**Primer Trabajo Estadística II**

Inferencia Estadística

**Integrantes:**

1. Juan David Mena Gamboa
2. Miguel Angel Bolaño López
3. Heyner David Marquez Garnica

* **Fuente de los datos (url):**
* **Descripción de los datos seleccionados:**
* **Variables seleccionadas:**

|  |  |  |
| --- | --- | --- |
| **Variables Cualitativas** | | |
| **Nombre Variable** | | **Categorías o Niveles** |
| **1** | P3101 |  |

|  |  |  |
| --- | --- | --- |
| **Variables Cuantitativas** | | |
| **Nombre Variable** | | **Unidad de medición** |
| **1** | P3094S3 |  |
| **2** | P3087S1 |  |
| **3** | P3095S3 |  |

**Desarrollo**

**Variable 1**

* ***Análisis Descriptivo***

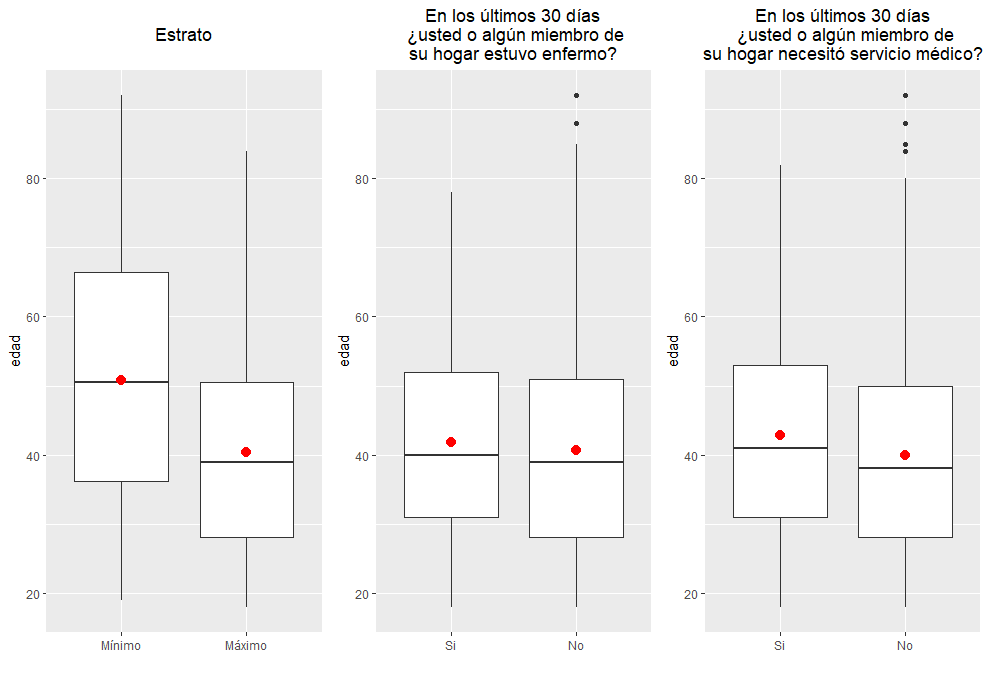
*Descriptivos Básicos*

|  |  |  |
| --- | --- | --- |
|  | Medida | Valor |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |
| 6 |  |  |
| 7 |  |  |
| 8 |  |  |
| 9 |  |  |
| 10 |  |  |

Comentarios:

*Análisis Gráfico*

Utilice las gráficas que requiera para el análisis, la figura aquí mostrada es sólo un ejemplo



Comentarios:

* ***Cálculo de los estimadores***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Estimador** | **Estimadores Puntuales** | | **Estimadores por Intervalo** | |
| **Analogía** | **Máxima Verosimilitud** | **Límite Inferior** | **Límite Superior** |
| (Media o Proporción) | Escriba aquí el resultado obtenido | Escriba aquí el resultado obtenido | Escriba aquí el resultado obtenido | Escriba aquí el resultado obtenido |
| Comentario | Escriba aquí el comentario | Escriba aquí el comentario | Escriba aquí el comentario | Escriba aquí el comentario |

