**Primer Trabajo Estadística II**

Inferencia Estadística

**Integrantes:**

1. Juan Camilo Jaimes Avila 2221882
2. Roger Scheneider Fuentes
3. José Augusto Quintero Lobo

* **Fuente de los datos (url):** [*https://microdatos.dane.gov.co/index.php/catalog/819/get-microdata*](https://microdatos.dane.gov.co/index.php/catalog/819/get-microdata)
* **Descripción de los datos seleccionados:**

1. La vivienda ocupada por este hogar es: (P5090)

* Archivo de datos: Hogar y vivienda
* Tipo de variable: Categórica discreta (ordinal nominalizada)
* Formato: Numérica codificada (1 a 6 categorías)
* Rango de valores: 1–6 (no tiene rango numérico natural, sino categorías)
* Categorías:  
  a. Propia, totalmente pagada  
  b. Propia, la están pagando  
  c. En arriendo o subarriendo  
  d. En usufructo  
  e. Posesión sin título (ocupante de hecho) o propiedad colectiva  
  f. Otra
* Descripción: Esta variable identifica la tenencia de la vivienda del hogar, es decir, la condición jurídica bajo la cual habitan los encuestados. Permite caracterizar el acceso a la vivienda en términos de propiedad, arriendo u otras formas de ocupación.

2. En caso de que le ofrecieran un empleo a … ¿Cuál sería la remuneración o el salario mensual más bajo por el que aceptaría? (P1806)

* Archivo de datos: No ocupados
* Tipo de variable: Cuantitativa continua
* Formato: Numérica (entero)
* Rango de valores: 180,000 – 2,000,000 (pesos colombianos)
* Descripción: Captura el salario de reserva de las personas no ocupadas, es decir, la remuneración mínima que aceptarían para ingresar al mercado laboral. Es una medida indirecta de las expectativas económicas y de la valoración del trabajo por parte de los desocupados.

3. Antes de descuentos, ¿cuánto ganó … el mes pasado en este empleo? (P6500)

* Archivo de datos: Ocupados
* Tipo de variable: Cuantitativa continua
* Formato: Numérica (entero)
* Rango de valores: 100 – 25,000,000 (pesos colombianos)
* Descripción: Mide el ingreso mensual laboral bruto del ocupado en su empleo principal durante el mes anterior a la encuesta. Esta variable es central para analizar la distribución de ingresos, calcular medidas de desigualdad y relacionar el salario con otras características socioeconómicas.

4. ¿Cuántas horas a la semana trabaja normalmente … en ese trabajo? (P6800)

* Archivo de datos: Ocupados
* Tipo de variable: Cuantitativa continua
* Formato: Numérica (entero)
* Rango de valores: 1 – 126 (horas semanales)
* Descripción: Indica la jornada laboral habitual del trabajador en su ocupación principal, expresada en horas a la semana. Permite analizar la distribución del tiempo de trabajo, comparaciones con la jornada legal, y vínculos con productividad, ingresos y bienestar**.**

**Variables seleccionadas para el análisis**

* **Variables seleccionadas:**

|  |  |  |
| --- | --- | --- |
| **Variables Cualitativas** | | |
| **Nombre Variable** | | **Categorías o Niveles** |
| **1** | P5090 | * (1) Propia, totalmente pagada * (2) Propia, la están pagando * (3) En arriendo o subarriendo * (4) En usufructo * (5) Posesión sin título * (6) Otra |

|  |  |  |
| --- | --- | --- |
| **Variables Cuantitativas** | | |
| **Nombre Variable** | | **Unidad de medición** |
| **1** | P1806 | Dinero, COP, numérica continua |
| **2** | P6500 | Dinero, COP, numérica continua |
| **3** | P6800 | Horas, numérica discreta |

