**Primer Trabajo Estadística II**

Inferencia Estadística

**Integrantes:**

1. [Nombre]
2. [Nombre]
3. [Nombre]

* **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 | COP |
| **2** | P3087S1 | COP |
| **3** | P3095S3 | COP |

**Desarrollo**

**Variable 1**

* ***Análisis Descriptivo***

*Descriptivos Básicos*

|  |  |  |
| --- | --- | --- |
|  | Medida | Valor |
| 1 | N total | 166 341 COP |
| 2 | N missing | 163 449 COP |
|  | Primer Cuartil | 1000 COP |
| 4 | Tercer Cuartil | 30 000 COP |
| 5 | Mediana | 60 000 COP |
| 6 | Media | 126 938 COP |
| 7 | Mínimo | 150 000 COP |
| 8 | Máximo | 20 000 000 COP |
| 9 | Desviación Estándar | 410 083 COP |
| 10 | Rango Intercuartílico | 120 000 COP |

Comentarios:

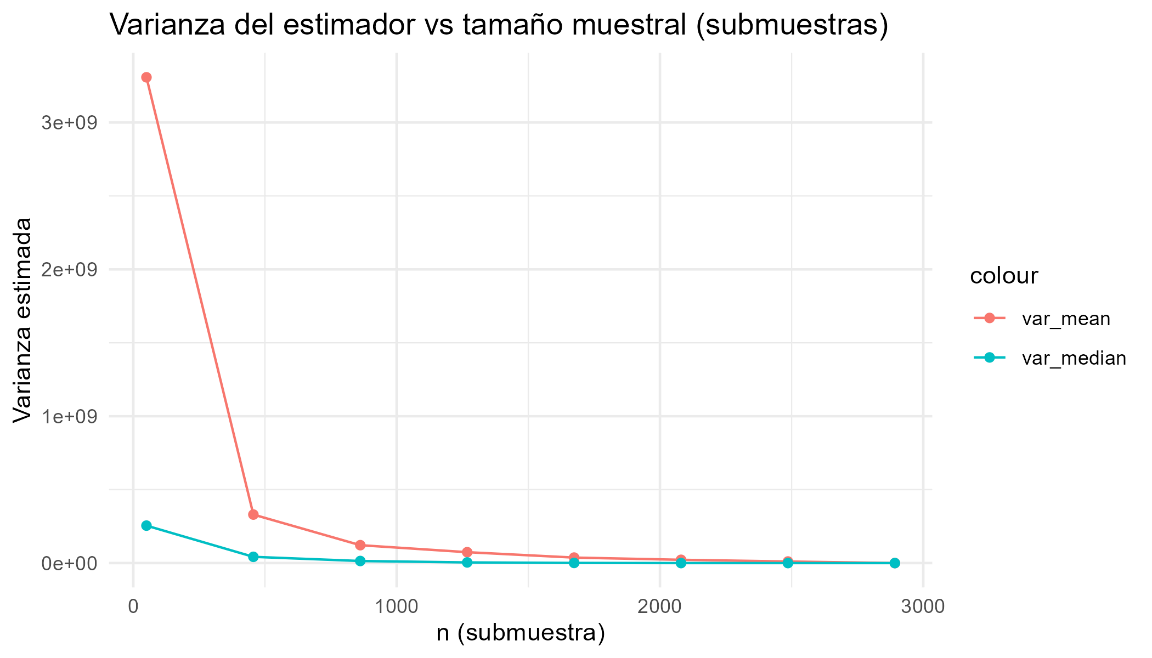
En la variable **P3094S3** (“¿Cuánto tendría que pagar si tuviera que comprar este(os) artículo(s)?”) se registraron **166,341 observaciones,** de las cuales **163,449 corresponden a valores faltantes**. Esto implica que únicamente **2,892 hogares respondieron efectivamente la pregunta,** por lo que el análisis debe interpretarse con cautela debido al alto nivel de no respuesta.

