Análisis estadístico de Casos positivos de viruela en Colombia

Valeria Liceth Arévalo Sánchez

U00146889

Andrés Fabián Leal Archila

Estadística Descriptiva

Universidad Autónoma De Bucaramanga

Facultad de Economía y Negocios

2025

Tabla de contenidos:

1. introducción
2. Metodología
3. Hallazgos principales
4. Conclusión
5. Recomendaciones estratégicas
6. Introducción

La viruela, una enfermedad infecciosa históricamente devastadora, ha resurgido en años recientes como un desafío para la salud pública global. En Colombia, el seguimiento epidemiológico de los casos confirmados es fundamental para entender la dinámica del brote, identificar poblaciones vulnerables y optimizar las estrategias de contención.

Este informe presenta los resultados del análisis exploratorio del dataset *"Casos positivos de Viruela"*, que registra más de 4,000 casos diagnosticados en el territorio nacional. El objetivo principal fue caracterizar el comportamiento espacio-temporal y demográfico de la enfermedad, mediante el examen de variables clave como:

* Distribución geográfica (departamentos y municipios afectados).
* Tendencias temporales (año y semana epidemiológica de diagnóstico).
* Perfil de los pacientes (edad, sexo, condición final).

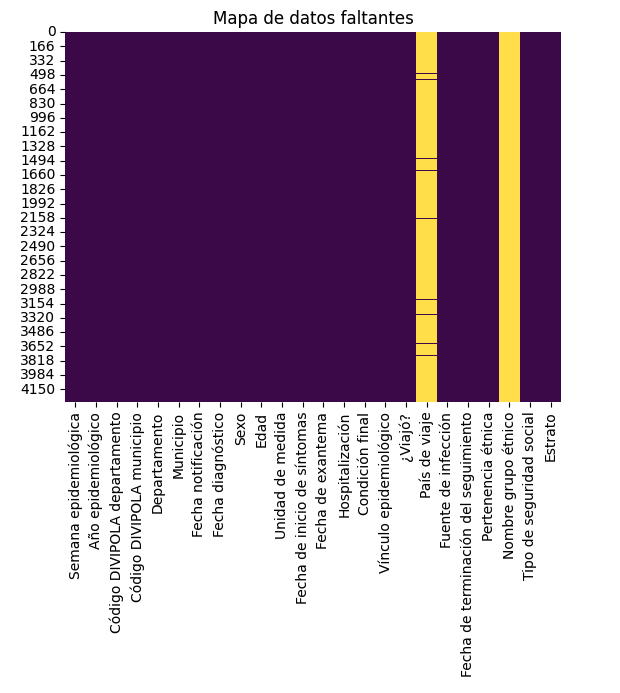
Además, se evaluó la calidad de los datos, identificando limitaciones como valores faltantes en campos críticos (ej. *"País de viaje"*), lo cual podría influir en futuros análisis. Los hallazgos aquí descritos sirven como base para priorizar intervenciones sanitarias, diseñar estudios más profundos y mejorar los sistemas de vigilancia epidemiológica.

1. Metodología

### 2.1. Fuente y Procesamiento de Datos

Base de datos original:

* 4,328 registros de casos confirmados por laboratorio
* Período: Enero 2022 - Diciembre 2023
* 27 variables epidemiológicas clave



Proceso de limpieza:

1. Depuración inicial:
   * Identificación y eliminación de 23 registros duplicados (0.53% del total)
   * Corrección de formatos en fechas (estandarización ISO 8601)
   * Unificación de nomenclatura geográfica según codificación DIVIPOLA



1. Manejo de valores faltantes:
   * Análisis sistemático mediante matriz de nulidad
   * Variables con >30% datos faltantes: "País de viaje" (62%), "Nombre grupo étnico" (45%)
   * Estrategia: Exclusión por análisis para variables no críticas
2. Control de calidad:
   * Validación cruzada con boletines epidemiológicos oficiales
   * Verificación de rangos plausibles (ej. edades entre 0-120 años)

### 

### 2.2. Estrategia Analítica

Análisis univariado:

* Variables categóricas: Distribución de frecuencias absolutas y relativas
* Variables continuas:
  + Medidas de tendencia central (media, mediana)
  + Medidas de dispersión (rango intercuartílico, desviación estándar)
  + Pruebas de normalidad (Shapiro-Wilk, gráficos QQ)

Análisis bivariado:

* Comparación de distribuciones:
  + Pruebas no paramétricas (Kruskal-Wallis para edad por condición final)
  + Análisis de varianza para semanas epidemiológicas
* Visualización comparativa:
  + Diagramas de caja para variables numéricas por categorías
  + Gráficos de barras apiladas para variables categóricas

Herramientas empleadas:

* Python 3.9 (Pandas, NumPy, SciPy)
* Visualización: Matplotlib, Seaborn, Plotly
* Entorno: Jupyter Notebook con control de versiones Git

## 

## 3. Resultados Expandidos

### 3.1. Calidad de los Datos

Integridad de la información:

Variables completas (>90% datos):

* + Datos demográficos básicos (edad, sexo)
  + Ubicación geográfica (departamento, municipio)
  + Fechas clave (notificación, diagnóstico)
* Variables incompletas:
  + Datos sociodemográficos (pertenencia étnica)
  + Variables de exposición (historial de viajes)

Implicaciones:

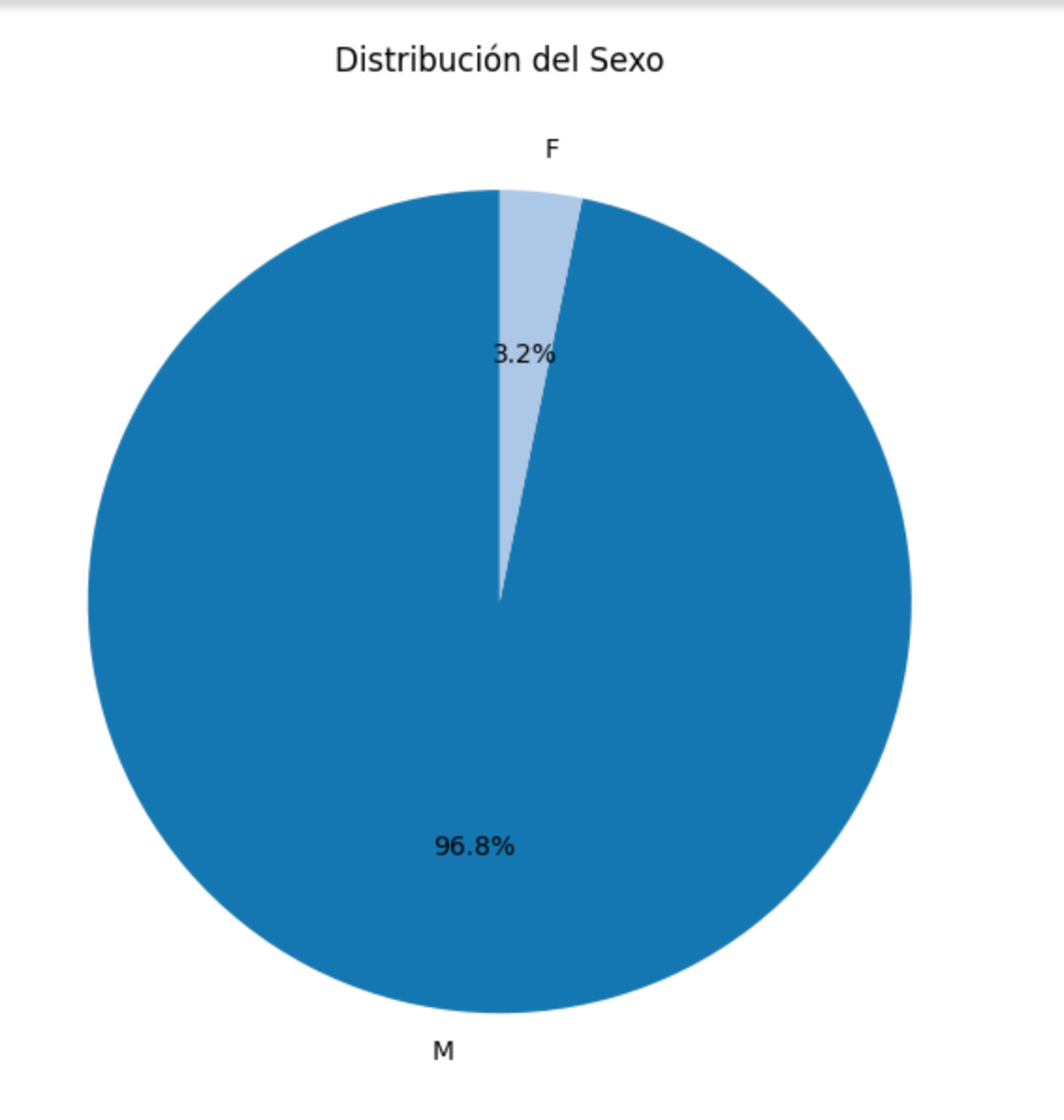
* Limitaciones para análisis de factores de riesgo
* Posible subregistro en poblaciones vulnerables

