

# Análisis de la Incidencia del Dengue usando Modelos de Poisson y Binomial Negativo

Este análisis tiene como objetivo modelar los casos de dengue semanales en los distritos del departamento de Loreto utilizando la temperatura mínima semanal y sus desfases (lags) como variables predictoras. Utilizaremos dos enfoques de modelado:

1. **Modelo de Poisson**, que es comúnmente utilizado para datos de conteo.
2. **Modelo Binomial Negativo**, que es más adecuado cuando hay evidencia de sobredispersión.

## Carga de datos

Primero, cargamos los datos del archivo CSV que contiene la información de los casos de dengue, la temperatura mínima y otras variables relevantes.

	ano	semana	distrito	ubigeo	total_casos	temp_min_semana	population	temp_min_lag1	ter
0	2017	1	ALTO NANAY	160102.0	0	22.098388	2855	NaN	
1	2017	1	ALTO TAPICHE	160502.0	0	20.769270	1515	NaN	
2	2017	1	ANDOAS	160706.0	0	21.732227	11714	NaN	
3	2017	1	BALSAPUERTO	160202.0	0	20.018025	13707	NaN	
4	2017	1	BARRANCA	160701.0	0	20.251695	12742	NaN	
...	...	...	...	...	...	...	...	...	...
16323	2022	52	URARINAS	160305.0	0	21.975947	14079	22.465956	
16324	2022	52	VARGAS GUERRA	160606.0	0	21.749073	7495	22.333004	
16325	2022	52	YAQUERANA	160511.0	0	21.580662	2232	22.079004	
16326	2022	52	YAVARI	160403.0	0	22.213742	9998	22.178608	
16327	2022	52	YURIMAGUAS	160201.0	43	21.598174	108870	22.324047	

16328 rows × 14 columns

## Definición del modelo y ajuste del Modelo de Poisson

A continuación, definimos la fórmula del modelo, que incluye las variables de temperatura mínima y sus desfases. Luego, ajustamos un modelo de Poisson.

### Generalized Linear Model Regression Results

Dep. Variable:	total_casos	No. Observations:	15964
Model:	GLM	Df Residuals:	15955
Model Family:	Poisson	Df Model:	8
Link Function:	Log	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-74385.
Date:	Tue, 22 Oct 2024	Deviance:	1.3884e+05
Time:	02:32:32	Pearson chi2:	6.80e+05
No. Iterations:	6	Pseudo R-squ. (CS):	0.08937
Covariance Type:	nonrobust		

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-4.3701	0.148	-29.501	0.000	-4.660	-4.080
temp_min_semana	0.1163	0.007	16.628	0.000	0.103	0.130
temp_min_lag1	0.0540	0.007	7.429	0.000	0.040	0.068
temp_min_lag2	0.0051	0.007	0.740	0.459	-0.008	0.019
temp_min_lag3	0.0295	0.007	4.296	0.000	0.016	0.043
temp_min_lag4	0.0047	0.007	0.687	0.492	-0.009	0.018
temp_min_lag5	0.0237	0.007	3.441	0.001	0.010	0.037
temp_min_lag6	0.0177	0.007	2.653	0.008	0.005	0.031
temp_min_lag7	-0.0270	0.006	-4.569	0.000	-0.039	-0.015

El resumen del modelo proporciona información sobre los coeficientes, su significancia y el ajuste general del modelo. Los coeficientes con p-valores bajos indican que son predictores significativos.

## Evaluación del ajuste del modelo de Poisson

Para evaluar la calidad del ajuste del modelo, calculamos dos métricas clave:

- **Devianza/Grados de Libertad (DF):** Mide la calidad del ajuste del modelo en relación con la variabilidad observada en los datos.
- **Pearson Chi2/DF:** Indica la sobredispersión en el modelo.

Devianza/DF: 8.702119903835104

Pearson Chi2/DF: 42.61988413750896

## Ajuste del Modelo Binomial Negativo

Dado que la evaluación del modelo de Poisson sugiere sobredispersión (Pearson Chi2/DF > 1), ajustamos un modelo Binomial Negativo, que es más adecuado para datos de conteo con sobredispersión.