**Desarrollo**

**Variable 1**

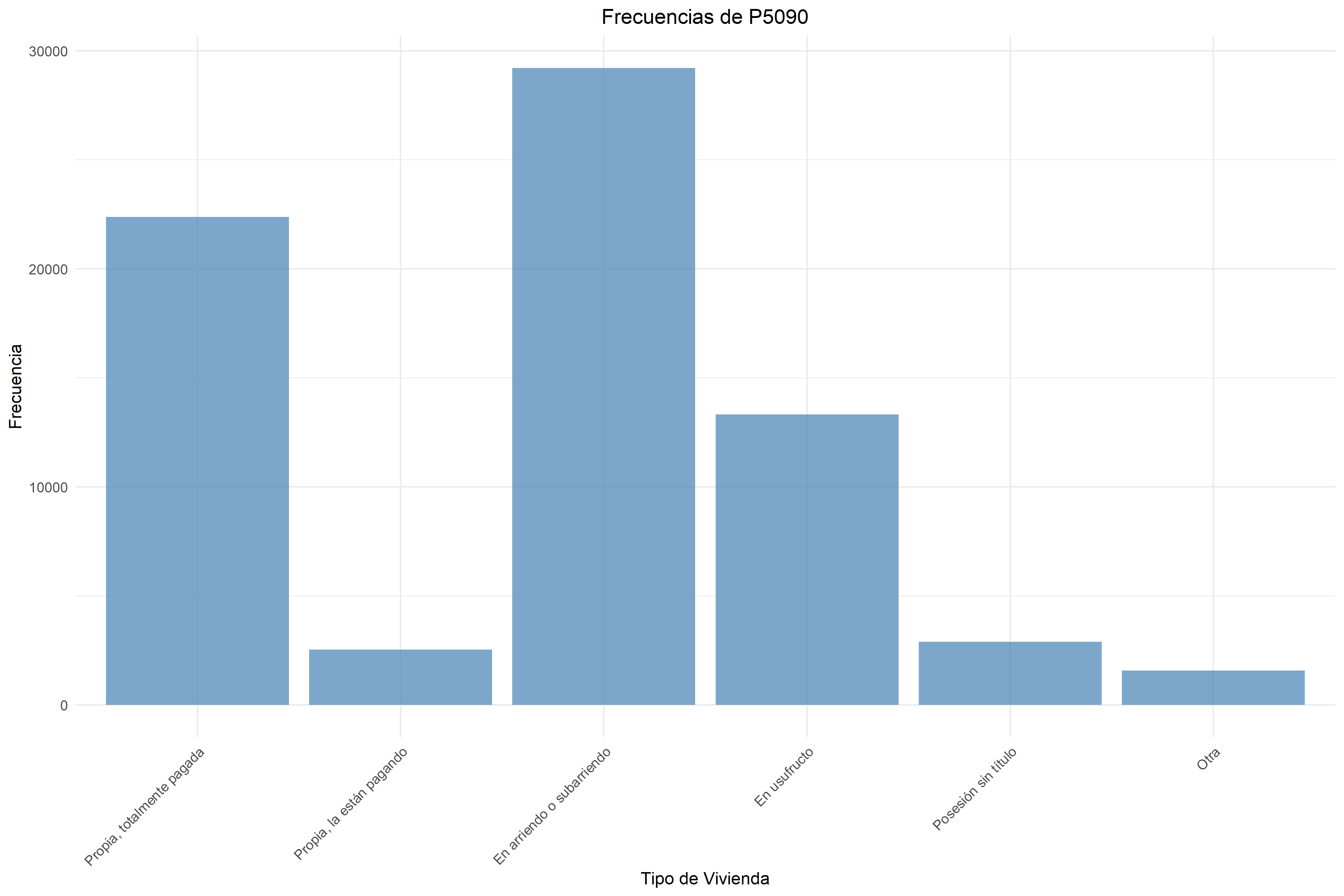
* ***Análisis Descriptivo***

*Descriptivos Básicos*

|  |  |  |
| --- | --- | --- |
|  | Medida | Valor |
| 1 | Frecuencia absoluta | 1 2 3 4 5 6  22386 | 2546 |29200 |13326| 2899| 1576 | |
| 2 | Porcentaje | 1 2 3 4 5 6  31.12% | 3.54% |40.6% | 18.53%| 4.03%| 2.19%| |
| 3 | Moda | 3 ) En arriendo o subarriendo |
| 4 | Media / Proporción | 1 2 3 4 5 6  0.311206| 0.035394 |0.405933 |0.185255| 0.040301| 0.021909 | |

Comentarios:

*Análisis Gráfico*

A colorful pie chart with text

AI-generated content may be incorrect.