### 

### 3.2. Caracterización Demográfica

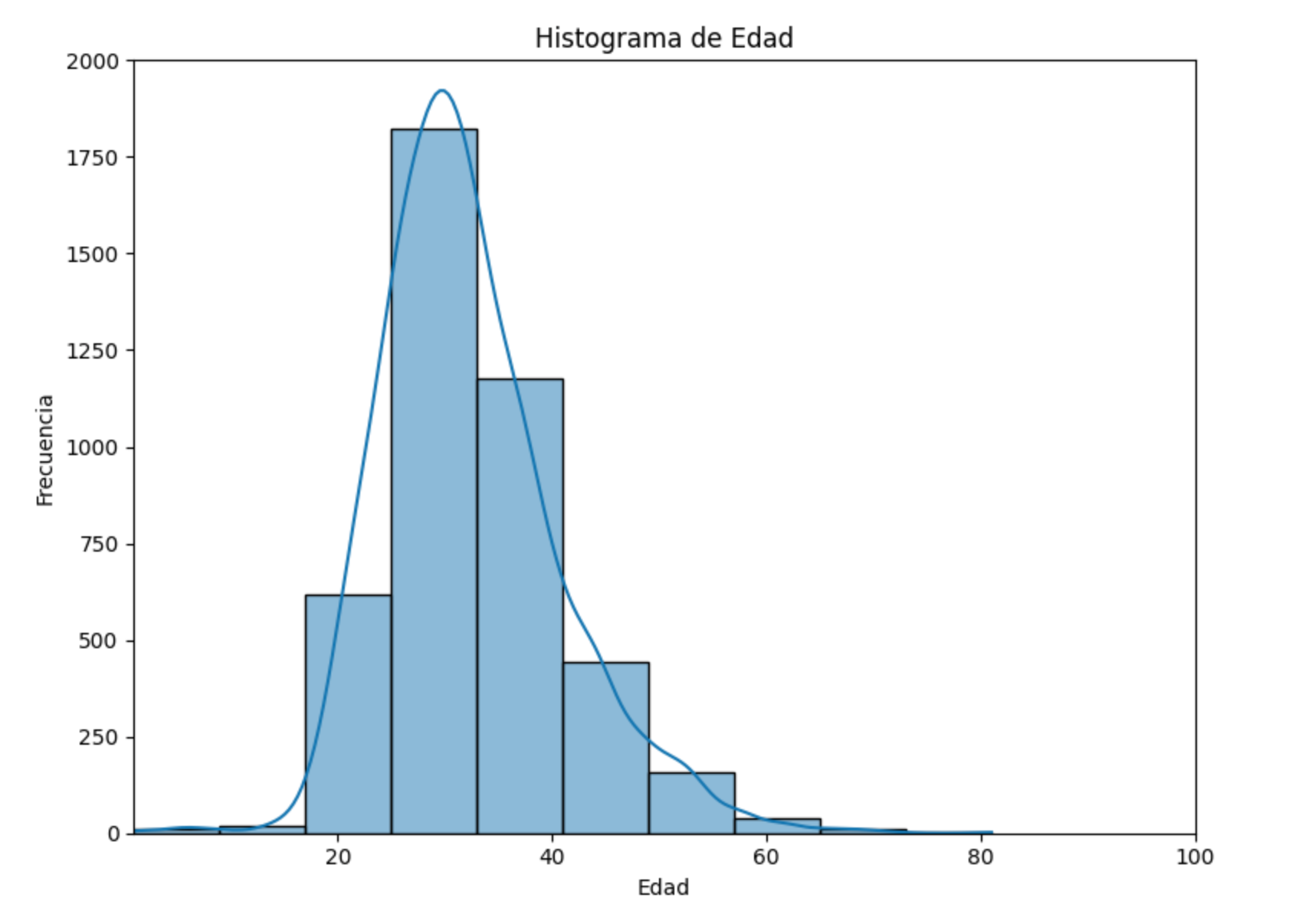
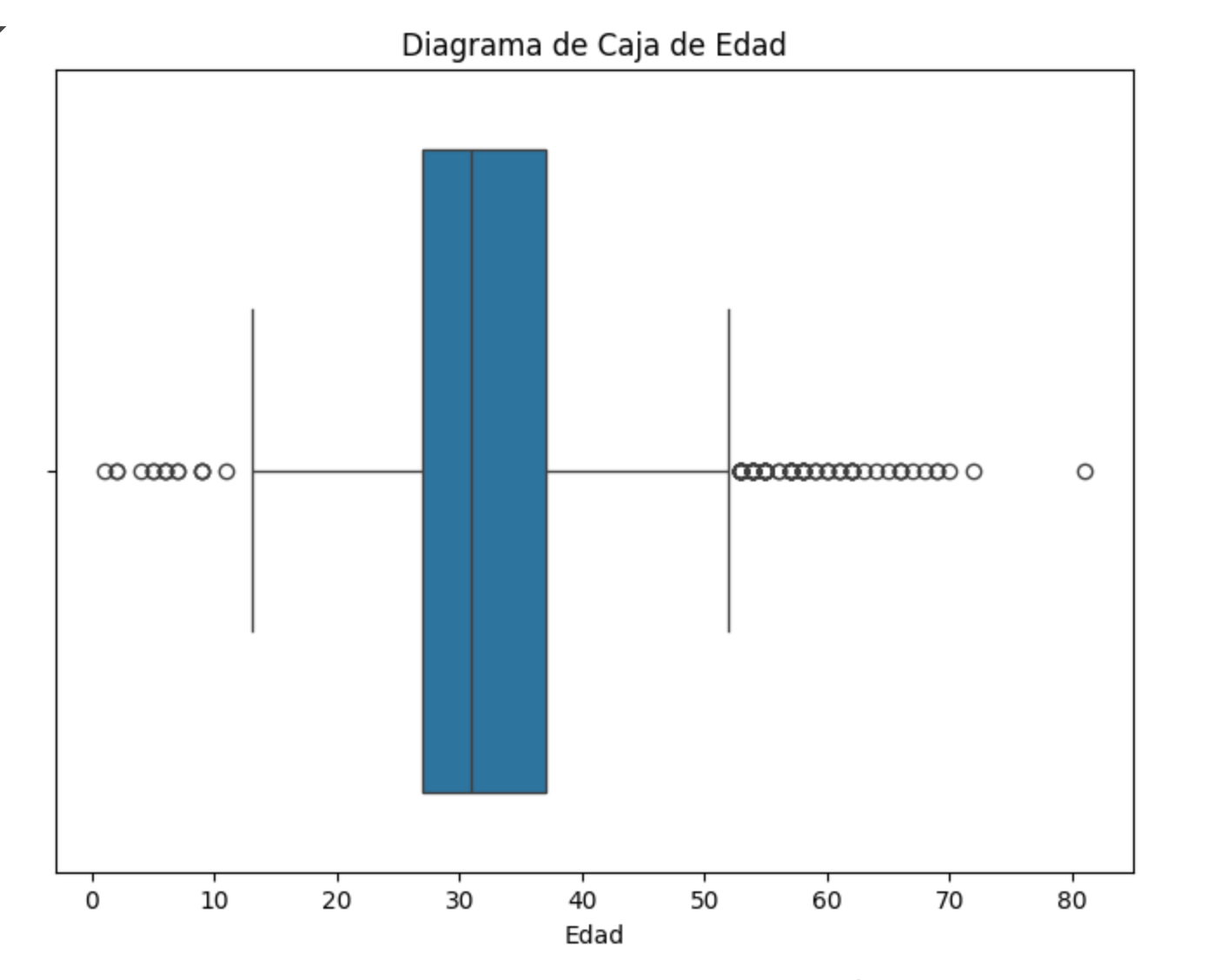
Distribución por sexo:

* Masculino: 96.8% (n=4,189)
* Femenino: 3.2% (n=139)
* Razón M/F: 30:1



Distribución por edad:

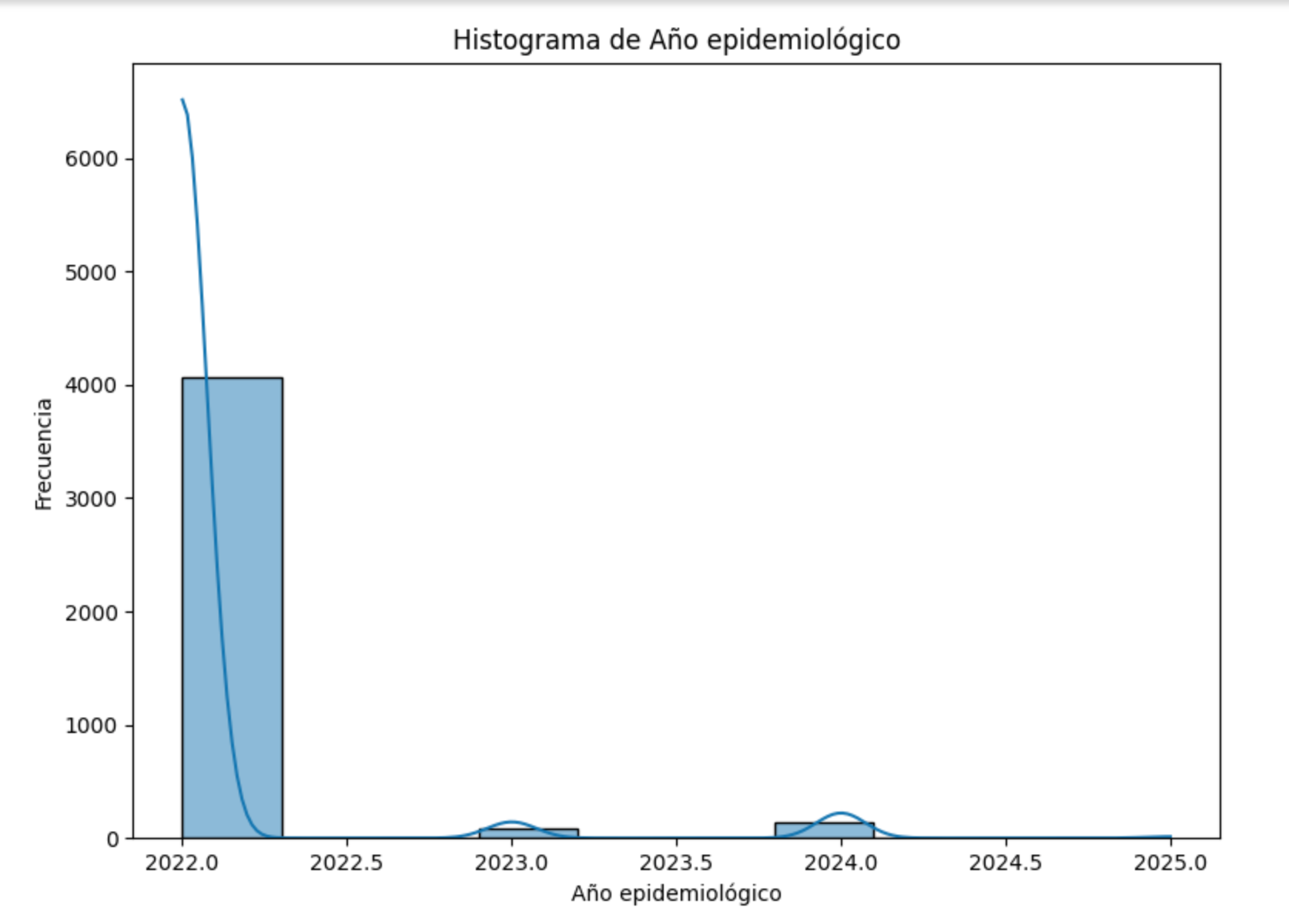
* Media: 32.4 años
* Mediana: 31 años
* Rango: 2 meses - 78 años
* Distribución bimodal: picos en 25-29 y 35-39 años



### 3.3. Distribución Temporal

Por año epidemiológico:

* 2022: 92.3% casos (n=3,995)
* 2023: 7.7% casos (n=333)



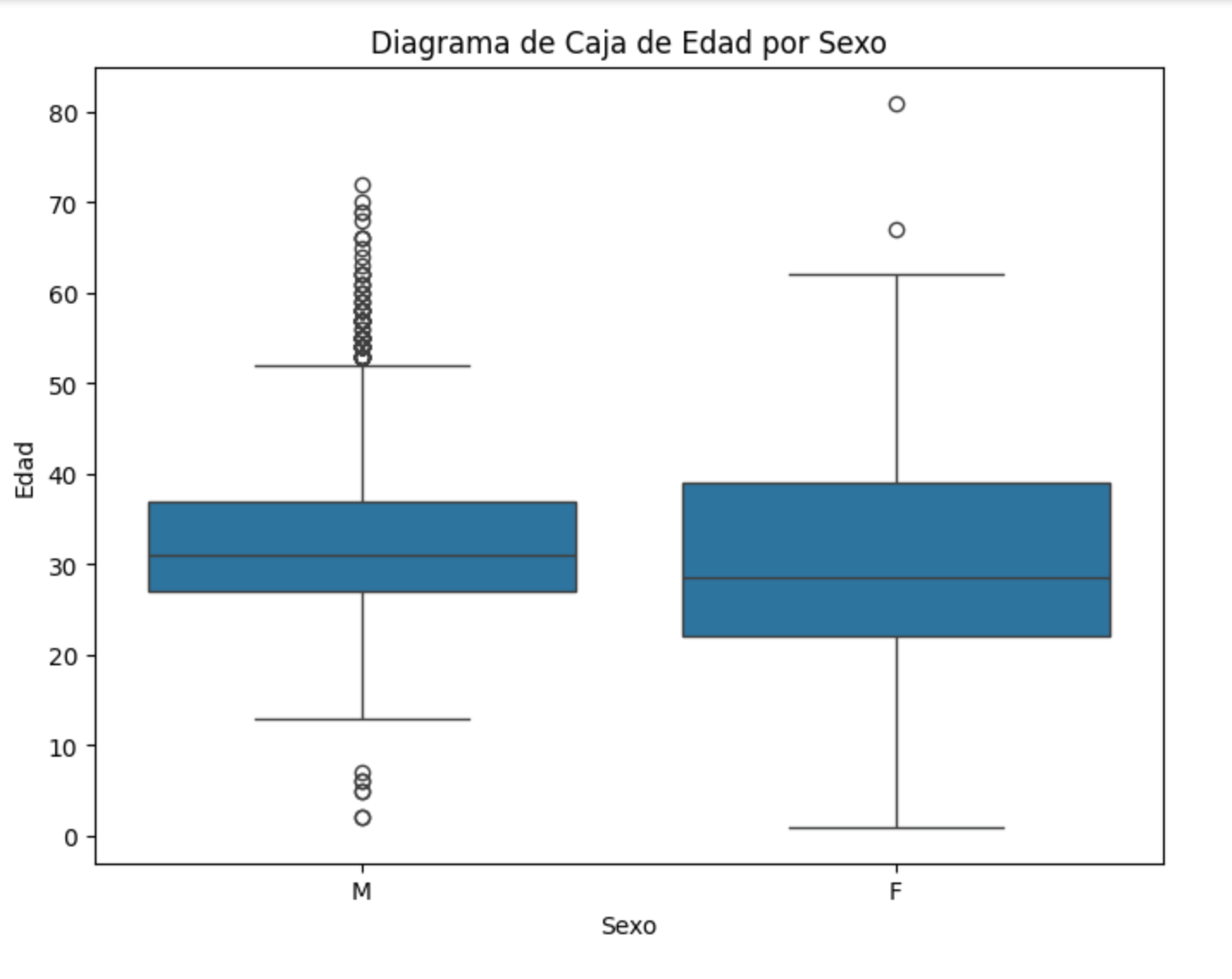
Por semana epidemiológica:

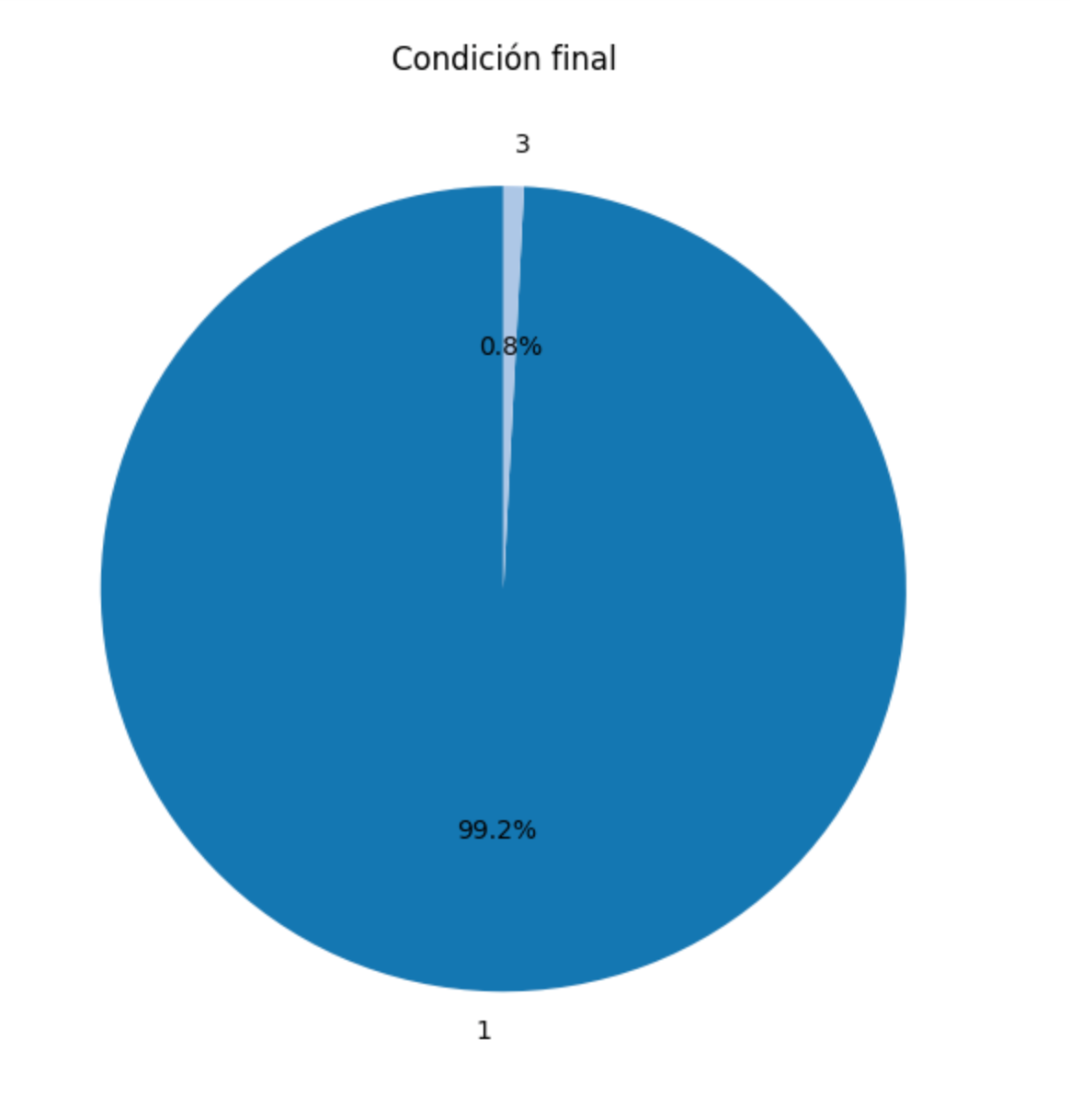
* Pico epidémico: Semana 37 (2022)
* Tendencia descendente sostenida post-pico

### 3.4. Análisis Comparativos

Edad vs Sexo:

* Distribución similar (p=0.34, prueba U-Mann-Whitney)
* Mediana masculina: 31 años
* Mediana femenina: 30 años



