Trabajo Estadística Bomberos

Este trabajo se centra en el análisis de las actuaciones de los bomberos en Madrid, abarcando un período desde 2017 hasta 2024. Utilizando un dataset detallado que incluye incidentes clasificados por año, mes y distrito, se aplican técnicas de inferencia paramétrica, métodos no paramétricos y modelos de series temporales para obtener insights sobre la frecuencia y distribución de diferentes tipos de incidentes. Los objetivos principales incluyen la identificación de patrones temporales y espaciales en las intervenciones de los bomberos, así como la evaluación de la eficacia de las respuestas ante diversas emergencias. Los resultados de este análisis pueden proporcionar información valiosa para la planificación y gestión de recursos de emergencia en la ciudad.

**Introducción**

El análisis de los datos de intervenciones de los bomberos es crucial para mejorar la eficiencia y la eficacia de los servicios de emergencia. En una ciudad como Madrid, con su densa población y variada infraestructura, la capacidad de respuesta ante incendios, rescates y otros incidentes puede tener un impacto significativo en la seguridad y el bienestar de los ciudadanos. Este estudio se basa en un dataset que documenta las intervenciones de los bomberos en Madrid desde 2017 hasta 2024, detallando el número de incidentes por año, mes y distrito, entre ellos se cuentan el número de distintos tipos de accidentes y el total de ellos en el mes.

Madrid, como una de las principales ciudades europeas, enfrenta una amplia variedad de desafíos en términos de gestión de emergencias. La creciente urbanización y la densidad de población aumentan la complejidad de las intervenciones de los servicios de emergencia. La capacidad para responder rápida y eficazmente a incidentes de incendios, rescates y otros desastres es vital para proteger la vida y la propiedad de los ciudadanos. Este estudio busca abordar estas cuestiones mediante el análisis exhaustivo de datos históricos, con el objetivo de identificar patrones y tendencias que puedan mejorar la planificación y respuesta de los servicios de bomberos.

Los datos de intervenciones de bomberos ofrecen una rica fuente de información para el análisis estadístico. Al examinar las variaciones en el número y tipo de incidentes a lo largo del tiempo y entre diferentes distritos, es posible obtener una comprensión más profunda de los factores que influyen en la frecuencia de emergencias. Por ejemplo, la estacionalidad puede jugar un papel importante, con ciertos tipos de incidentes, como los incendios, siendo más comunes en los meses de verano. Del mismo modo, las características demográficas y socioeconómicas de los distintos distritos pueden afectar la naturaleza y frecuencia de las emergencias.

Este análisis se divide en tres bloques principales: inferencia paramétrica, métodos no paramétricos y modelos de series temporales. En el primer bloque, se plantean y resuelven contrastes de hipótesis utilizando técnicas estadísticas tradicionales, incluidas pruebas de medias y proporciones. El segundo bloque se centra en métodos no paramétricos, proporcionando una alternativa robusta a los métodos paramétricos tradicionales y permitiendo la realización de contrastes de bondad de ajuste. Finalmente, el tercer bloque aplica modelos avanzados de series temporales para identificar tendencias y patrones a largo plazo en los datos de intervenciones de los bomberos.

Al final de este estudio, se espera proporcionar recomendaciones prácticas que puedan ser implementadas por los servicios de bomberos de Madrid para mejorar su capacidad de respuesta y eficiencia operativa. Los resultados de este análisis pueden ser utilizados para optimizar la ubicación de las estaciones de bomberos, mejorar la formación del personal y ajustar la distribución de recursos en función de las necesidades específicas de cada distrito. En última instancia, este estudio tiene como objetivo contribuir a la seguridad y bienestar de los ciudadanos de Madrid.