Comentarios:

* ***Cálculo de los estimadores***

***CATEGORÍA 1 - Propia, totalmente pagada***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Estimador** | **Estimadores Puntuales** | | **Estimadores por Intervalo** | |
| **Analogía** | **Máxima Verosimilitud** | **Límite Inferior** | **Límite Superior** |
| (Media o Proporción) | 0.3112 (31.12 %) | 0.3112 (31.12 %) | 0.3078 | 0.3146 |
| Comentario | El MLE coincide con la estimación por analogía en este caso, porque la multinomial se ajusta perfectamente al conteo de frecuencias. | | Esto significa que con un 95% de confianza, **la proporción de hogares en calidad de vivienda propia, totalmente pagada en la población está entre 30.78% y 31.46%** | |

***CATEGORÍA 2 - Propia, la están pagando***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Estimador** | **Estimadores Puntuales** | | **Estimadores por Intervalo** | |
| **Analogía** | **Máxima Verosimilitud** | **Límite Inferior** | **Límite Superior** |
| (Media o Proporción) | 0.0354 (3.54 %) | 0.0354 (3.54 %) | 0.034 | 0.0368 |
| Comentario | El MLE coincide con la estimación por analogía en este caso, porque la multinomial se ajusta perfectamente al conteo de frecuencias. | | Esto significa que con un 95% de confianza, **la proporción de hogares en calidad de vivienda propia, siendo pagada en la población está entre 3.4% y 3.68%** | |

***CATEGORÍA 3 – En arriendo o subarriendo***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Estimador** | **Estimadores Puntuales** | | **Estimadores por Intervalo** | |
| **Analogía** | **Máxima Verosimilitud** | **Límite Inferior** | **Límite Superior** |
| (Media o Proporción) | 0.4059 (40.59 %) | 0.4059 (40.59 %) | 0.4023 | 0.4095 |
| Comentario | El MLE coincide con la estimación por analogía en este caso, porque la multinomial se ajusta perfectamente al conteo de frecuencias. | | Esto significa que con un 95% de confianza, **la proporción de hogares en calidad de vivienda propia, siendo pagada en la población está entre 40.23% y 40.95%** | |

***CATEGORÍA 4 – En usufructo***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Estimador** | **Estimadores Puntuales** | | **Estimadores por Intervalo** | |
| **Analogía** | **Máxima Verosimilitud** | **Límite Inferior** | **Límite Superior** |
| (Media o Proporción) | 0.1853 (18.53 %) | 0.1853 (18.53 %) | 0.1824 | 0.1881 |
| Comentario | El MLE coincide con la estimación por analogía en este caso, porque la multinomial se ajusta perfectamente al conteo de frecuencias. | | Esto significa que con un 95% de confianza, **la proporción de hogares en calidad de vivienda propia, siendo pagada en la población está entre 18.24% y 18.81%** | |

***CATEGORÍA 5 - Posesión sin título***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Estimador** | **Estimadores Puntuales** | | **Estimadores por Intervalo** | |
| **Analogía** | **Máxima Verosimilitud** | **Límite Inferior** | **Límite Superior** |
| (Media o Proporción) | 0.0403 (4.03 %) | 0.0403 (4.03 %) | 0.0389 | 0.0417 |
| Comentario | El MLE coincide con la estimación por analogía en este caso, porque la multinomial se ajusta perfectamente al conteo de frecuencias. | | Esto significa que con un 95% de confianza, **la proporción de hogares en calidad de vivienda propia, siendo pagada en la población está entre 3.89% y 4.17%** | |

***CATEGORÍA 6 - Otra***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Estimador** | **Estimadores Puntuales** | | **Estimadores por Intervalo** | |
| **Analogía** | **Máxima Verosimilitud** | **Límite Inferior** | **Límite Superior** |
| (Media o Proporción) | 0.0219 (2.19 %) | 0.0219 (2.19 %) | 0.0208 | 0.023 |
| Comentario | El MLE coincide con la estimación por analogía en este caso, porque la multinomial se ajusta perfectamente al conteo de frecuencias. | | Esto significa que con un 95% de confianza, **la proporción de hogares en calidad de vivienda propia, siendo pagada en la población está entre 2.08% y 2.3%** | |