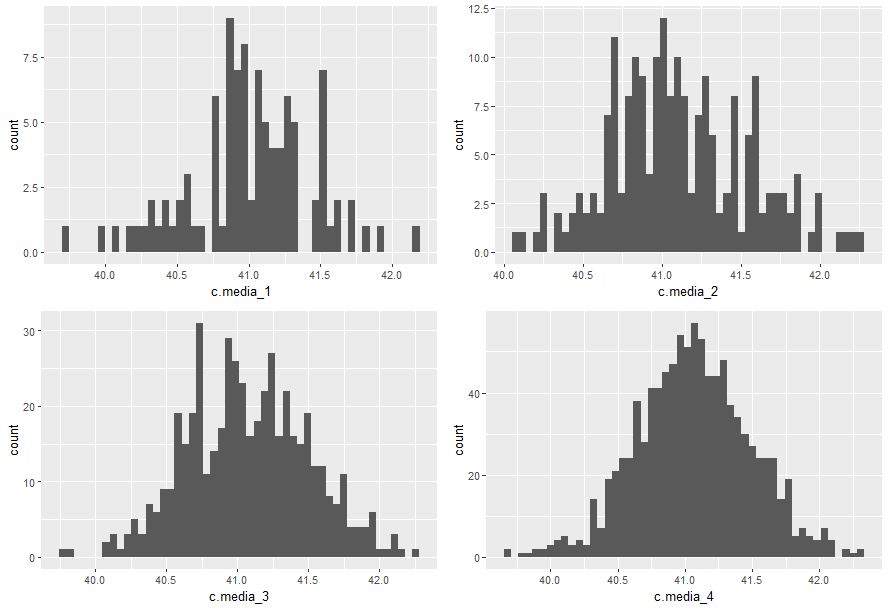
* ***Evaluación del estimador:***

*Insesgamiento*

|  |  |  |
| --- | --- | --- |
| **Media** | **Mediana** | **Sesgo** |
| Escriba aquí el resultado Obtenido | Escriba aquí el resultado Obtenido | Escriba aquí el resultado Obtenido |

Comentario:

*Consistencia*



Comentario:

*Eficiencia*

|  |  |
| --- | --- |
| **Medida** | **Valor** |
| Media |  |
| Mediana |  |

Comentario:

Sintaxis empleada con esta variable:

Coloque aquí la sintaxis empleada para procesar esta variable

**Variable 2**

* ***Análisis Descriptivo***

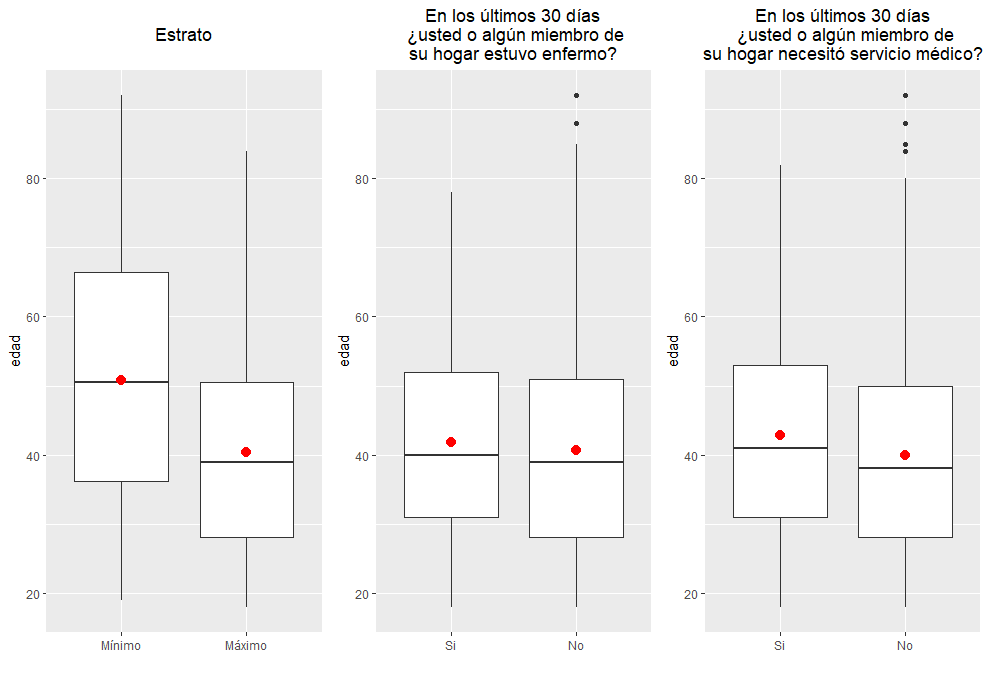
*Descriptivos Básicos*

|  |  |  |
| --- | --- | --- |
|  | Medida | Valor |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |
| 6 |  |  |
| 7 |  |  |
| 8 |  |  |
| 9 |  |  |
| 10 |  |  |

Comentarios:

*Análisis Gráfico*

Utilice las gráficas que requiera para el análisis, la figura aquí mostrada es sólo un ejemplo



Comentarios:

* ***Cálculo de los estimadores***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Estimador** | **Estimadores Puntuales** | | **Estimadores por Intervalo** | |
| **Analogía** | **Máxima Verosimilitud** | **Límite Inferior** | **Límite Superior** |
| (Media o Proporción) | Escriba aquí el resultado obtenido | Escriba aquí el resultado obtenido | Escriba aquí el resultado obtenido | Escriba aquí el resultado obtenido |
| Comentario | Escriba aquí el comentario | Escriba aquí el comentario | Escriba aquí el comentario | Escriba aquí el comentario |

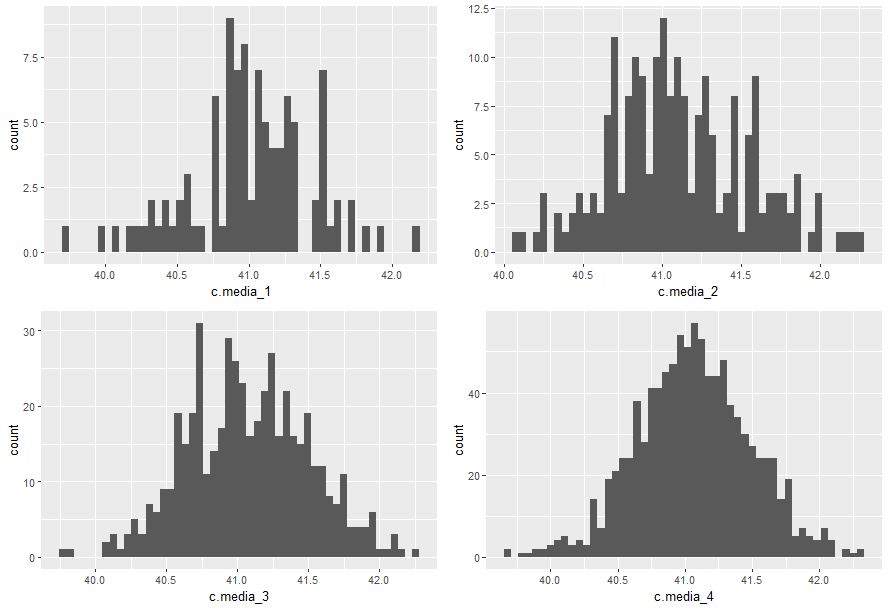
* ***Evaluación del estimador:***

*Insesgamiento*

|  |  |  |
| --- | --- | --- |
| **Media** | **Mediana** | **Sesgo** |
| Escriba aquí el resultado Obtenido | Escriba aquí el resultado Obtenido | Escriba aquí el resultado Obtenido |

Comentario:

*Consistencia*



Comentario:

*Eficiencia*

|  |  |
| --- | --- |
| **Medida** | **Valor** |
| Media |  |
| Mediana |  |

Comentario:

Sintaxis empleada con esta variable:

Coloque aquí la sintaxis empleada para procesar esta variable

**Variable 3**

* ***Análisis Descriptivo***

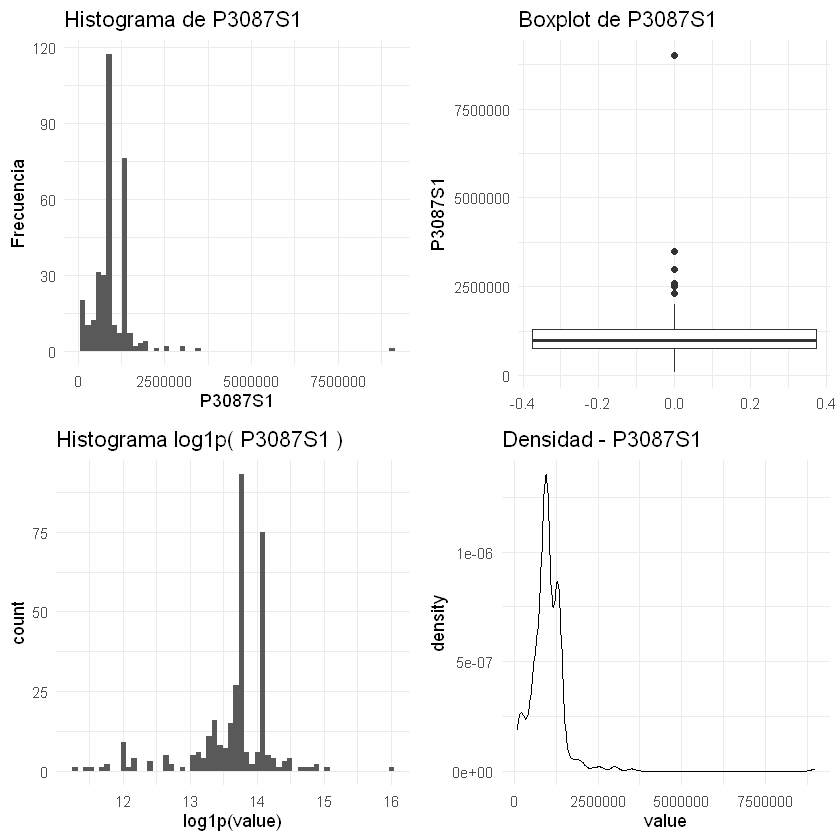
*Descriptivos Básicos*

|  |  |  |
| --- | --- | --- |
|  | Medida | Valor |
| 1 | Registros | 166341 |
| 2 | Valores faltantes | 166005 |
| 3 | Mínimo | 80000 |
| 4 | Primer cuartil | 750000 |
| 5 | Mediana | 975000 |
| 6 | Media | 998708 |
| 7 | Tercer cuartil | 1300000 |
| 8 | Máximo | 9000000 |
| 9 | Desviación Estándar | 621977 |
| 10 | Rango intercuartílico | 550000 |

Comentarios:

El conjunto de datos presenta una **alta proporción de valores faltantes** (más del 99%), y entre los datos válidos se observa **gran dispersión** y **valores atípicos elevados** que influyen en la media. La mediana es una mejor representación del valor típico, ya que no se ve tan afectada por los extremos.

*Análisis Gráfico*



Comentarios:

Las visualizaciones muestran que la variable presenta una **distribución altamente asimétrica a la derecha**, con varios **valores atípicos extremos** evidenciados en el boxplot. El histograma y la curva de densidad confirman una concentración significativa de observaciones en valores bajos, mientras que la transformación logarítmica (log1p) reduce la dispersión, revelando una distribución más cercana a la normalidad.

* ***Cálculo de los estimadores***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Estimador** | **Estimadores Puntuales** | | **Estimadores por Intervalo** | |
| **Analogía** | **Máxima Verosimilitud** | **Límite Inferior** | **Límite Superior** |
| (Media o Proporción) | 998,708.3 | 998,708.3 | 931,962.4 | 1,065,454 |
| Comentario | La media es el valor promedio de los datos, representando el centro de la distribución. | El estimador de máxima verosimilitud para la media coincide con la media muestral, ya que bajo normalidad ambos son equivalentes. | El límite inferior del IC al 95% indica que, con alta confianza, la media poblacional no es menor que este valor | El límite superior del IC al 95% establece que, con alta confianza, la media poblacional no excede este valor. |