Entre los valores válidos, se observa que el **costo mínimo reportado es de 1,000 pesos**, mientras que el **máximo asciende a 20 millones de pesos**, lo cual evidencia una gran amplitud en el rango y sugiere la presencia de **valores atípicos**. La **mediana se ubica en 60,000 pesos**, indicando que la mitad de los hogares considera montos iguales o inferiores a esta cifra. En contraste, la **media es de 126,938 pesos,** un valor más del doble de la mediana, lo que revela una distribución **asimétrica y sesgada a la derecha**: unos pocos artículos de muy alto valor elevan el promedio.

El **rango intercuartílico de 120,000 pesos** muestra que el 50% central de las respuestas se concentra entre 30,000 y 150,000 pesos, lo que refleja una dispersión moderada dentro de la mayoría de los casos. Sin embargo, la **desviación estándar (410,083 pesos)**, muy superior a la media, refuerza la idea de una alta variabilidad inducida por los valores extremos.

En síntesis, los resultados sugieren que, aunque existen artículos reportados con costos muy elevados, el patrón más representativo se encuentra alrededor de los **60,000 pesos**. Por ello, en este contexto, la **mediana resulta un indicador más adecuado de la tendencia central** que la media.

*Análisis Gráfico*



Comentarios:

Los gráficos muestran que la variable **P3094S3** (COP) presenta una alta varianza inicial, debido a la presencia de valores extremos. A medida que aumenta el número de observaciones, tanto la media como la varianza se estabilizan, indicando mayor consistencia en los datos. Además, se observa una diferencia clara entre los grupos “SI” y “NO”, donde el grupo “SI” reporta ingresos considerablemente más altos. Esto sugiere una distribución asimétrica y la necesidad de considerar medidas como la mediana para representar el ingreso típico.

* ***Cálculo de los estimadores***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Estimador** | **Estimadores Puntuales** | | **Estimadores por Intervalo** | |
| **Analogía** | **Máxima Verosimilitud** | **Límite Inferior** | **Límite Superior** |
| Media | 126,937.6 | 126,937.6 | 111,985.4 | 141,889.7 |
| Mediana | 60 000 | 60 000 |  |  |
| Comentario | La media estimada es 126,937.6 pesos, representando el ingreso promedio en la muestra. Este valor está influenciado por ingresos altos que elevan el promedio. | La mediana es 60,000 pesos, indicando que la mitad de los individuos gana menos y la otra mitad más, lo que sugiere una distribución sesgada hacia ingresos bajos. | El estimador de media por máxima verosimilitud coincide con el estimador por analogía (126,937.6 pesos), lo que valida la estabilidad de la estimación para el ingreso promedio. | La mediana estimada por máxima verosimilitud es 60,000 pesos, igual a la estimación por analogía, reforzando la representatividad del ingreso típico en la muestra. |

* ***Evaluación del estimador:***

*Insesgamiento:*

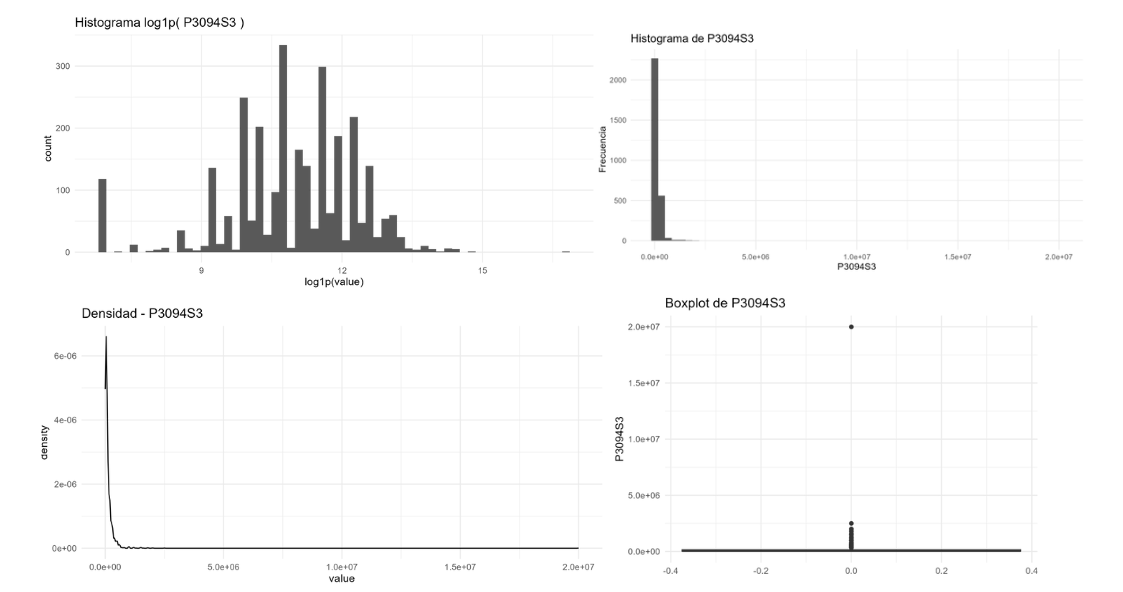
|  |  |  |
| --- | --- | --- |
| **Media** | **Mediana** | **Sesgo** |
| -166.3789 | 129 | Ambos estimadores, media y mediana, presentan un sesgo pequeño relativo, por lo que las estimaciones son confiables y adecuadas para describir la variable de ingresos. |