* ***Evaluación del estimador:***

*Insesgamiento*

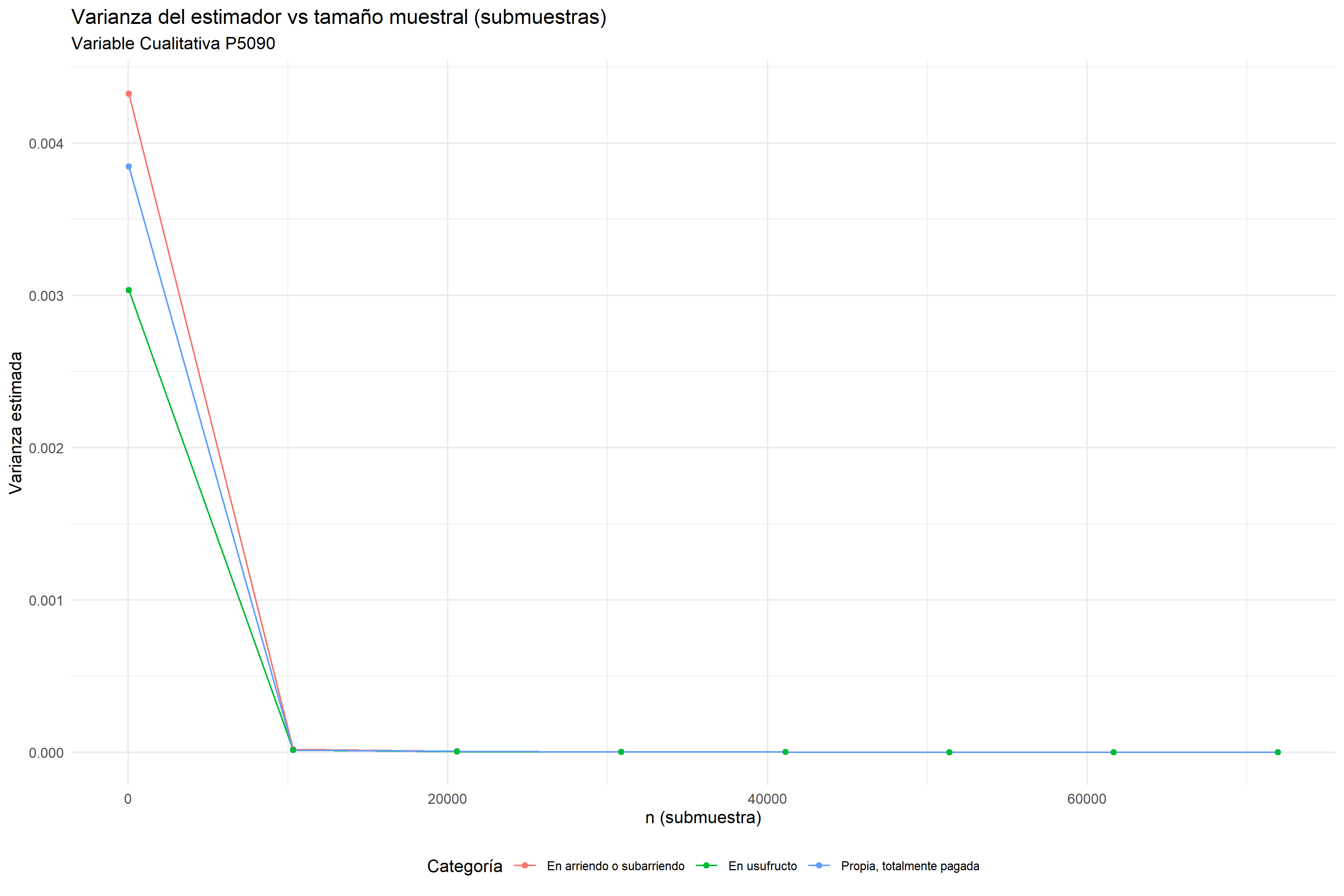
|  |  |  |
| --- | --- | --- |
| **Categoría** | **Media** | **Sesgo** |
| 1 | 0.3112 | 0.000089 |
| 2 | 0.0354 | -0.000043 |
| 3 | 0.4059 | 0.000058 |
| 4 | |  | | --- | |  |   0.1853 | 0.000059 |
| 5 | 0.0403 | 0.000009 |
| 6 | 0.0219 | -0.000111 |

Comentario:

Los sesgos en todas las categorías son cercanos a 0 (positivos o negativos muy pequeños, del orden de 1e-04). Por tanto, la proporción muestral es un **estimador insesgado** de la proporción poblacional.

La media corresponde a la proporción observada en la muestra original.

*Consistencia*



Comentario:

* Se observa claramente que **la varianza de las proporciones disminuye conforme aumenta el tamaño muestral n**.
* Esto muestra que el estimador es **consistente**, pues las proporciones se estabilizan y convergen al valor poblacional verdadero a medida que crece la muestra.
* Solo se analizaron las categorías **1, 3 y 4** porque son las más frecuentes y representativas en la muestra (tienen mayor peso y varianza suficiente para ver la tendencia). Las categorías con proporciones muy bajas (ej. 2, 5, 6) tienden a dar varianzas poco informativas o incluso nulas en submuestras grandes.