* ***Evaluación del estimador:***

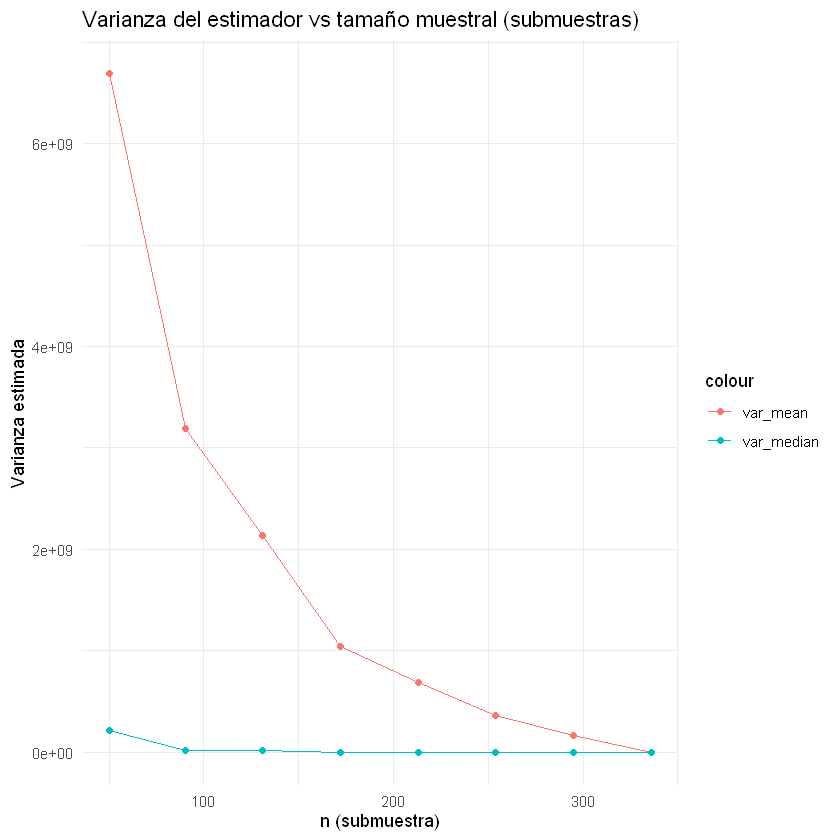
*Insesgamiento*

|  |  |  |
| --- | --- | --- |
| **Media** | **Mediana** | **Sesgo** |
| 996,995.5 | 975,000 | Media: sesgo ≈ +333.15  Mediana: sesgo ≈ -282.5 |

Comentario:

Tanto la media como la mediana muestran **sesgos pequeños y poco significativos** respecto a sus verdaderos valores poblacionales, ambos estimadores son **casi insesgados**, lo que respalda su fiabilidad en el análisis.

*Consistencia*



Comentario:

Se observa una disminución notable de la varianza de ambos estimadores a medida que aumenta el tamaño muestral, lo que confirma que muestras más grandes proporcionan estimaciones más estables. Sin embargo, la varianza de la media es significativamente mayor en tamaños de muestra pequeños y desciende de forma más lenta, lo que evidencia su sensibilidad a valores extremos presentes en los datos, estos resultados sugieren que, para este conjunto de datos con alta dispersión, la mediana es un estimador más confiable y eficiente que la media.

*Eficiencia*

|  |  |
| --- | --- |
| **Medida** | **Valor** |
| Media | 1120377604 |
| Mediana | 0.001749369 |

Comentario:

La mediana es más eficiente en esta muestra.

Sintaxis empleada con esta variable:

1) Limpieza money: clean\_numeric2(...)

2) Descriptivos: tabla\_vertical (N, N missing, Min, Q1, Median, Mean, Q3, Max, SD, IQR)

3) Gráficas guardadas en: analisis\_P3087S1\_check/plots

4) IC mean (t): t.test(x)$conf.int (o calculado manualmente)

5) Bootstrap bias y CI: uso package 'boot' (boot(...), boot.ci(...))

