UNIVERSIDAD TECNOLOGICA DE PANAMA

FACULTAD DE INGENIERIA EN SISTEMAS

MAESTRIA EN ANALISIS DE DATOS

MATERIA: MODELOS DE PRONOSTICOS

PROFESOR: JUAN CASTILLO

**“Análisis Predictivo de Delitos Sexuales en Panamá (2020-2021) Utilizando Métodos de Holt, SARIMAX y Promedio Móvil”**

**Realizado Por: Jimenez Victoria**

**20-70-8021**

**INTRODUCCION**

**El presente informe tiene como objetivo analizar los delitos sexuales ocurridos en Panamá durante los años 2020 y 2021, utilizando métodos predictivos como Holt, SARIMAX y Promedio Móvil. Este análisis se enfoca en identificar patrones y tendencias en la ocurrencia de estos delitos, con el fin de proporcionar información valiosa para la formulación de políticas públicas y estrategias de prevención.**

**Para una mejor comprensión y análisis de los datos, se han agrupado las provincias de Panamá en tres categorías:**

1. **Panamá Centro: San Miguelito, Panamá Centro, Panamá Oeste, Colón.**
2. **Provincias Centrales: Coclé, Chiriquí, Veraguas, Herrera, Los Santos.**
3. **Provincias Restantes: Bocas del Toro, Darién, Comarca Ngäbe-Buglé, Comarca Guna Yala, Comarca Emberá.**

**Los delitos analizados corresponden al Capítulo 1, Título III del Código Penal de Panamá, que trata sobre la libertad e integridad sexual. Entre los delitos específicos se incluyen: acceso sexual con una persona mayor de catorce (14) años y menor de dieciocho (18) años, violación, corrupción de personas menores de edad, actos libidinosos, posesión de pornografía infantil, explotación sexual, entre otros.**

**PROBLEMÁTICA**

**La problemática de los delitos sexuales en Panamá es una preocupación creciente debido a su impacto negativo en la sociedad, especialmente en las víctimas, que en su mayoría son menores de edad. Los datos demuestran que la mayoría de estos delitos se llevan a cabo en Panamá Centro. Sin embargo, la diferencia entre el año 2020 y 2021 es apenas del -1.53%, lo que indica que no hay una significancia estadística notable en la variación anual. Los números de casos por año se mantienen relativamente constantes, sin mostrar estacionalidad y con una leve tendencia a disminuir.**

**Un aspecto preocupante es el aumento del 50% en los casos reportados en la Comarca Guna Yala. Este incremento es significativo y sugiere la necesidad de una atención especial en esta región. Además, es importante considerar que a partir de marzo de 2021 se decretó el cierre de Panamá debido a la pandemia de COVID-19, lo cual pudo haber influido en la dinámica de los delitos sexuales.**

**El análisis predictivo de los delitos sexuales utilizando métodos como Holt, SARIMAX y Promedio Móvil permitirá identificar tendencias y patrones que pueden ser utilizados para anticipar futuros incidentes y diseñar estrategias de intervención más efectivas. Este enfoque no solo contribuirá a una mejor comprensión de la dinámica de los delitos sexuales en Panamá, sino que también proporcionará una base sólida para la toma de decisiones informadas por parte de las autoridades competentes.**

**RESULTADOS**

**Panamá Centro**

**Grafico de Periodo y Demanda de los Dos anos**

**A graph with blue dots

Description automatically generated**

**Regresión Lineal**

Estadísticas de Regresión  
Multiple R (0.281): Indica una relación positiva débil entre los meses y la cantidad de denuncias presentadas.

R Square (0.079): Aproximadamente el 7.89% de la variabilidad en la cantidad de denuncias es explicada por los meses.

Adjusted R Square (0.037): Después de ajustar por los grados de libertad, sugiere que el modelo tiene una capacidad predictiva limitada.

Standard Error (32.10): En promedio, los valores reales se desvían de los valores predichos por aproximadamente 32.10 denuncias.

Observations (24): Número de observaciones en el análisis.

**ANOVA (Análisis de Varianza)**  
Regression df (1): Grados de libertad para la regresión.

Residual df (22): Grados de libertad para los residuos.  
Regression SS (1943.5): Suma de cuadrados de la regresión.  
Residual SS (22675.83): Suma de cuadrados de los residuos.  
F (1.886): Estadístico F.  
Significance F (0.184): Valor p para el test F, indicando que no hay una relación estadísticamente significativa entre los meses y la cantidad de denuncias.  
Coeficientes  
Intercept (279.58): Cuando X es 0 (lo cual no es posible en este contexto), Y sería aproximadamente 279.58 denuncias. Este coeficiente es estadísticamente significativo (p-valor muy bajo).  
X Variable 1 (-1.3): Cada incremento de un mes se asocia con una disminución promedio de 1.3 denuncias, pero este coeficiente no es estadísticamente significativo (p-valor alto).