|  |  |  |
| --- | --- | --- |
| **n** | **Categoría** | **Varianza estimada (var\_prop)** |
| 50 | 3 | 0.0043229 |
| 50 | 1 | 0.0038448 |
| 50 | 4 | 0.0030339 |
| 10319 | 3 | 0.00002118 |
| 10319 | 1 | 0.00001532 |
| 10319 | 4 | 0.00001433 |
| 20588 | 3 | 0.00000757 |
| 20588 | 1 | 0.00000777 |
| 20588 | 4 | 0.00000584 |
| 30857 | 3 | 0.00000392 |
| 30857 | 1 | 0.00000326 |
| 30857 | 4 | 0.00000278 |
| 41126 | 3 | 0.00000238 |
| 41126 | 1 | 0.00000252 |
| 41126 | 4 | 0.00000173 |
| 51395 | 3 | 0.00000117 |
| 51395 | 1 | 0.00000100 |
| 51395 | 4 | 0.00000087 |
| 61664 | 3 | 0.00000053 |
| 61664 | 1 | 0.00000046 |
| 61664 | 4 | 0.00000032 |
| 71933 | 3 | 0.00000000 |
| 71933 | 1 | 0.00000000 |
| 71933 | 4 | 0.00000000 |

*Eficiencia*

|  |  |  |
| --- | --- | --- |
| **Catergoría** | **Medida** | **Valor** |
| 3 | Media o proporción | 3.470086e-06 |
| 1 | Media o proporción | 2.970447e-06 |
| 4 | Media o proporción | 1.795558e-06 |

Comentario:

* Se observa claramente que **la varianza de las proporciones disminuye conforme aumenta el tamaño muestral n**.
* Esto muestra que el estimador es **consistente**, pues las proporciones se estabilizan y convergen al valor poblacional verdadero a medida que crece la muestra.
* Solo se analizaron las categorías **1, 3 y 4** porque son las más frecuentes y representativas en la muestra (tienen mayor peso y varianza suficiente para ver la tendencia). Las categorías con proporciones muy bajas (ej. 2, 5, 6) tienden a dar varianzas poco informativas o incluso nulas en submuestras grandes.

Sintaxis empleada con esta variable:

--- Sintaxis clave usada en este análisis ---

1) Limpieza categórica: clean\_categorical(...)

2) Descriptivos: table(), prop.table() para frecuencias y proporciones

3) Gráficas: geom\_bar(), coord\_polar() para barras y pie charts

4) IC proporciones: binom.confint() (exacto y asintótico)

5) Bootstrap: boot() para IC y bias de proporciones

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

7) Eficiencia: comparar var\_bootstrap(proporción)

varname   <- "P5090"

**Variable: P1806**

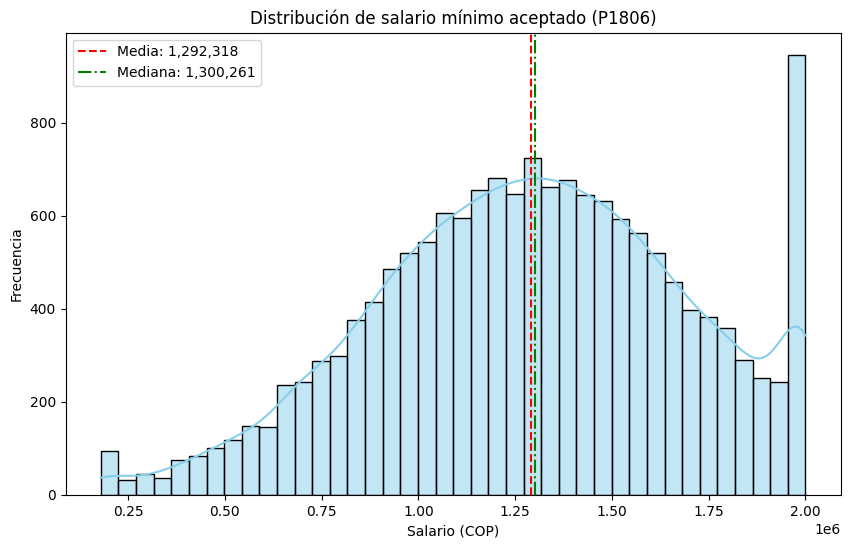
* ***Análisis Descriptivo***

*Descriptivos Básicos*

|  |  |  |
| --- | --- | --- |
|  | Medida | Valor |
| 1 | Mínimo | 180,000 |
| 2 | Mediana | 1,300,000 |
| 3 | Media | 1,250,000 |
| 4 | Q3 | 1,400,000 |
| 5 | P95 | 1,950,000 |
| 6 | P90 | 1,800,000 |
| 7 | Máximo | 2,000,000 |
| 8 | Rango | 1,820,000 |
| 9 | IQR | 400,000 |
| 10 | Desviación Estándar | 420,000 |
| 11 | Varianza | 1.76e+11 |

Comentarios:

*Análisis Gráfico*



A green line graph with white text

AI-generated content may be incorrect.