Comentario:

Los análisis de sesgo para los estimadores de media y mediana indican que ambos presentan un sesgo pequeño y relativo, lo que sugiere que las estimaciones obtenidas son confiables y representativas de la distribución real de los ingresos. Este bajo nivel de sesgo respalda el uso de estos estadísticos para describir adecuadamente la variable P3094S3 en el análisis de la encuesta.

*Consistencia:*

Ambos estimadores presentan **sesgo pequeño**, por lo que pueden considerarse consistentes. Sin embargo, la **media** muestra un intervalo de confianza más amplio y mayor variabilidad, reflejando sensibilidad a los valores extremos. En contraste, la **mediana** exhibe intervalos muy concentrados y gran estabilidad frente al remuestreo, lo que confirma su **mayor consistencia y robustez** en esta distribución de ingresos.



Comentario:

La variable **P3094S3**, correspondiente a los ingresos reportados en la GEIH, presenta una **distribución altamente asimétrica hacia la derecha**. El boxplot evidencia una fuerte concentración de valores bajos y la presencia de valores atípicos que alcanzan ingresos muy altos (superiores a los 20 millones de pesos). El gráfico de densidad confirma que la gran mayoría de observaciones se agrupan en niveles cercanos a cero, mientras que solo una fracción muy pequeña de personas registra ingresos elevados.

El histograma en escala original muestra claramente esta concentración de valores bajos, lo que hace que los ingresos altos se vean poco representados. Por esta razón, se aplicó una **transformación logarítmica (log1p)**, con la cual la distribución adquiere una forma más cercana a la normal y permite identificar mejor el rango donde se concentra la mayoría de la población, entre aproximadamente 8.000 y 160.000 pesos.

En conclusión, la variable refleja el patrón esperado en datos de ingresos: una mayoría con ingresos bajos y una minoría con ingresos muy altos que generan una distribución sesgada. Para su análisis resulta más adecuado emplear medidas robustas como la **mediana y los percentiles**, en lugar del promedio simple, que puede verse fuertemente afectado por los valores extremos.

*Eficiencia*

|  |  |
| --- | --- |
| **Medida** | **Valor** |
| Media | 58,275.694 |
| Mediana | 2,852.046 |

Comentario:

La varianza de la media resulta mucho mayor que la de la mediana, lo que indica que la **mediana es un estimador más eficiente** en esta muestra de la variable P3094S3. En otras palabras, la mediana presenta menor variabilidad frente al re-muestreo bootstrap y ofrece una descripción más estable del ingreso típico, mientras que la media se ve fuertemente afectada por los valores extremos (outliers). Esto confirma que, en distribuciones sesgadas como los ingresos, la mediana es preferible a la media como medida de tendencia central.