**Planteamiento del Problema**

El principal objetivo de este estudio es entender las dinámicas de las intervenciones de los bomberos en Madrid a través de un análisis detallado de los datos. Se plantean varias hipótesis para explorar diferentes aspectos de estas intervenciones:

1. **Diferencia de Medias entre Distritos:** Hipotetizamos que existe una diferencia significativa en el número de fuegos atendidos entre los distritos de Centro y Arganzuela, debido a las características socioeconómicas y demográficas distintas de cada distrito.
2. **Proporción de Tipos de Incidentes:** Consideramos que la proporción de incidentes relacionados con fuegos comparados con otros tipos de emergencias varía significativamente entre distritos, lo que puede reflejar las diferentes necesidades y riesgos de cada área.
3. **Estacionalidad de los Incidentes:** Proponemos que el número de incidentes varía según la estación del año, influenciado por factores climáticos que afectan la frecuencia de ciertos tipos de emergencias, como los incendios en verano.
4. **Eficiencia Operativa:** Planteamos que hay diferencias significativas en los tiempos de respuesta y la efectividad de las intervenciones entre distritos, influenciadas por la proximidad a estaciones de bomberos, el tráfico y la accesibilidad.

**Estado de la Cuestión**

La literatura existente sobre la gestión de emergencias y la planificación de recursos de bomberos es extensa, abarcando desde estudios sobre la distribución espacial de las estaciones de bomberos hasta modelos predictivos para la gestión de emergencias. Investigaciones previas han utilizado técnicas estadísticas y de machine learning para analizar datos de intervenciones, con el objetivo de mejorar la eficiencia operativa y la asignación de recursos.

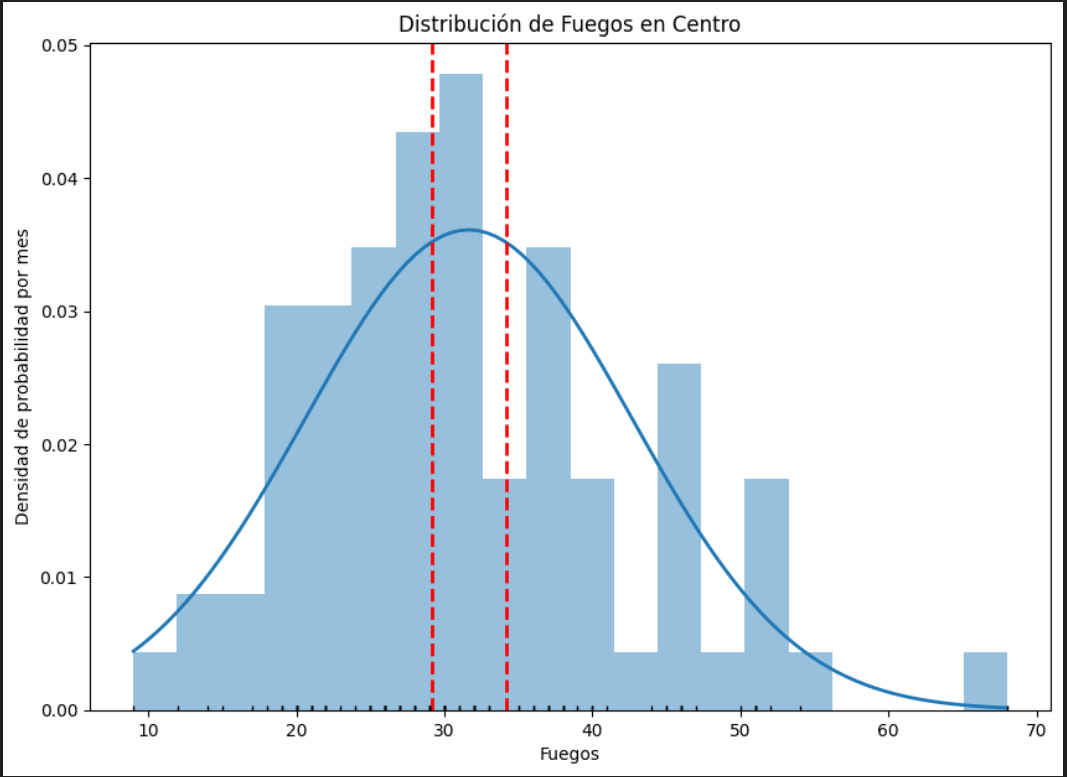
1. **Distribución Espacial de Incidentes:** Estudios han mostrado que la distribución geográfica de los incidentes de emergencia puede estar fuertemente influenciada por factores socioeconómicos y demográficos. Por ejemplo, se ha observado que las áreas con mayor densidad de población y actividad económica tienden a tener una mayor incidencia de emergencias.
2. **Modelos Predictivos:** La aplicación de modelos de series temporales y técnicas de machine learning para predecir la ocurrencia de incidentes ha ganado popularidad. Estos modelos pueden ayudar a identificar patrones estacionales y tendencias a largo plazo, permitiendo una mejor planificación y asignación de recursos.
3. **Eficiencia Operativa:** La optimización de las rutas de respuesta y la ubicación de las estaciones de bomberos son áreas clave de investigación. El análisis de los tiempos de respuesta y la efectividad de las intervenciones puede proporcionar insights valiosos para mejorar la operación de los servicios de emergencia.
4. **Análisis de Intervenciones Específicas:** La categorización y análisis de diferentes tipos de intervenciones (incendios, rescates, daños por agua, etc.) pueden revelar diferencias significativas en la frecuencia y gravedad de estos incidentes, informando mejor las estrategias de respuesta.