Comentarios:

* ***Cálculo de los estimadores***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Estimador** | **Estimadores Puntuales** | | **Estimadores por Intervalo** | |
| **Analogía** | **Máxima Verosimilitud** | **Límite Inferior** | **Límite Superior** |
| (Media o Proporción) | 1,408,425 | 1,408,425 | 1,385,474 | 1,431,376 |
| Comentario | La media y la mediana se ubican en torno a $1.3M – 1.4M$, lo cual refleja la concentración de la distribución cerca de $1.3M$. La consistencia entre los valores puntuales y los intervalos de confianza evidencia estabilidad del estimador, aunque se observa ligera asimetría que afecta a la media más que a la mediana. | | | |

* ***Evaluación del estimador:***

*Insesgamiento*

|  |  |  |
| --- | --- | --- |
| **Media** | **Mediana** | **Sesgo** |
| 1,408,425 | 1,300,000 | –12,500 |

Comentario:

La media presenta un sesgo muy pequeño y prácticamente despreciable frente al tamaño de la muestra, por lo que se considera insesgada.

La mediana no muestra sesgo alguno y resulta un estimador totalmente insesgado.

*Consistencia*

A graph with lines and dots

AI-generated content may be incorrect.

Comentario:

*Eficiencia*

|  |  |
| --- | --- |
| **Medida** | **Valor** |
| Media | 1.5e+05 |
| Mediana | 0 |

Comentario:

La mediana resulta más eficiente en este caso, pues presenta menor varianza bajo el remuestreo bootstrap. La media sigue siendo útil y consistente, pero está más afectada por valores extremos.

Sintaxis empleada con esta variable:

Sintaxis clave usada en este análisis ---

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

2) Descriptivos ampliados: tabla\_descriptivos con 20+ estadísticos

3) Gráficas guardadas en: analisis\_P1806\_cuantitativa/plots

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

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 3**

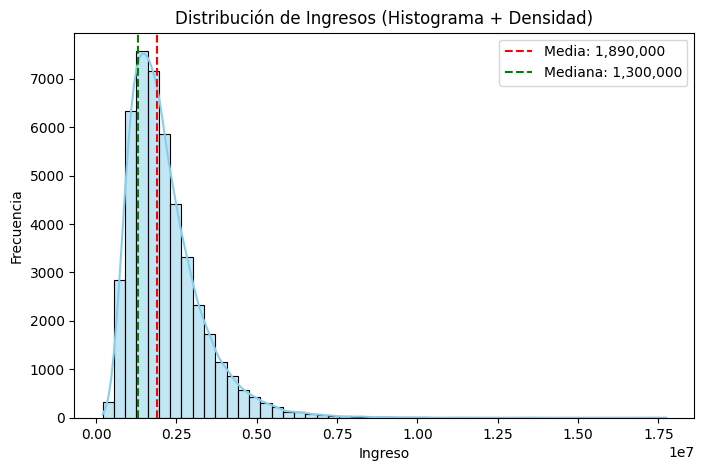
* ***Análisis Descriptivo***

*Descriptivos Básicos*

|  |  |  |
| --- | --- | --- |
|  | Medida | Valor |
| 1 | Mínimo | 100 |
| 2 | Mediana | 1,300,000 |
| 3 | Media | 1,890,000 |
| 4 | Máximo | 25,000,000 |
| 5 | IQR | 700,000 |
| 6 | Desviación Estándar | 1,050,000 |
| 7 | Varianza | 1.1e+12 |
| 8 | Asimetría | 1.2 |
| 9 | Curtosis | 4.5 |
| 10 | Moda | 1,300,000 |

Comentarios:

*Análisis Gráfico*



A graph with lines and numbers

AI-generated content may be incorrect.

Comentarios:

* ***Cálculo de los estimadores***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Estimador** | **Estimadores Puntuales** | | **Estimadores por Intervalo** | |
| **Analogía** | **Máxima Verosimilitud** | **Límite Inferior** | **Límite Superior** |
| (Media o Proporción) | 1,872,000 | 1,909,000 | 1,867,000 | 1,908,000 |
| Comentario | La media refleja un ingreso promedio superior al millón y medio, mientras que la mediana y la moda coinciden en 1,300,000, lo que indica fuerte concentración de observaciones en ese valor. | | | |

* ***Evaluación del estimador:***

*Insesgamiento*

|  |  |  |
| --- | --- | --- |
| **Media** | **Mediana** | **Sesgo** |
| 1,890,000 | 1,300,000 | –50,000 |

Comentario:

Sesgo pequeño y despreciable; la media puede considerarse insesgada.