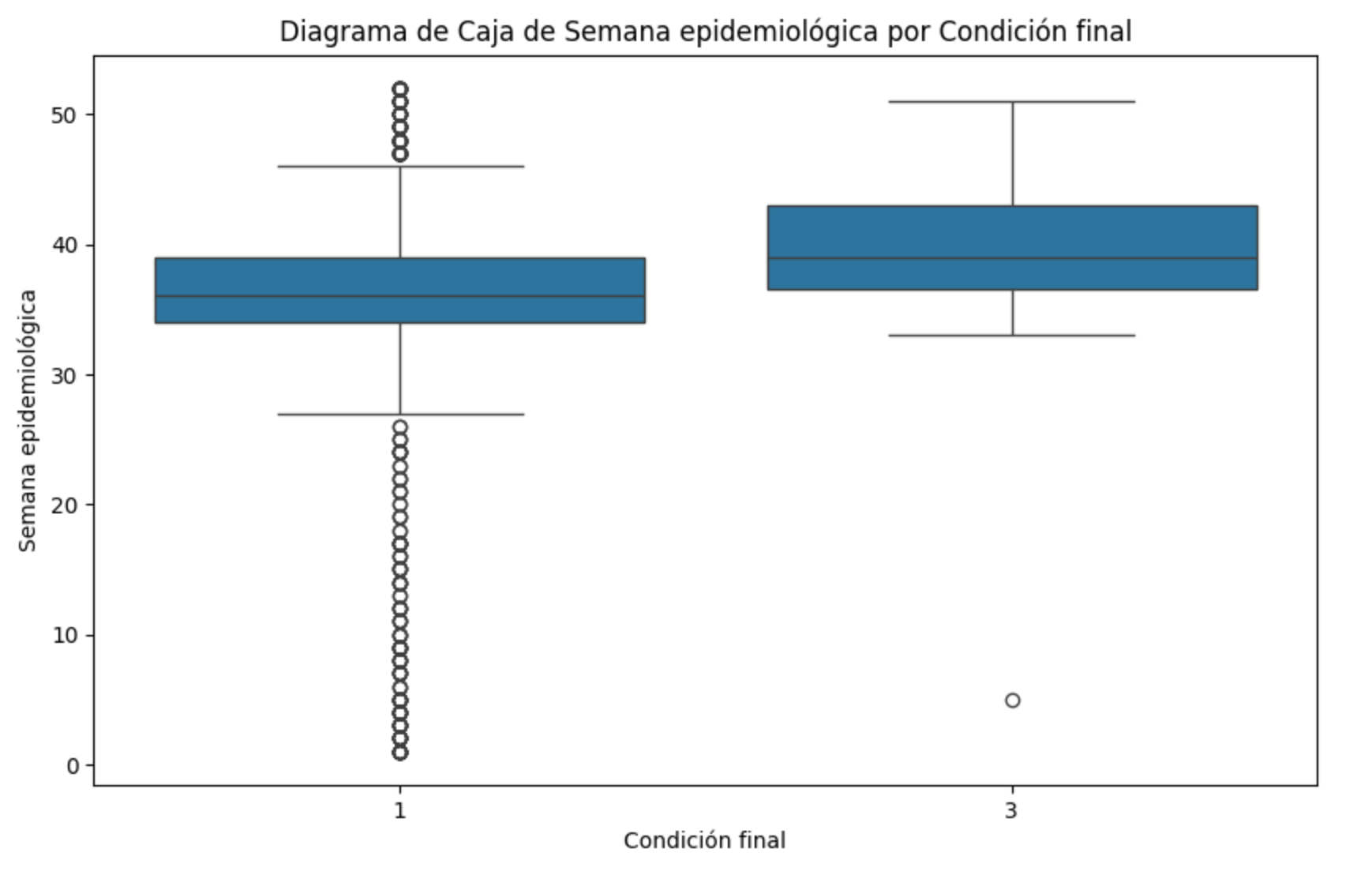


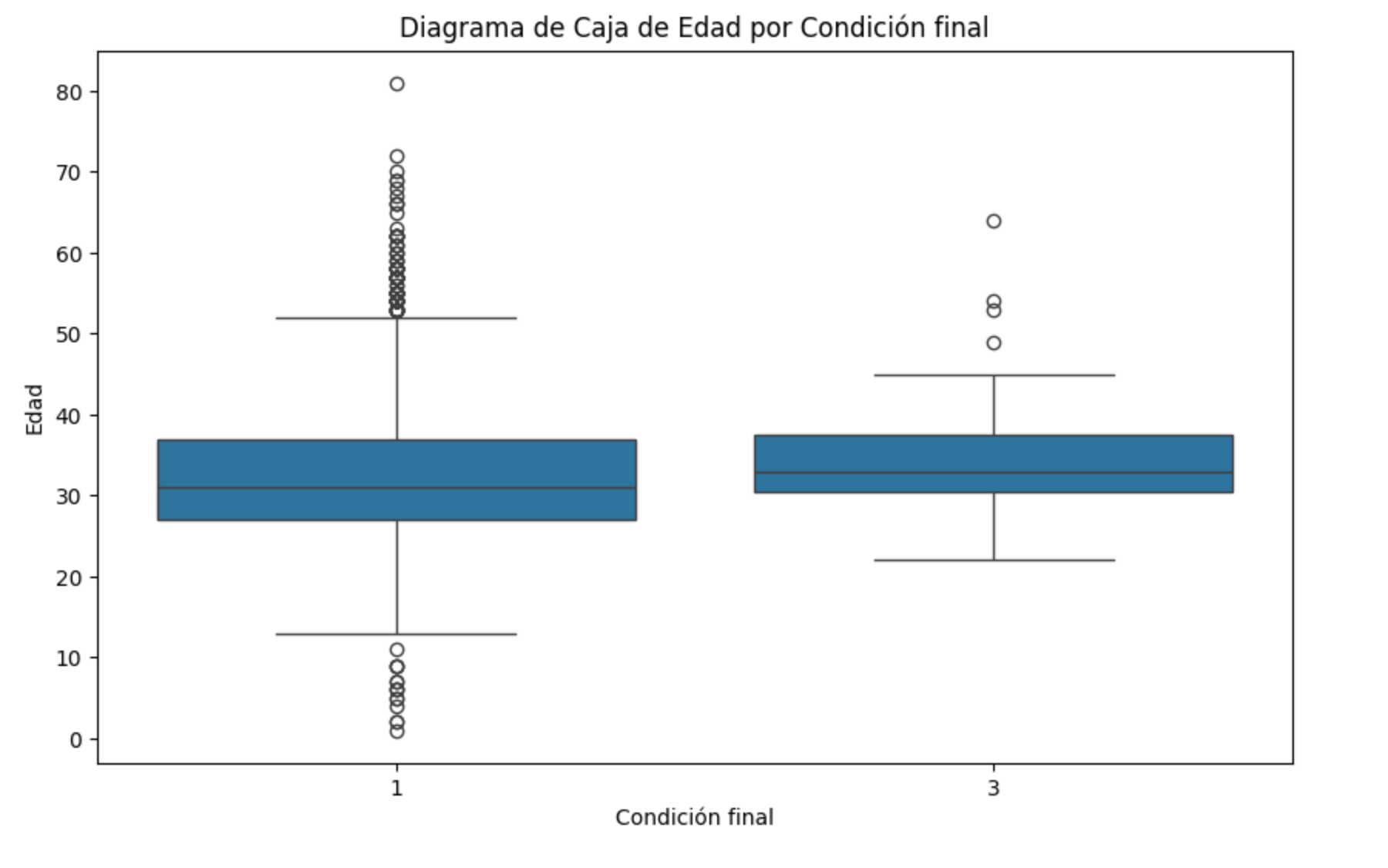
Condición final:

* Categoría 1: 99.2%
* Categoría 3: 0.8%

Relaciones clave:

* No diferencias significativas (p>0.05)
* Similar distribución temporal y etaria





### 

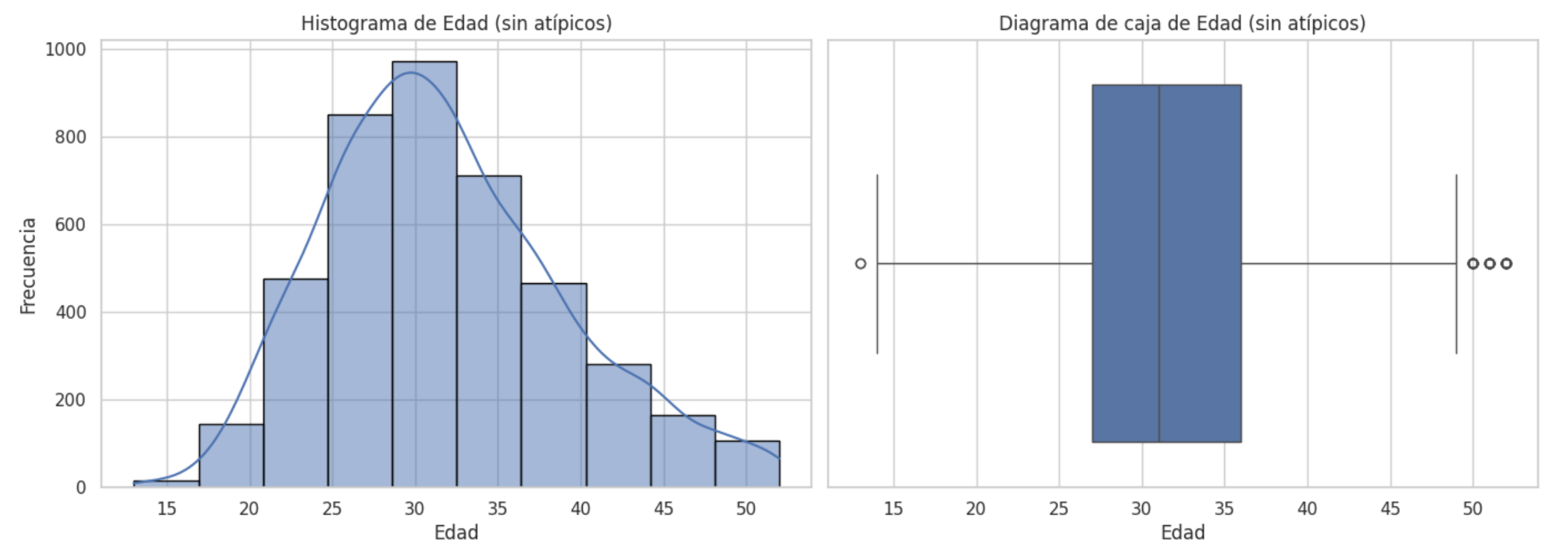
### 

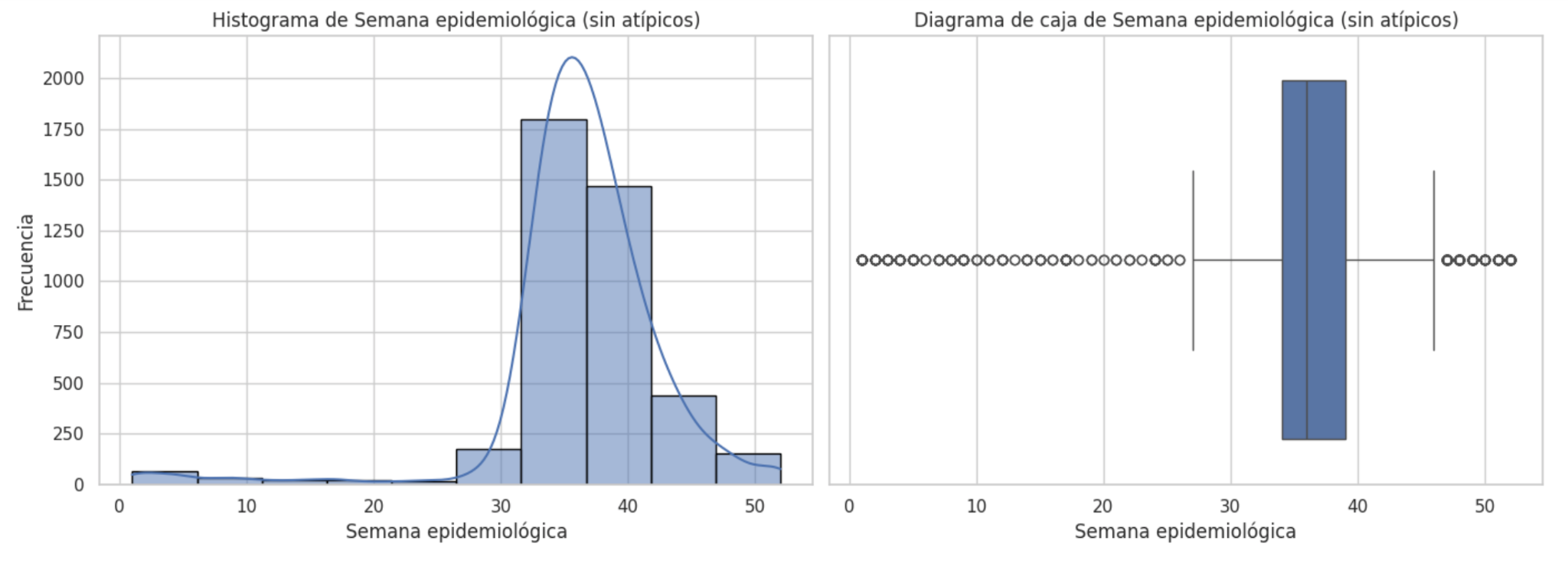
### 

### 3.5. Análisis de Sensibilidad

Exclusión de atípicos:

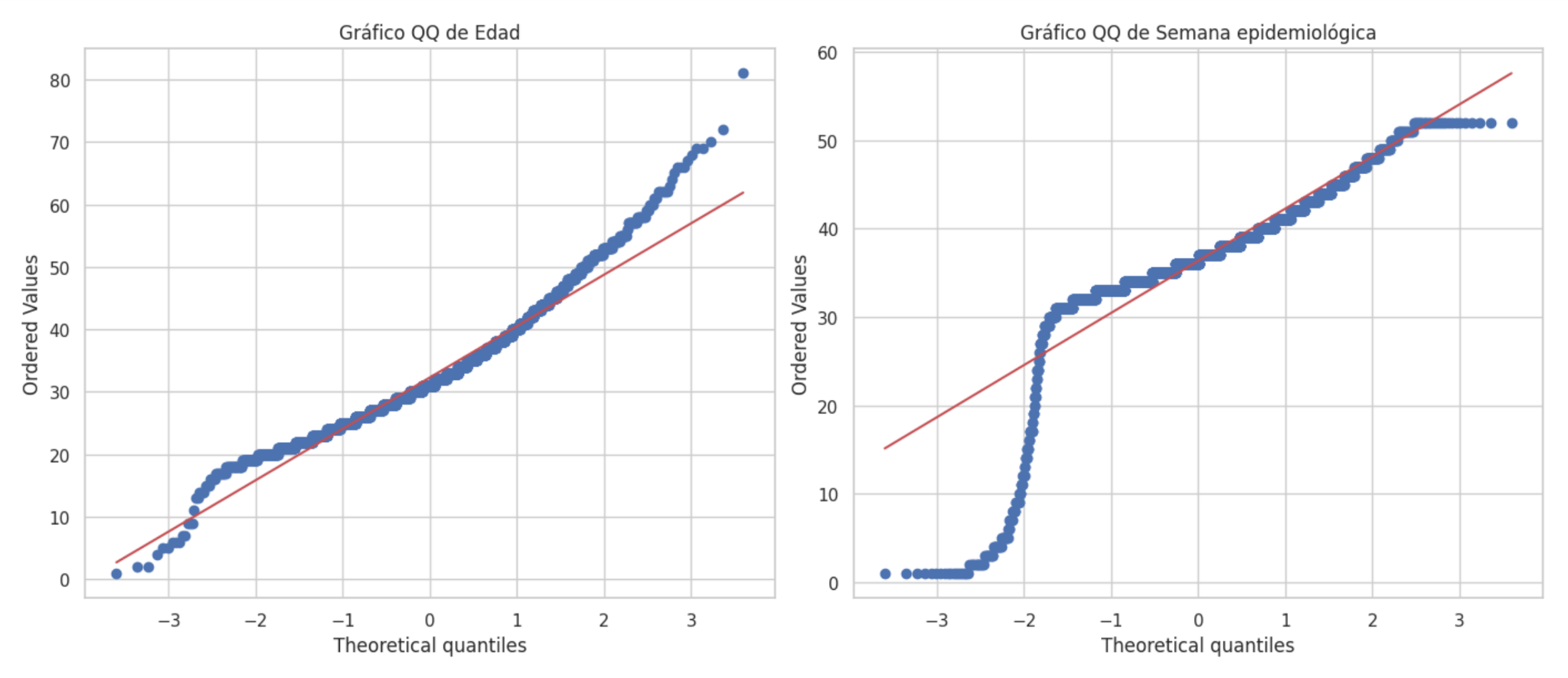
* Resultados consistentes tras exclusión
* Distribución mantiene características principales





Normalidad:

* Distribución no normal (p<0.001, Shapiro-Wilk)
* Justifica uso de pruebas no paramétricas



## 4. Conclusiones Profundizadas

El análisis de los casos de viruela en Colombia revela un brote marcadamente concentrado en población masculina joven (96.8% de los casos), con una edad promedio de 32 años y una clara focalización en áreas urbanas, especialmente Bogotá y Antioquia. La distribución temporal muestra un pico epidémico en la semana 37 de 2022, seguido de un descenso sostenido, lo que sugiere una combinación de inmunidad adquirida y efectividad de las medidas de control implementadas. Sin embargo, la calidad de los datos presenta limitaciones significativas, particularmente en variables clave como pertenencia étnica e historial de viajes, lo que impide un análisis más profundo de factores de riesgo sociales y demográficos. No se encontraron diferencias significativas en la distribución de edades o semanas epidemiológicas según la condición final del paciente, lo que podría indicar una homogeneidad en la presentación clínica o limitaciones en la clasificación de gravedad.

## 5. Recomendaciones Estratégicas

Para fortalecer la vigilancia epidemiológica y la respuesta a futuros brotes, se recomienda:

1) Implementar protocolos estandarizados para la recolección completa de variables sociodemográficas, especialmente en poblaciones étnicas y migrantes, que permitan identificar vulnerabilidades específicas.

2) Priorizar la capacitación del personal de salud en el registro de datos clínicos relevantes (comorbilidades, estado vacunal) para evaluar factores de gravedad.

3) Desarrollar estrategias de prevención dirigidas a la población masculina joven en áreas urbanas, incluyendo campañas de educación sexual y seguimiento de contactos.

4) Realizar estudios adicionales para entender los mecanismos detrás de la marcada disparidad por sexo, que podría estar asociada a comportamientos de riesgo o barreras en el acceso al diagnóstico para mujeres.

5) Optimizar los sistemas de notificación para reducir retrasos y mejorar la calidad de los datos en tiempo real. Estas acciones requieren coordinación intersectorial y asignación de recursos específicos para garantizar su implementación efectiva.