6) Consistencia: submuestreo comparando var(estimator) vs n

7) Eficiencia: comparar var\_bootstrap(estimator)

**Variable 4**

* ***Análisis Descriptivo***

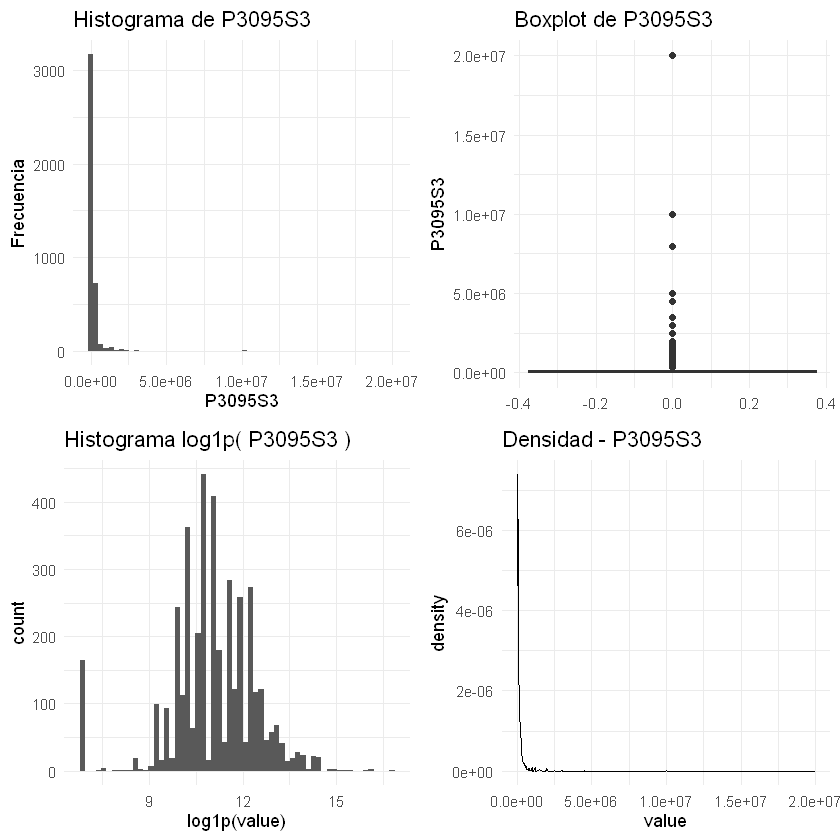
*Descriptivos Básicos*

|  |  |  |
| --- | --- | --- |
|  | Medida | Valor |
| 1 | Registros | 166341 |
| 2 | Valores faltantes | 162261 |
| 3 | Mínimo | 1000 |
| 4 | Primer cuartil | 30000 |
| 5 | Mediana | 60000 |
| 6 | Media | 159471 |
| 7 | Tercer cuartil | 150000 |
| 8 | Máximo | 20000000 |
| 9 | Desviación Estándar | 512519 |
| 10 | Rango intercuartílico | 120000 |

Comentarios:

El conjunto de datos contiene 166,341 registros, de los cuales 162,261 presentan valores faltantes, evidenciando una disponibilidad limitada de información. Los valores observados oscilan entre 1,000 y 2,000,000, con una media de 159,471 y mediana de 60,000, lo que refleja una distribución asimétrica positiva influenciada por valores extremos. El rango intercuartílico (120,000) indica una alta dispersión incluso en los datos centrales, mientras que la desviación estándar (512,519) confirma una marcada variabilidad. Se sugiere aplicar técnicas robustas o transformaciones para su análisis.

*Análisis Gráfico*



Comentarios:

La variable **P3095S3** presenta una distribución altamente sesgada a la derecha con numerosos valores atípicos. Tras aplicar la transformación logarítmica (log1p), los datos muestran una distribución más equilibrada y adecuada para análisis estadísticos y modelado.