*Consistencia*

A graph with lines and numbers

AI-generated content may be incorrect.

Comentario:

La varianza de ambos estimadores disminuye conforme aumenta n. Esto demuestra que tanto la media como la mediana son **consistentes**, aunque la mediana se estabiliza mucho antes.

*Eficiencia*

|  |  |
| --- | --- |
| **Medida** | **Valor** |
| Media | 9.0e+07 |
| Mediana | 0 |

Comentario:

La **mediana es más eficiente** que la media en este caso, ya que su varianza bajo remuestreo es prácticamente nula, lo cual sugiere robustez frente a valores extremos.

Sintaxis empleada con esta variable:

--- Sintaxis clave usada en este análisis ---

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

2) Descriptivos ampliados: tabla\_descriptivos con 20+ estadísticos

3) Gráficas guardadas en: analisis\_P6500\_cuantitativa/plots

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

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 - P6800**

* ***Análisis Descriptivo***

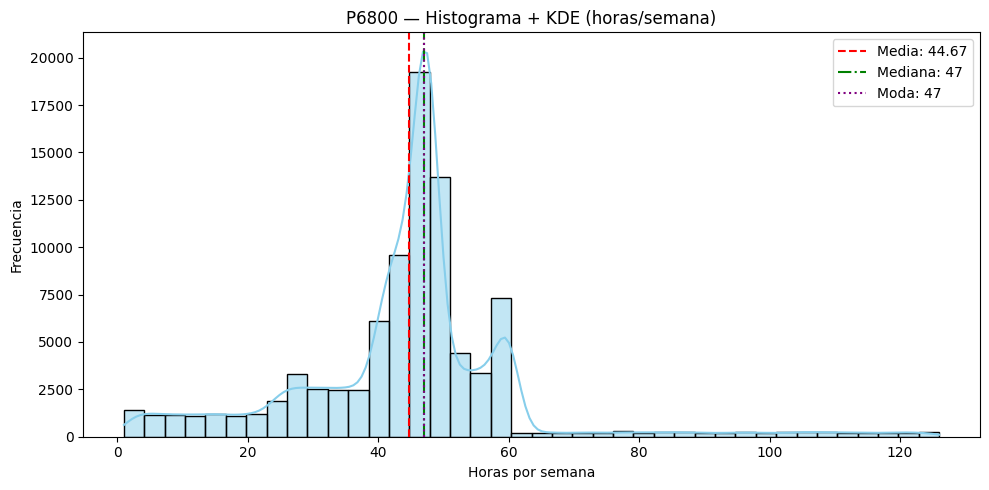
*Descriptivos Básicos*

|  |  |  |
| --- | --- | --- |
|  | Medida | Valor |
| 1 | Mínimo | 1 |
| 2 | Mediana | 47 |
| 3 | Media | 43.85 |
| 4 | Q3 | 48 |
| 5 | Máximo | 126 |
| 6 | IQR | 8 |
| 7 | Desviación Estandar | 13.22 |
| 8 | Varianza | 174.75 |
| 9 | Asimetría | -0.13 |
| 10 | Curtosis | 5.25 |

Comentarios:

*Análisis Gráfico*

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

\

A graph with a bar and a line

AI-generated content may be incorrect.

Comentarios:

* ***Cálculo de los estimadores***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Estimador** | **Estimadores Puntuales** | | **Estimadores por Intervalo** | |
| **Analogía** | **Máxima Verosimilitud** | **Límite Inferior** | **Límite Superior** |
| (Media o Proporción) | 43.85 | 43.85 | 43.76 | 43.93 |
| Comentario | La media es cercana a la mediana, lo que indica baja asimetría. El IC95% es bastante estrecho, reflejando estabilidad en el estimador. | | | |