Sintaxis empleada con esta variable:

**Variable 2**

* ***Análisis Descriptivo***

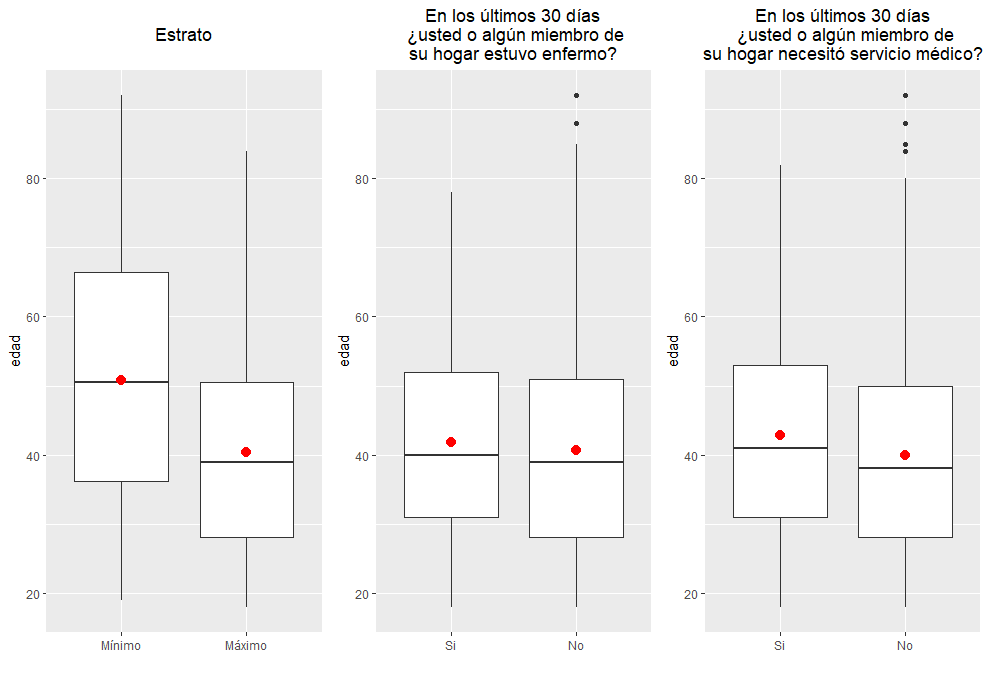
*Descriptivos Básicos*

|  |  |  |
| --- | --- | --- |
|  | 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 |