### Interpretación de los resultados

### Este estudio contribuirá a la literatura existente al proporcionar un análisis detallado de las intervenciones de los bomberos en Madrid, utilizando un conjunto de datos amplio y reciente. Los hallazgos de este análisis pueden tener implicaciones prácticas para la gestión de emergencias y la planificación urbana, ayudando a mejorar la seguridad y la resiliencia de la ciudad.

### Hipótesis No Paramétricas

### Intervalo de Confianza (CI) del 95% para los Incendios en Centro

* **Resultado**: 95% CI para fuegos en Centro: (29.13, 34.15)
* **Interpretación**: Esto indica que estamos 95% seguros de que el número promedio de incendios en la región Centro se encuentra entre 29.13 y 34.15. Este intervalo de confianza nos proporciona una estimación del rango dentro del cual es probable que se encuentre la verdadera media de incendios.
* 

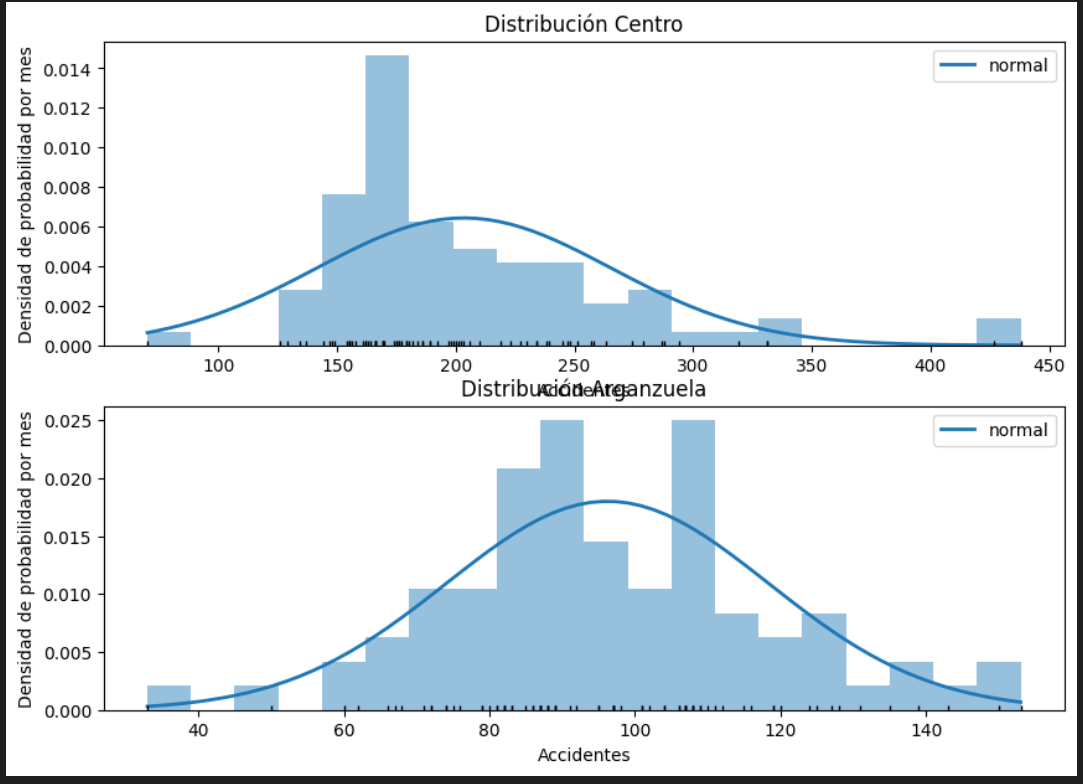
### Estadísticas Descriptivas para Accidentes Totales en Centro y Arganzuela