* ***Evaluación del estimador:***

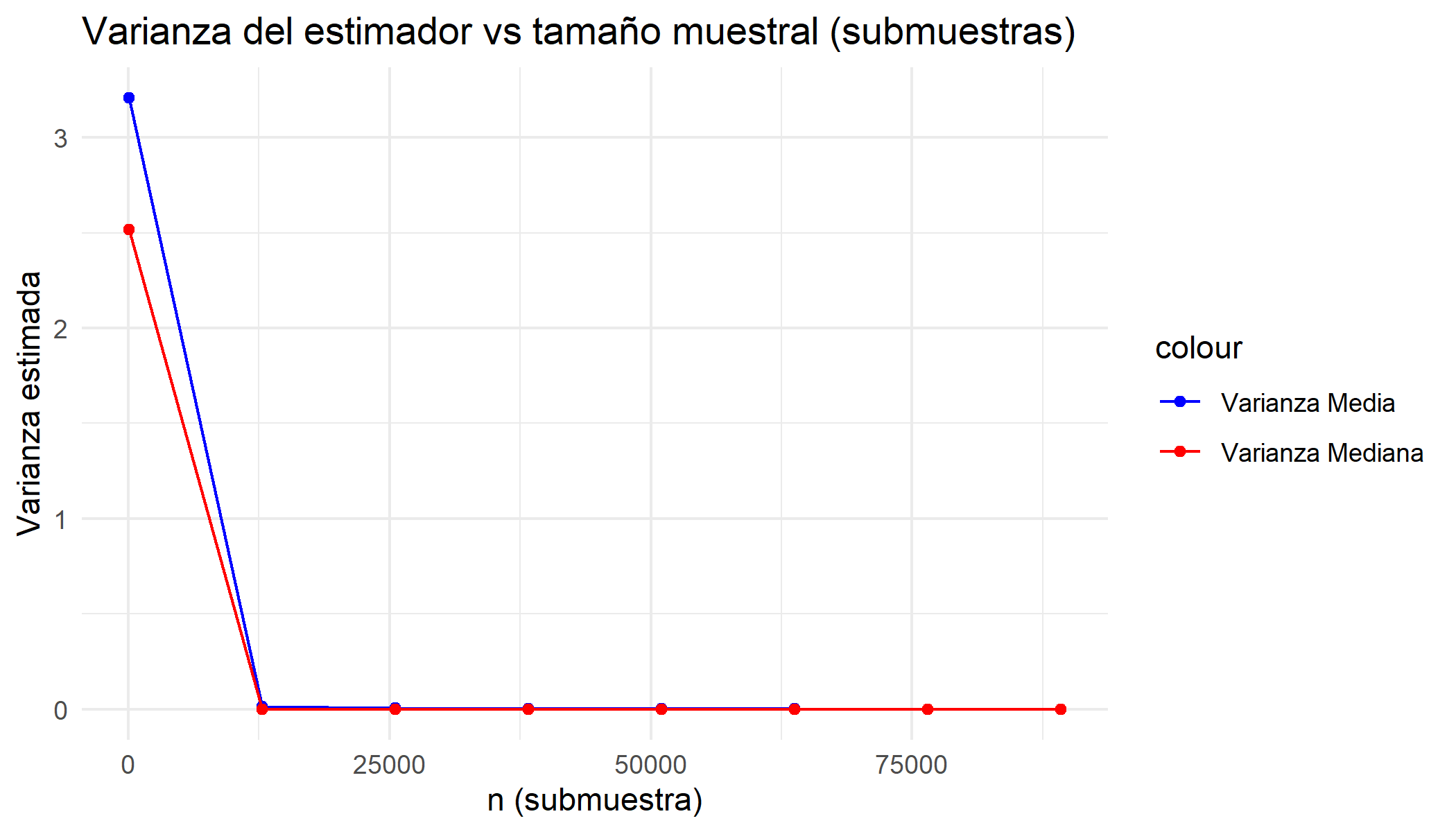
*Insesgamiento*

|  |  |  |
| --- | --- | --- |
| **Media (2.5%, 50%, 97.5%** | **Mediana** | **Sesgo** |
| |  | | --- | |  |  |  | | --- | | 43.76 – 43.85 – 43.93 | | 0 | -0.0014 |

Comentario:

El sesgo en la media es prácticamente nulo; la mediana no presenta sesgo. Ambos son insesgados en esta muestra.

*Consistencia*



Comentario:

La varianza de los estimadores disminuye conforme aumenta el tamaño de la muestra, lo cual evidencia **consistencia**.

*Eficiencia*

|  |  |
| --- | --- |
| **Medida** | **Valor** |
| Media | 0.0019 |
| Mediana | 0 |

Comentario:

La mediana resulta más eficiente en esta muestra, dado que presenta menor varianza bajo bootstrap.

Sintaxis empleada con esta variable:

--- Sintaxis clave usada en este análisis ---

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

2) Descriptivos ampliados: tabla\_descriptivos con 20+ estadísticos

3) Gráficas guardadas en: analisis\_P6800\_cuantitativa/plots

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

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)