**Modelo Holt**

**Para el periodo 2023 de Enero a Diciembre arrojo estos resultados de Periodo y Demanda**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 25 |  |  |  | **246.879375** |
| 26 |  |  |  | **245.297269** |
| 27 |  |  |  | **243.715163** |
| 28 |  |  |  | **242.133057** |
| 29 |  |  |  | **240.55095** |
| 30 |  |  |  | **238.968844** |
| 31 |  |  |  | **237.386738** |
| 32 |  |  |  | **235.804632** |
| 33 |  |  |  | **234.222526** |
| 34 |  |  |  | **232.64042** |
| 35 |  |  |  | **231.058314** |
| 36 |  |  |  | **229.476208** |

**MAD : 25.36337**

**TST: -0.035, 3.343**

**Modelo Sarimax:**

**Realizado en Python arrojo los siguientes valores :**

|  |  |
| --- | --- |
| **Periodo futuro** | **Promedio de Denuncias** |
| 24 | 221.463769 |
| 25 | 214.896924 |
| 26 | 210.708231 |
| 27 | 286.397498 |
| 28 | 199.313546 |
| 29 | 239.440772 |
| 30 | 222.406092 |
| 31 | 175.468397 |
| 32 | 238.156196 |
| 33 | 243.097909 |
| 34 | 148.11411 |
| 35 | 156.111905 |

**GRAFICA COMPARATIVA DEMANDA 2023 SARIMAX Y HOLT**

**A graph with red and blue dots

Description automatically generated**

**PROVINCIAS CENTRALES**

**Grafica de periodo y Demanda en provincias centrales**

**A graph with blue dots

Description automatically generated**

**Resultados de la regresión lineal:**

**Estadísticas de la Regresión**

* **Multiple R (Coeficiente de Correlación): 0.347. Esto indica una correlación positiva débil entre los meses y la cantidad de denuncias.**
* **R Square (Coeficiente de Determinación): 0.120. Esto sugiere que aproximadamente el 12% de la variabilidad en la cantidad de denuncias puede explicarse por los meses del año.**
* **Adjusted R Square: 0.080. Ajusta el R Square teniendo en cuenta el número de predictores en el modelo.**
* **Standard Error (Error Estándar): 25.25. Esto da una idea de la distancia típica que los valores observados caen desde la línea de regresión.**
* **Observations (Observaciones): 24. El número de puntos de datos utilizados en el análisis.**

**ANOVA (Análisis de Varianza)**

* **df (Grados de Libertad): 1 para la regresión y 22 para el residual, con un total de 23.**
* **SS (Suma de Cuadrados): 1920.17 para la regresión y 14023.66 para el residual. La suma de cuadrados residual es mucho mayor, lo que indica que gran parte de la variabilidad no es explicada por el modelo.**
* **MS (Media de Cuadrados): 1920.17 para la regresión y 637.44 para el residual.**
* **F (Estadístico F): 3.01. Compara la media de cuadrados entre grupos con la media de cuadrados dentro de los grupos.**
* **Significance F (Valor P del F): 0.097. Este valor es mayor que el nivel de significancia común (0.05), lo que sugiere que no hay suficiente evidencia para rechazar la hipótesis nula de que no hay relación lineal.**

**Coeficientes**

* **Intercept (Intersección): 198.07. Este es el valor predicho de Y cuando X es 0.**
* **X Variable 1 (Pendiente): -1.29. Esto sugiere que por cada incremento de un mes, la cantidad de denuncias disminuye en promedio en 1.29.**
* **Standard Error (Error Estándar): 10.64 para la intersección y 0.74 para la pendiente.**
* **t Stat (Estadístico t): 18.62 para la intersección y -1.74 para la pendiente.**
* **P-value (Valor P): 5.91E-15 para la intersección y 0.097 para la pendiente. El valor P para la pendiente es mayor que 0.05, lo que indica que no es estadísticamente significativa.**
* **Lower 95% y Upper 95% (Intervalo de Confianza del 95%): Para la intersección, el intervalo es de 176.01 a 220.13, y para la pendiente, de -2.84 a 0.25.**

**Modelo HOLT**

**Para el periodo 2023 género como resultado una demanda de :**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 25 |  |  |  | **165.7377805** |
| 26 |  |  |  | **164.0866538** |
| 27 |  |  |  | **162.4355272** |
| 28 |  |  |  | **160.7844005** |
| 29 |  |  |  | **159.1332738** |
| 30 |  |  |  | **157.4821472** |
| 31 |  |  |  | **155.8310205** |
| 32 |  |  |  | **154.1798939** |
| 33 |  |  |  | **152.5287672** |
| 34 |  |  |  | **150.8776406** |