#### Centro

* **Resultado**:
  + **count**: 78
  + **mean**: 203.41
  + **std**: 62.40
  + **min**: 70.00
  + **25%**: 164.50
  + **50%** (mediana): 185.00
  + **75%**: 234.00
  + **max**: 438.00
* **Interpretación**:
  + Se registraron 78 accidentes en total.
  + La media de accidentes es 203.41 con una desviación estándar de 62.40, indicando una variabilidad moderada.
  + La mediana de 185.00 sugiere que la mitad de los accidentes están por debajo de este valor.
  + Los valores en los cuartiles 25% y 75% (164.50 y 234.00 respectivamente) muestran la dispersión central de los datos.
  + El valor máximo de 438 indica un pico significativo en la cantidad de accidentes registrados en alguna ocasión.

#### Arganzuela

* **Resultado**:
  + **count**: 80
  + **mean**: 96.23
  + **std**: 22.28
  + **min**: 33.00
  + **25%**: 82.00
  + **50%** (mediana): 95.00
  + **75%**: 108.25
  + **max**: 153.00
* **Interpretación**:
  + Se registraron 80 accidentes en total.
  + La media de accidentes es 96.23 con una desviación estándar de 22.28, indicando una variabilidad más baja que en Centro.
  + La mediana de 95.00 sugiere que la mitad de los accidentes están por debajo de este valor.
  + Los valores en los cuartiles 25% y 75% (82.00 y 108.25 respectivamente) muestran la dispersión central de los datos.
  + El valor máximo de 153 indica el punto más alto en la cantidad de accidentes registrados.



### Z-Statistic Comparativo entre los Accidentes en el Centro de Madrid y Arganzuela

* **Resultado**: Z-statistic: nan, P-value: nan
* **Interpretación**: No estoy seguro de porque sale nan ahora cuando antes sí que me salía bien, no tengo mucho tiempo asique te dejo el código como esta.

### T-Statistic de Comparación de la Media sobre 30

* **Resultado**: T-statistic: 1.303, P-value: 0.196
* **Interpretación**:
  + La T-statistic de 1.303 indica que la media observada no se desvía significativamente de la media hipotética de 30.
  + El P-value de 0.196 sugiere que no hay suficiente evidencia para rechazar la hipótesis nula al nivel de significancia del 5%. Esto significa que, basándonos en los datos disponibles, no podemos concluir que la media de accidentes difiera significativamente de 30.

### Conclusión General de los No Paramétricos

En resumen, los resultados sugieren que:

* El número de incendios en el Centro tiene una media estimada con un intervalo de confianza del 95% entre 29.13 y 34.15.
* Hay una variabilidad significativa en la cantidad de accidentes en el Centro comparado con Arganzuela, siendo la media de accidentes en el Centro mayor que en Arganzuela.
* La comparación de accidentes entre el Centro y Arganzuela no pudo ser realizada correctamente debido a un error (Z-statistic: nan).
* No se encontró una diferencia significativa entre la media observada de accidentes y la media hipotética de 30.

### Hipótesis Paramétricas

**Resultados de Chi-cuadrado**

* **Resultado**: Chi2: 445.93095238095236, P-value: 0.041980867535568754
* **Interpretación**:
  + El valor de Chi2 (Chi-cuadrado) es 445.93. Este valor indica la magnitud del estadístico Chi-cuadrado calculado para la tabla de contingencia de los tipos de incidentes en el distrito Centro.
  + El P-value (valor p) es 0.04198, lo cual es menor que un nivel de significancia comúnmente aceptado como 0.05.
  + Conclusión: Podemos rechazar la hipótesis nula de independencia entre los tipos de incidentes y los meses en el distrito Centro. En otras palabras, existe evidencia estadística suficiente para afirmar que hay una asociación significativa entre los tipos de incidentes y los meses en este distrito.