## Generalized Linear Model Regression Results

=====						
Dep. Variable:	total_casos	No. Observations:	15964			
Model:	GLM	Df Residuals:	15955			
Model Family:	NegativeBinomial	Df Model:	8			
Link Function:	Log	Scale:	1.0000			
Method:	IRLS	Log-Likelihood:	-28116.			
Date:	Tue, 22 Oct 2024	Deviance:	40893.			
Time:	02:37:20	Pearson chi2:	2.54e+05			
No. Iterations:	9	Pseudo R-squ. (CS):	0.03357			
Covariance Type:	nonrobust					
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
Intercept	-3.9646	0.225	-17.588	0.000	-4.406	-3.523
temp_min_semana	0.1069	0.011	10.102	0.000	0.086	0.128
temp_min_lag1	0.0503	0.011	4.493	0.000	0.028	0.072
temp_min_lag2	0.0117	0.011	1.069	0.285	-0.010	0.033
temp_min_lag3	0.0305	0.011	2.775	0.006	0.009	0.052
temp_min_lag4	0.0062	0.011	0.569	0.569	-0.015	0.028
temp_min_lag5	0.0198	0.011	1.807	0.071	-0.002	0.041
temp_min_lag6	0.0150	0.011	1.386	0.166	-0.006	0.036
temp_min_lag7	-0.0350	0.010	-3.537	0.000	-0.054	-0.016
=====						

```
c:\Users\USUARIO\OneDrive\Documentos\Portafolio\Dengue-Incidence-Analysis\env\Lib\site-packages\statsmodels\genmod\families\family.py:1367: ValueWarning: Negative binomial dispersion parameter alpha not set. Using default value alpha=1.0.
warnings.warn("Negative binomial dispersion parameter alpha not ")
```

## Interpretación de los Coeficientes

- **Intercepto (Intercept):** -3.96. Este es el valor de la intersección cuando todas las variables predictoras son cero, lo que significa que cuando la temperatura mínima y sus desfases son cero, el logaritmo de los casos esperados es -3.96.
- **temp\_min\_semana:** 0.1069, **p-valor < 0.0001**. Este coeficiente es significativo, lo que indica que la temperatura mínima de la semana actual tiene una relación positiva y significativa con los casos de dengue. Por cada incremento unitario en la temperatura mínima semanal, el logaritmo del número de casos esperados aumenta en 0.1069.
- **temp\_min\_lag1:** 0.0503, **p-valor < 0.0001**. La temperatura mínima de la semana anterior también tiene una relación positiva con los casos actuales de dengue. Este coeficiente indica que un aumento en la temperatura mínima hace una semana se asocia con un aumento en los casos actuales de dengue.
- **temp\_min\_lag2:** 0.0117, **p-valor = 0.285**. Este coeficiente no es significativo, lo que sugiere que la temperatura mínima hace dos semanas no tiene un efecto estadísticamente significativo sobre los casos actuales de dengue.
- **temp\_min\_lag3:** 0.0305, **p-valor = 0.006**. La temperatura mínima hace tres semanas muestra un efecto significativo, aunque menor, sobre los casos actuales. Un aumento en la temperatura de hace

tres semanas se asocia con un aumento moderado en los casos de dengue.

- **temp\_min\_lag7**: -0.0350, **p-valor** < **0.0001**. Este coeficiente es significativo y negativo, lo que indica que la temperatura mínima de hace siete semanas tiene un efecto negativo sobre los casos actuales de dengue. Esto sugiere que una temperatura alta hace siete semanas puede estar relacionada con una disminución en los casos actuales.

## Conclusiones del Modelo

- Las temperaturas mínimas actuales y con desfases de 1 y 3 semanas parecen tener un impacto positivo significativo en la incidencia de dengue.
- La temperatura mínima de hace 7 semanas muestra un impacto negativo, lo que podría estar relacionado con efectos biológicos, como el ciclo de vida del mosquito vector del dengue.
- Aunque algunos coeficientes son estadísticamente significativos, los valores de los coeficientes son relativamente bajos, lo que indica que, aunque la temperatura influye en los casos de dengue, es probable que existan otros factores relevantes que no se están considerando en este modelo.

Este análisis sugiere que el uso de un modelo binomial negativo es adecuado debido a la sobredispersión presente en los datos, y que la temperatura mínima semanal tiene un efecto significativo pero moderado en los casos de dengue.