* ***Cálculo de los estimadores***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Estimador** | **Estimadores Puntuales** | | **Estimadores por Intervalo** | |
| **Analogía** | **Máxima Verosimilitud** | **Límite Inferior** | **Límite Superior** |
| (Media o Proporción) | 159,471.4 | |  | | --- | |  |  |  | | --- | | IC 95% t: [143,740.4, 175,202.4] | | 143,740.4 | |  | | --- | |  |  |  | | --- | | 175,202.4 | |
| Comentario | Muestra una amplia dispersión, lo que refleja variabilidad. | Con un 95% de confianza, la media poblacional se encuentra entre los límites estimados. | Es poco probable que la media real sea menor que este valor. | Valor máximo estimado de la media, reforzando la amplitud del rango y la posible variabilidad de los datos. |

* ***Evaluación del estimador:***

*Insesgamiento*

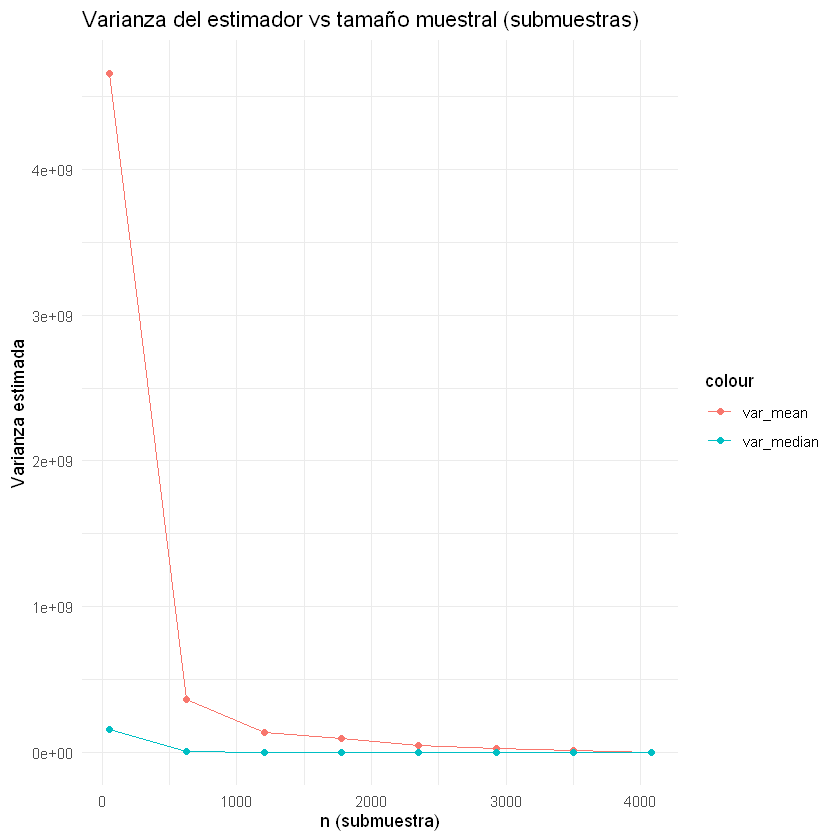
|  |  |  |
| --- | --- | --- |
| **Media** | **Mediana** | **Sesgo** |
| 159172.9 | 60000 | |  | | --- | |  |  |  | | --- | | Media: 166.16  Mediana: 0 | |

Comentario:

La media muestra sesgo pequeño relativo.

La mediana muestra sesgo pequeño relativo.

*Consistencia*



Comentario:

La varianza del estimador (mean o median) disminuye al crecer n, evidenciando consistencia.

*Eficiencia*

|  |  |
| --- | --- |
| **Medida** | **Valor** |
| Media | 64150092 |
| Mediana | 0 |

Comentario:

La mediana es más eficiente en esta muestra.

Sintaxis empleada con esta variable:

Coloque aquí la sintaxis empleada para procesar esta variable

1) Limpieza money: clean\_numeric2(...)

2) Descriptivos: tabla\_vertical (N, N missing, Min, Q1, Median, Mean, Q3, Max, SD, IQR)

3) Gráficas guardadas en: analisis\_P3095S3\_check/plots

4) IC mean (t): t.test(x)$conf.int (o calculado manualmente)

5) Bootstrap bias y CI: uso package 'boot' (boot(...), boot.ci(...))

6) Consistencia: submuestreo comparando var(estimator) vs n

7) Eficiencia: comparar var\_bootstrap(estimator)