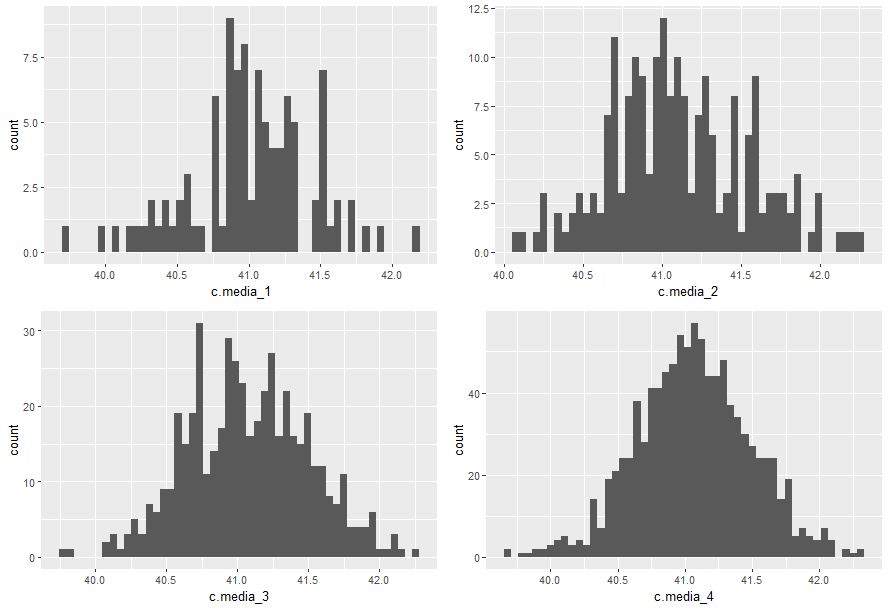
* ***Evaluación del estimador:***

*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***

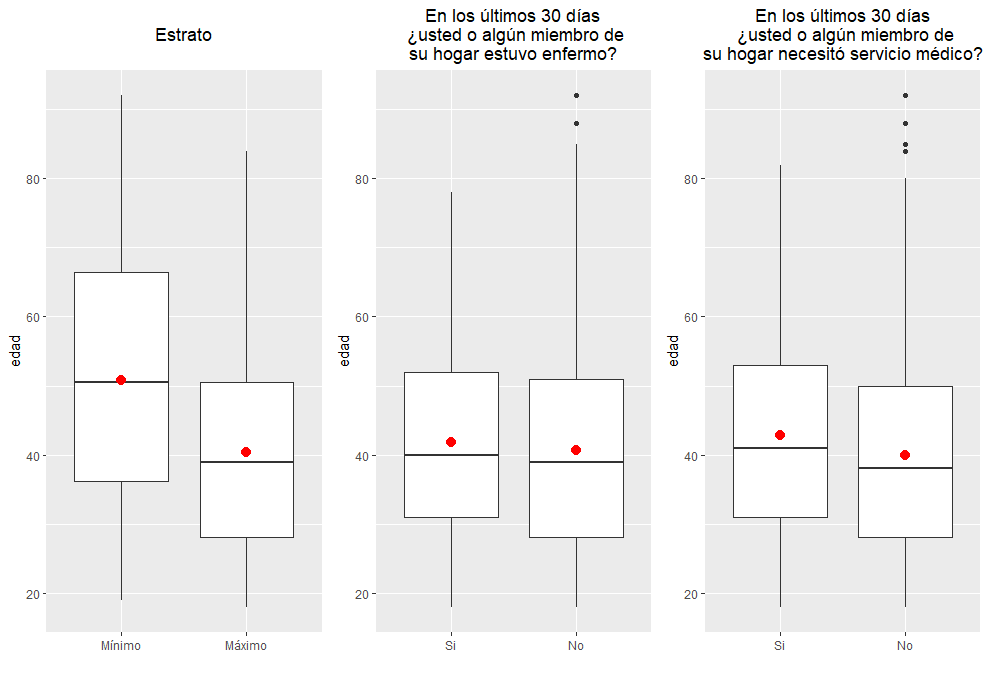
*Descriptivos Básicos*

|  |  |  |
| --- | --- | --- |
|  | Medida | Valor |
| 1 |  |  |
| 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 |