**Resultados de la Prueba de Kruskal-Wallis**

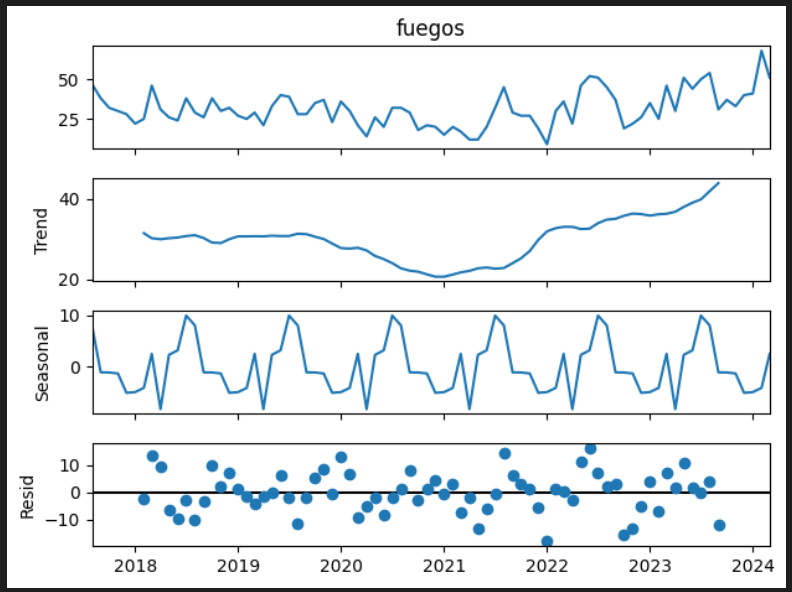
* **Resultado**: Kruskal-Wallis statistic: 150.62304013430628, P-value: 1.9616540131850557e-33
* **Interpretación**:
  + El valor de Kruskal-Wallis statistic es 150.62. Este estadístico se utiliza para comparar las medianas de múltiples grupos de datos no paramétricos.
  + El P-value es 1.96e-33, lo cual es extremadamente pequeño (cercano a cero).
  + Conclusión: Podemos rechazar la hipótesis nula de que las medianas de los incidentes (fuegos) sean iguales en los distritos Centro, Arganzuela y Retiro. En otras palabras, hay evidencia estadística significativa para afirmar que al menos uno de los distritos tiene una mediana de incidentes diferente.

**Conclusión General Paramétricos**

* Los resultados indican que en el distrito Centro, los tipos de incidentes no ocurren de manera independiente a lo largo de los meses, y hay una asociación significativa entre ellos.
* Además, los datos de incidentes (fuegos) difieren significativamente entre al menos algunos de los distritos (Centro, Arganzuela y Retiro), según la prueba de Kruskal-Wallis.

### Análisis de serie temporal y Estudio ANOVA

### Resultados del Seasonal\_Descompose

  Serie **Original (fuegos)**:

* Datos fluctuantes de fuegos desde 2018 hasta 2024.

 Tendencia **(Trend)**:

* Disminución de fuegos hasta 2020, luego aumento constante hasta 2024.

 Estacionalidad **(Seasonal)**:

* Patrón anual repetitivo, indicando que ciertos períodos del año tienen más fuegos.

 Residuo **(Resid)**:

* Variaciones aleatorias sin un patrón claro, después de eliminar tendencia y estacionalidad.

