	-8.089	41.5	-15.731	-2.778
Previa 4	-8.092	42.4	-15.969	-2.710

Tabla 3: Estadísticas distribución posterior

Las distribuciones posteriores revelan diferencias significativas en los tiempos de reparación entre ILEC y CLEC, respaldadas por IC del 95%. Las diferencias observadas sugieren que los tiempos de reparación de CLEC son, en promedio, más largos que los de ILEC. Además, el modelo es robusto dado que el estado de información previo, no está influyendo de manera sustancial en las inferencias de η .

2.3. Bondad de ajuste

En cada población, ILEC y CLEC, evaluaremos la bondad de ajuste con el estado de información 1 (previa 1), usando como estadísticos de pruebas la media y la desviación estándar.

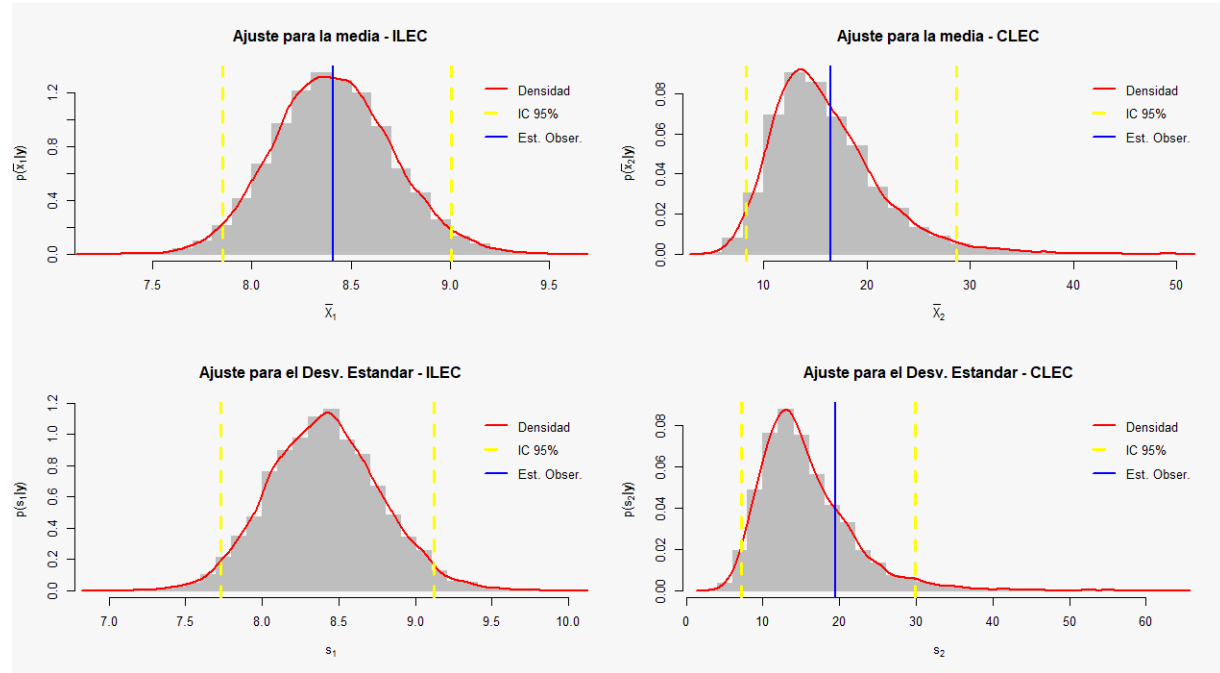


Figura 3: Densidad bondad de ajuste para \bar{X}_k y S_k

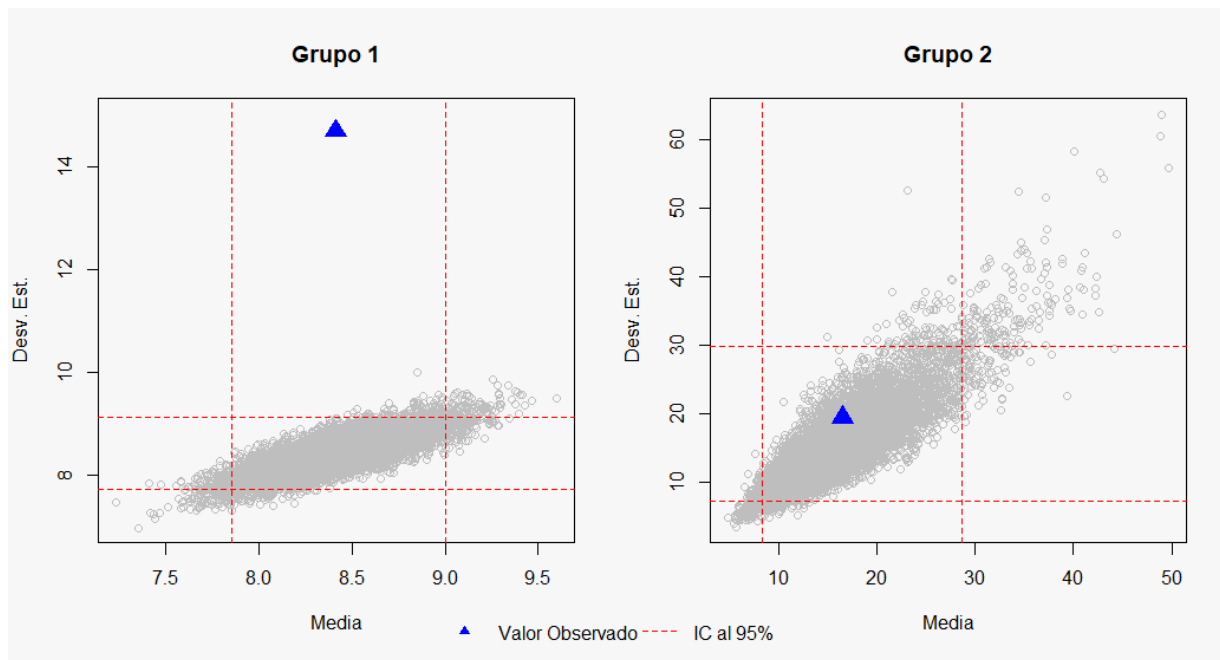


Figura 4: Dispersograma por grupos de la bondad de ajuste

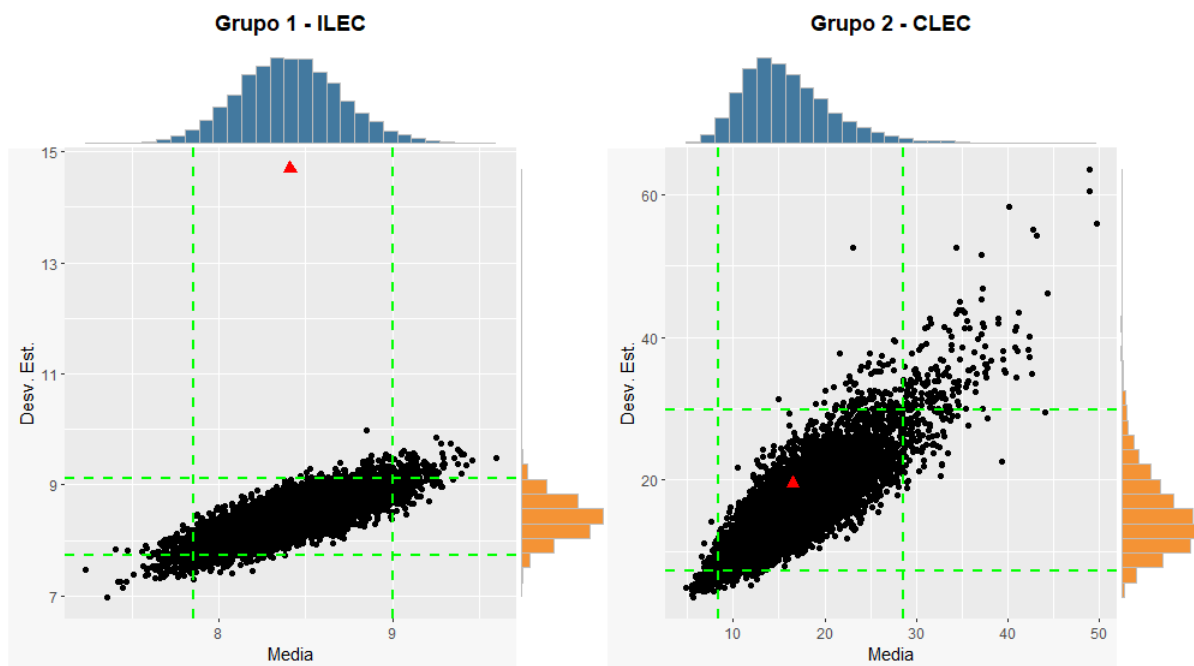


Figura 5: Distribución predictiva de los estadístico de prueba de manera conjunta

Grupo	SD Obs.	Media Obs.	PPP Media	PPP SD
ILEC	14.690	8.412	0.511	1.000
CLEC	19.504	16.509	0.607	0.800

Tabla 4: Estadísticas de la Bondad de Ajuste

Para CLEC se obtiene que tanto la media como el desvío estandar están siendo estimados correctamente dado que el valor p predictivo posterior - PPP se encuentra cercano a 0.5 y su valor observado está contenido en la región de credibilidad. Acto que solo sucede para la media de ILEC, ya que para el desvío estandar el PPP es 1 y el valor observado se encuentra en la cola derecha de la distribución, en consecuencia el modelo subestimada el desvío estandar para ILEC.

3. Análisis Frecuentista

Con el estado de información de la Sección 2.1, se realizará análisis, usando la Normalidad asintótica del MLE, descrito en la Introducción, además del *Bootstrap* paramétrico, no paramétrico y por supuesto el Bayesiano. Para esto consideremos un total de 10000 replicaciones, por lo que se obtuvieron los resultados descritos a continuación:

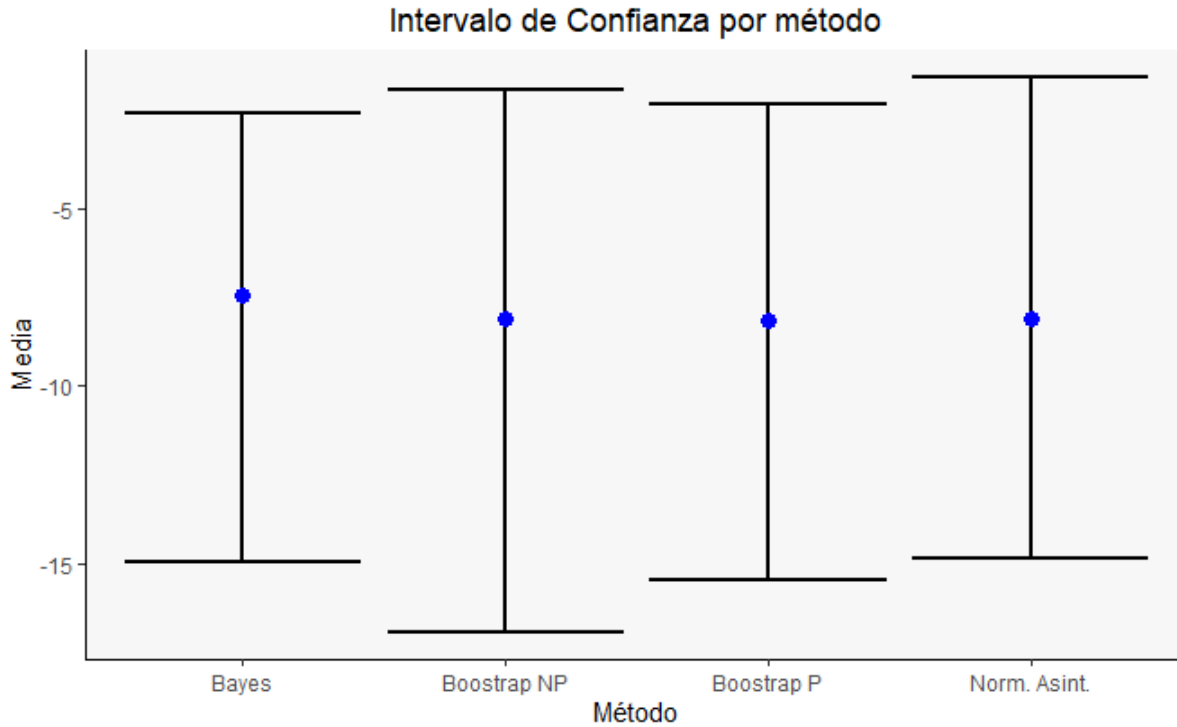


Figura 6: IC y , estado de información 1

Metodo	Media	CV	IC Inf.	IC Sup.
Bayes	-7.462	0.435	-14.960	-2.316
Norm. Asint.	-8.098	0.426	-14.857	-1.338
Bootstrap P	-8.172	0.423	-15.490	-2.057
Bootstrap NP	-8.111	0.492	-16.929	-1.662

Tabla 5: Análisis frecuentista y Bayesiano

Con la información de la Figura 6 y Tabla 5, es preciso mencionar la diferencia promedio de tiempo de reparación sigue siendo mayor para CLEC, además que se tiene una variabilidad relativa considerable. Por otro lado, la diferencia promedio y del coeficiente de variación no varían sustancialmente, se mantienen en el mismo rango de valor para cada método. Así también, sus intervalos de confianza y de credibilidad.

4. Simulación

A continuación simularemos 100000 muestras exponenciales teniendo en cuenta los escenarios y ajustaremos un modelo de manera Bayesiana y Frecuentista (Normalidad asintótica, *Bootstrap*

paramétrico y no paramétrico) y calcularemos la proporción de veces que el Intervalo de Confianza/Credibilidad 95 % contiene el valor real:

- Escenario 1: $n_1 = 10, n_2 = 10, \lambda_1 = \bar{y}_1$ y $\lambda_2 = \bar{y}_2$
- Escenario 2: $n_1 = 20, n_2 = 20, \lambda_1 = \bar{y}_1$ y $\lambda_2 = \bar{y}_2$
- Escenario 3: $n_1 = 50, n_2 = 50, \lambda_1 = \bar{y}_1$ y $\lambda_2 = \bar{y}_2$
- Escenario 4: $n_1 = 100, n_2 = 100, \lambda_1 = \bar{y}_1$ y $\lambda_2 = \bar{y}_2$

Ajustado los modelos correspondientes, definido el valor real de $\eta = \bar{y}_1 - \bar{y}_2$, se obtuvieron los siguientes resultados:

Escenarios	n	Bayes	Norm. Asint	Bootstrap Par.	Bootstrap No Par.
1	10	0.930	0.945	0.938	0.890
2	20	0.938	0.948	0.943	0.917
3	50	0.944	0.949	0.946	0.935
4	100	0.946	0.950	0.947	0.942

Tabla 6: Cobertura empírica por método y tamaño de la muestra

Con la información de la Tabla 6, podemos observar que a medida que aumenta el tamaño de la muestra (de 10 a 100), la cobertura empírica tiende a mejorar en todos los métodos. Esto es consistente con lo que esperaríamos, ya que con muestras más grandes, los estimadores y los intervalos de confianza suelen ser más precisos. Ahora bien, de los 4 escenarios la cobertura más robusta es el de la normalidad sintótica, seguido del *Bootstrap* paramétrico, los cuáles se comportan muy similar. En tercer lugar se tiene, al método Bayesiano y por último el *Bootstrap* no paramétrico el cuál inicia con la cobertura más baja que los demás métodos con un 89 %.