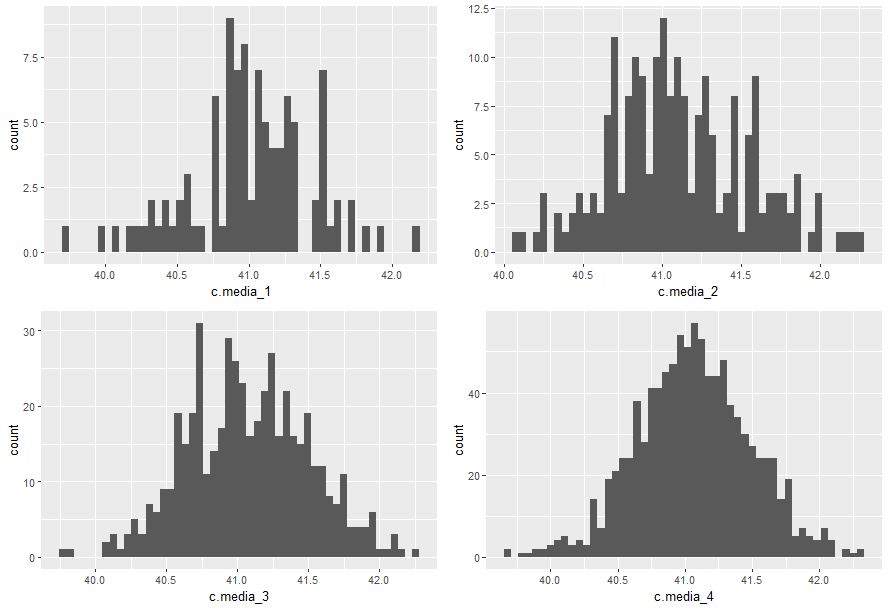
* ***Evaluación del estimador:***

*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 4**

* ***Análisis Descriptivo***

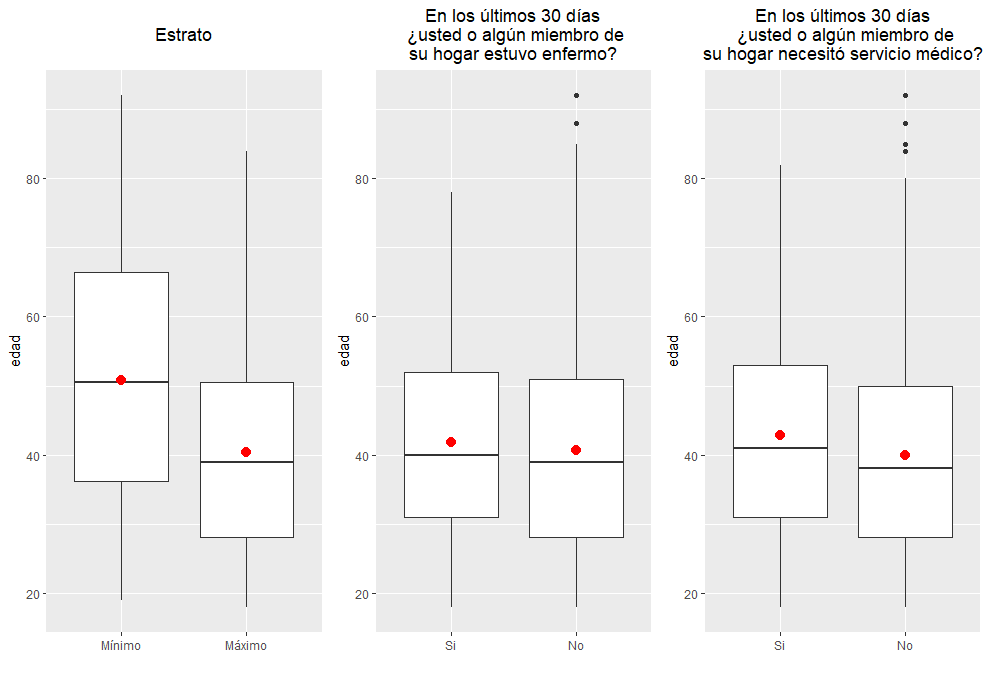
*Descriptivos Básicos*

|  |  |  |
| --- | --- | --- |
|  | 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 |

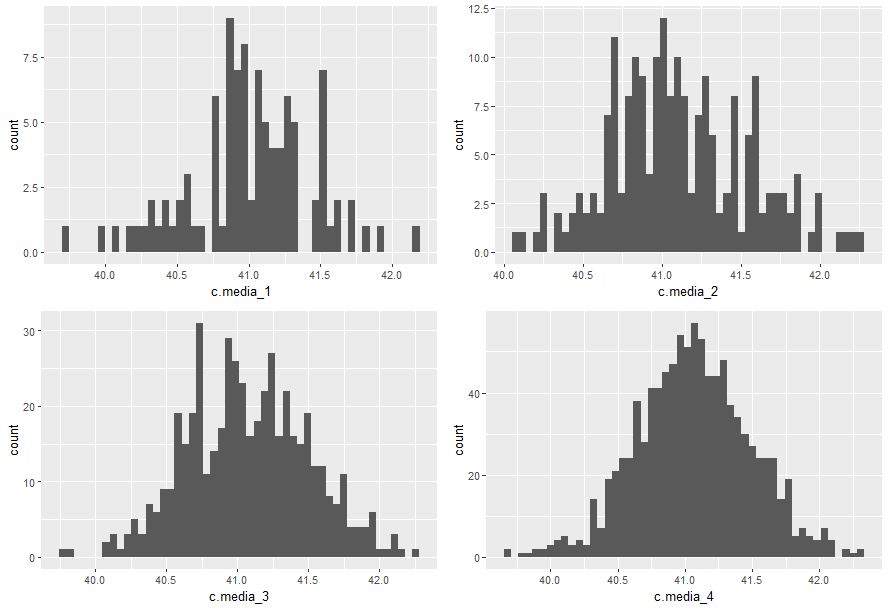
* ***Evaluación del estimador:***

*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