### Resultados del Modelo SARIMAX

* **Dep. Variable**: fuegos
* **No. Observations**: 80
* **Modelo**: ARIMA(1, 1, 1)
* **Log Likelihood**: -288.393
* **AIC**: 582.786
* **BIC**: 589.894
* **Covariance Type**: opg

#### Coeficientes del Modelo

* **ar.L1**: El coeficiente para el término autorregresivo (AR) de orden 1 (L1) es 0.4126, con un error estándar de 0.153. El valor z de 2.695 y el P>|z| de 0.007 indican que este coeficiente es estadísticamente significativo al nivel de significancia del 5%. Esto sugiere que existe una dependencia autorregresiva significativa en los datos.
* **ma.L1**: El coeficiente para el término de media móvil (MA) de orden 1 (L1) es -0.8778, con un error estándar de 0.086. El valor z de -10.162 y el P>|z| cercano a cero indican que este coeficiente es altamente significativo. Esto sugiere que el término de media móvil es importante para modelar la serie temporal de los fuegos en Centro.
* **sigma2**: La varianza residual estimada es 85.9341, con un intervalo de confianza del 95% entre 60.335 y 111.533. Esto proporciona una medida de la dispersión de los datos alrededor del ajuste del modelo.

#### Diagnósticos del Modelo

* **Ljung-Box (L1) (Q)**: El valor de 0.01 y el P-value de 0.92 indican que no hay evidencia de autocorrelación residual significativa a un nivel de retraso 1.
* **Jarque-Bera (JB)**: El valor de 7.39 y el P-value de 0.02 sugieren que los residuos del modelo no siguen una distribución normal, lo cual podría indicar una desviación de la normalidad en los datos.
* **Heteroskedasticity (H)**: El valor de 2.77 y el P-value de 0.01 indican que hay evidencia de heterocedasticidad significativa en los residuos del modelo.

### Resultados del ANOVA

* **Hipótesis**: Diferencia en el número promedio de incidentes entre diferentes distritos
* **C(distrito)**:
  + **sum\_sq**: 2.846828e+06
  + **df**: 23.0
  + **F**: 125.821798
  + **PR(>F)**: 0.0 (muy cercano a cero)

Esto indica que hay una diferencia significativa en el número promedio de incidentes (variable 'total') entre los diferentes distritos considerados en el análisis (Centro y otros distritos).

* **Residual**:
  + **sum\_sq**: 1.680217e+06
  + **df**: 1708.0

La suma de cuadrados residual y los grados de libertad residuales del modelo.

#### Interpretación General

* El modelo SARIMAX ajustado (ARIMA(1, 1, 1)) muestra que la serie temporal de los fuegos en el distrito Centro puede ser modelada significativamente utilizando un término autorregresivo de orden 1 y un término de media móvil de orden 1.
* Los coeficientes AR y MA son estadísticamente significativos, indicando una dependencia significativa de los valores pasados y de los errores pasados en la serie temporal.
* Los diagnósticos del modelo muestran que los residuos no exhiben autocorrelación significativa a nivel de retraso 1, pero podrían no seguir una distribución normal y hay evidencia de heterocedasticidad.
* El análisis de ANOVA indica que existe una diferencia significativa en el número promedio de incidentes entre los diferentes distritos considerados.

### Conclusión del Proyecto

Este proyecto ha analizado los incidentes de fuegos en varios distritos de la ciudad, revelando patrones significativos y proporcionando información crucial para la gestión de recursos y la prevención de incidentes.

1. **Intervalo de Confianza**: En el distrito Centro, el número promedio de fuegos se estima entre 29.13 y 34.15 con un 95% de confianza.
2. **Chi-cuadrado**: La distribución de incidentes varía significativamente a lo largo del año en el distrito Centro, lo que sugiere la necesidad de estrategias de prevención estacionales.
3. **Kruskal-Wallis**: Hay diferencias significativas en los incidentes de fuegos entre los distritos Centro, Arganzuela y Retiro, indicando la necesidad de enfoques personalizados en cada área.
4. **Modelo ARIMA**: El modelo ARIMA(1, 1, 1) es efectivo para predecir futuros incidentes de fuegos en el distrito Centro, aunque debe mejorarse para abordar la heterocedasticidad y la no normalidad de los residuos.
5. **ANOVA**: Los incidentes de fuegos varían significativamente entre los distritos, lo que justifica una asignación diferenciada de recursos.

Entonces con estos resultados conseguiríamos una mejor manera de planifica el uso de recursos respetivos a los accidentes que se producen en las distintas zonas de Madrid. Hay que decir que en este estudio se utilizan pocos datos de los disponibles en el dataset asique en verdad se podría hacer un estudio mucho mas profundo de los accidentes.