**MAD: 21.599**

**TST: -1.502, 4.178**

**MODELO SARIMAX**

|  |  |
| --- | --- |
| **Periodo** | **Denuncias** |
| 24 | 95.876296 |
| 25 | 126.420876 |
| 26 | 179.076488 |
| 27 | 119.420667 |
| 28 | 154.331445 |
| 29 | 158.310842 |
| 30 | 124.805468 |
| 31 | 168.363163 |
| 32 | 133.118409 |
| 33 | 153.683297 |
| 34 | 141.731355 |
| 35 | 157.624347 |

**GRAFICA COMPARATIVA 2023 SARIMAX Y HOLT**

**A graph with dots and numbers

Description automatically generated**

**PROVINCIAS RESTANTES**

**Grafica de periodo y Demanda**

**A graph with blue dots and numbers

Description automatically generated**

**Regresión lineal**

**Estadísticas de la Regresión**

* **Multiple R (Coeficiente de Correlación): 0.3308. Esto indica una correlación positiva débil entre los meses y la cantidad de denuncias.**
* **R Square (Coeficiente de Determinación): 0.1094. Esto sugiere que aproximadamente el 10.94% de la variabilidad en la cantidad de denuncias puede explicarse por los meses del año.**
* **Adjusted R Square: 0.0689. Ajusta el R Square teniendo en cuenta el número de predictores en el modelo.**
* **Standard Error (Error Estándar): 21.43. Esto da una idea de la distancia típica que los valores observados caen desde la línea de regresión.**
* **Observations (Observaciones): 24. El número de puntos de datos utilizados en el análisis.**

**ANOVA (Análisis de Varianza)**

* **df (Grados de Libertad): 1 para la regresión y 22 para el residual, con un total de 23.**
* **SS (Suma de Cuadrados): 1240.72 para la regresión y 10099.24 para el residual. La suma de cuadrados residual es mucho mayor, lo que indica que gran parte de la variabilidad no es explicada por el modelo.**
* **MS (Media de Cuadrados): 1240.72 para la regresión y 459.06 para el residual.**
* **F (Estadístico F): 2.70. Compara la media de cuadrados entre grupos con la media de cuadrados dentro de los grupos.**
* **Significance F (Valor P del F): 0.114. Este valor es mayor que el nivel de significancia común (0.05), lo que sugiere que no hay suficiente evidencia para rechazar la hipótesis nula de que no hay relación lineal.**

**Modelo Holt**

**Resultados para el periodo 2023**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 25 |  |  |  | 104.1707 |
| 26 |  |  |  | 102.9269 |
| 27 |  |  |  | 101.6831 |
| 28 |  |  |  | 100.4392 |
| 29 |  |  |  | 99.19544 |
| 30 |  |  |  | 97.95162 |
| 31 |  |  |  | 96.70781 |
| 32 |  |  |  | 95.464 |
| 33 |  |  |  | 94.22019 |
| 34 |  |  |  | 92.97637 |

**MAD: 18.184**

**TST: -2.047, 3.65**

**MODELO SARIMAX**

|  |  |
| --- | --- |
| Periodo | Denunias de Pronosticos |
| 24 | 582.48456 |
| 25 | 801.018482 |
| 26 | 671.624766 |
| 27 | 785.281636 |
| 28 | 665.258985 |
| 29 | 724.329929 |
| 30 | 712.744622 |
| 31 | 757.114548 |
| 32 | 686.533896 |
| 33 | 726.21277 |
| 34 | 695.863371 |
| 35 | 1607.994989 |

Medidas de Precisión  
MAD (Desviación Absoluta Media): 66.95881327175188  
MAPE (Error Porcentual Absoluto Medio): 34.79180458329103%

**GRAFICA DE COMPARACION ENTRE AMBOS MODELOS**

**A graph with numbers and dots

Description automatically generated**

CONCLUSION

Viendo las graficas, se nota que la diferencia entre cada Modelo predictivo es poca, observando la data real sumnistrada por el ministerio publico Para los Delitos Sexuales 2023 en mayo dieron un total de 568 para todos los grupos, para el modelo Holt dio como resultados el mes de mayo 2023 507 y para el modelo Sarimax dio un estimado de 481 casos.

La data real recaudada da un valor aproximado al modelo Holt pero mucho mas alto de lo esperado, mostrando un aumento del 12.02% mas alto de lo esperado.

El resultado real de los primeros 10 meses del 2023 dan como resultado 5,389 Denuncias

En el modelo Holt, los resultados fueron:

-Provincias centro: 2397 denuncias

-Provincia Restante: 985

-Provincias Centrales: 1583

Para un total de 4965 denuncias, con una diferencia del 7.87% entre la data real a la data pronosticada en Holt.