## Comparación entre los Modelos de Poisson y Binomial Negativo

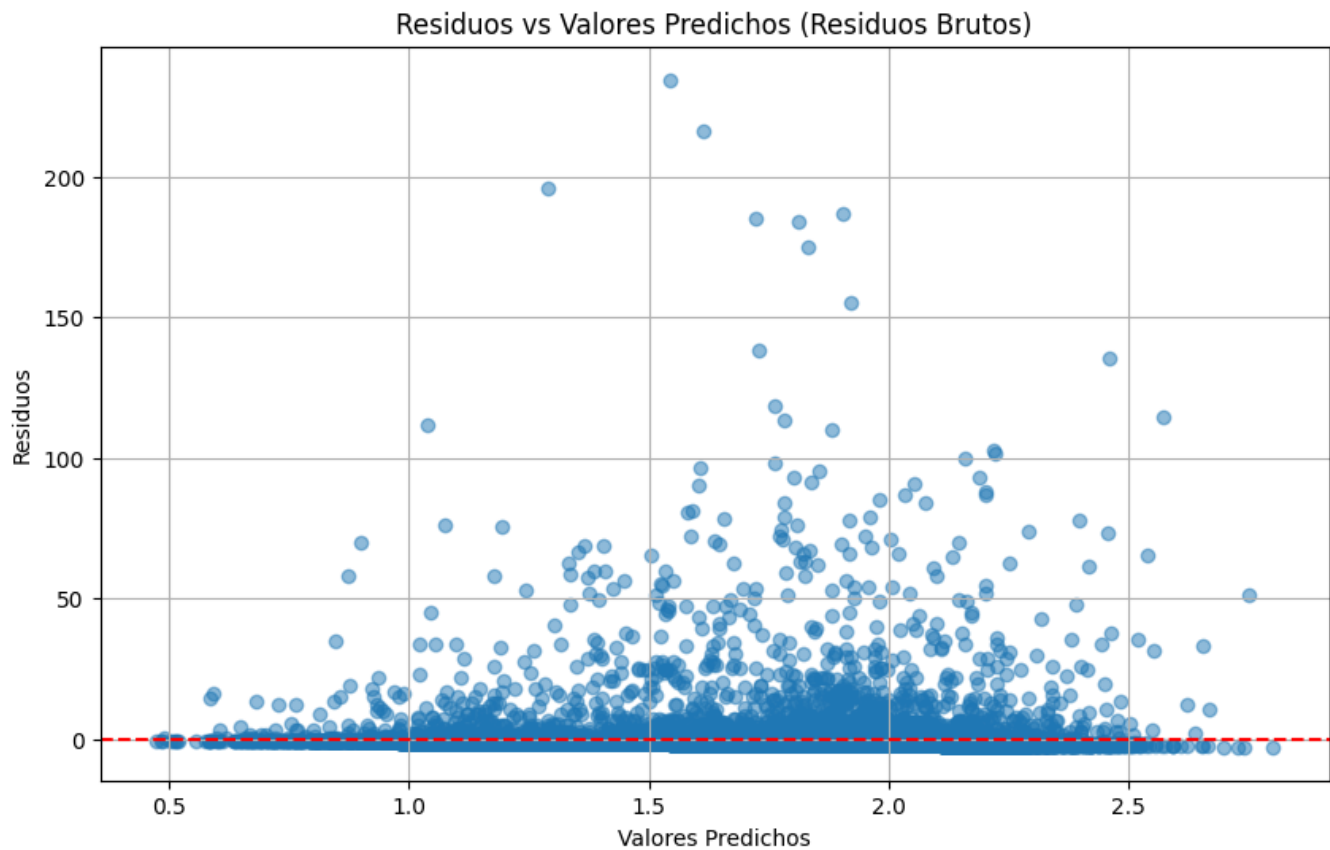
El criterio de información de Akaike (**AIC**) es utilizado para comparar modelos. Un AIC más bajo indica un mejor ajuste. A continuación, comparamos el AIC de ambos modelos.

AIC Poisson: 148788.749073087

AIC Binomial Negativo: 56249.34373379561

## Análisis de Residuos

El análisis de residuos nos permite verificar si el modelo se ajusta bien a los datos o si quedan patrones no explicados por el modelo. Comenzamos con los residuos brutos y los residuos de deviance.



## Residuos de Deviance

Los residuos de deviance proporcionan otra métrica para evaluar el ajuste del modelo. Si los residuos no muestran patrones claros y están distribuidos aleatoriamente alrededor de cero, esto indica un buen ajuste del modelo.

Residuos de Deviance vs Valores Predichos

