

Aplicación de modelos lineales generalizados: Caso Regresión Gamma para analizar información generada de estaciones meteorológicas.

Karla Reyes Maya
Marcelo Sanchez Zaragoza

Computo Estadístico, CIMAT



Diciembre 7, 2021

Section 1

1 Descripción del proyecto.

- Objetivos del proyecto.
- Información utilizada.

2 Modelo GML, caso: Gamma.

3 Implementación.

4 Resultados.

5 Conclusiones.

Descripción del problema

El caso consiste en el análisis de datos sobre información generada en estaciones meteorológicas. En particular sobre la precipitación en milímetros de 54 estaciones meteorológicas con información histórica de 302 semanas (aprox 6 años).

ESTACIONES METEOROLOGICAS

10VCZ Boca del Rio	Ciudad Juarez	La Paz	Mexico Oriente	Morelia	Saltillo
10VCZ Coatzacoalcos	Colima	Laguna	Mexico Reforma	Nogales	San Luis Potosi
10VCZ Cordoba	Cuernavaca	Laredo	Mexico Satellite	Oaxaca	Tampico
10VCZ Poza Rica	Culiacan	Leon	Mexico Valle	Obregon	Tijuana
10VCZ Xalapa	Durango	Los Mochis	Monclova	Pachuca	Toluca
Acapulco	Guadalajara Altos	Matamoros	Monterrey Centro	Piedras Negras	Tuxtla
Aguascalientes	Guadalajara Chapala	Merida	Monterrey Norte	Puebla	Vallarta
Cancun	Guadalajara Tequila	Mexicali	Monterrey Oriente	Queretaro	Villahermosa
Chihuahua	Hermosillo	Mexico Ajusco	Monterrey Sur	Reynosa	Zacatecas

Objetivos

Desarrollar un Modelo Lineal Generalizado Gamma con variables de regresión como: Media de la Precipitación, Tendencia y Estacionalidad que permita establecer un pronóstico anual sobre el valor Máximo de Precipitación en cada una de las 54 estaciones meteorológicas para el siguiente año.

Lo anterior se traduce en objetivos particulares como proporcionar una herramienta, a través de las estimaciones, que auxilie en la gestión y toma de decisiones como:

- Diseño pluvial en las ciudades
- Detección temprana de sequías

Datos Historicos

Semana	Plaza	Precipitación_MM
201553	Acapulco	0.007536232
201601	Acapulco	2.184099353
201602	Acapulco	0.132298137
201603	Acapulco	0
201604	Acapulco	0.188819876
201605	Acapulco	0.643229808
201606	Acapulco	0
⋮	⋮	⋮
202136	Acapulco	19.75043478
202137	Acapulco	32.45565217
202138	Acapulco	9.164720497
202139	Acapulco	36.24149068
202140	Acapulco	0

Cuadro: Visualización de los datos sin procesar para la Plaza meteorológica: Acapulco. El total de la semanas contempla de la semana 53 de 2015 a la semana 40 de 2021. Un aproximado de 6 años.

Tabla Por Año y Meses

Plaza	Año	Mes	Max_Precipitación_MM
Acapulco	2016	1	2.18409935
Acapulco	2016	2	1.20670807
Acapulco	2016	3	3.52956518
Acapulco	2016	4	0.53751553
Acapulco	2016	5	3.38608695
Acapulco	2016	6	8.00931668
Acapulco	2016	7	5.88645958
Acapulco	2016	8	17.1673294
Acapulco	2016	9	14.3450931
Acapulco	2016	10	4.02149069
Acapulco	2016	11	3.23018638
Acapulco	2016	12	0.64496894

Cuadro: Visualiación de los datos reducido por Año y Mes (con base en las semanas) y tomando el máximo de la precipitación mensual.

Section 2

1 Descripción del proyecto.

- Objetivos del proyecto.
- Información utilizada.

2 Modelo GML, caso: Gamma.

3 Implementación.

4 Resultados.

5 Conclusiones.

GML Gamma

La distribución de densidad gamma viene dada por :

$$f(y) = \frac{1}{\Gamma(r)} \left(\frac{1}{\lambda} \right)^r e^{-y/\lambda} y^{r-1}; \quad y \geq 0, r > 0, \lambda > 0$$

$$\theta = \frac{1}{\lambda r} = -\frac{1}{\mu}$$

$$\mu = r\lambda$$

$$Var \ y = \frac{\mu^2}{r}$$

Función Liga

En la siguiente tabla se muestra algunos enlaces canonicos utilizados regularmente:

Distribución	Enlace Canonico
Normal	$\eta_i = \mu_i$ (identity link)
Binomial	$\eta_i = \ln(\frac{\pi_i}{1-\pi_i})$ (logistic link)
Poisson	$\eta_i = \ln(\mu_i)$ (log link)
Exponencial	$\eta_i = \frac{1}{\mu_i}$ (reciprocal link)
Gamma	$\eta_i = \frac{1}{\mu_i}$ (reciprocal link)

- **log link** $\eta_i = \log \mu_i$
- **identity link** $\eta_i = \mu_i$

Section 3

- 1 Descripción del proyecto.
 - Objetivos del proyecto.
 - Información utilizada.
- 2 Modelo GML, caso: Gamma.
- 3 Implementación.
- 4 Resultados.
- 5 Conclusiones.

MTY
Modelando
el siglo XXI

Variables de regresión consideradas

Plaza	Max.Precipitacion_MM	Media.Precipitacion	Tendencia	mes_1	mes_2	mes_3	mes_4	mes_5	mes_6	mes_7	mes_8	mes_9	mes_10	mes_11	mes_12
Colima	6.79334039	1.70365196	2	1	0	0	0	0	0	0	0	0	0	0	0
Colima	0.17518797	0.08308271	3	0	1	0	0	0	0	0	0	0	0	0	0
Colima	3.14672401	0.94355533	4	0	0	1	0	0	0	0	0	0	0	0	0
Colima	0	0	5	0	0	0	1	0	0	0	0	0	0	0	0
Colima	2.09999997	1.29057465	6	0	0	0	0	1	0	0	0	0	0	0	0
Colima	9.01815243	6.61458105	7	0	0	0	0	0	1	0	0	0	0	0	0
Colima	14.4611171	11.9402792	8	0	0	0	0	0	0	1	0	0	0	0	0
Colima	14.4792693	11.8405745	9	0	0	0	0	0	0	0	1	0	0	0	0
Colima	20.6278197	11.9795059	10	0	0	0	0	0	0	0	0	1	0	0	0
Colima	5.56015044	4.15185284	11	0	0	0	0	0	0	0	0	0	1	0	0
Colima	8.46390972	3.55385604	12	0	0	0	0	0	0	0	0	0	0	1	0
Colima	0.04092374	0.01023094	13	0	0	0	0	0	0	0	0	0	0	0	1

Cuadro: Visualización de las variables de regresión, media mensual, tendencia(consecutiva) y estacionalidad(variable dummy).

Ejemplo GML Gamma

Ejemplo

```
plaza_model<-subset(plaza_ , select=c(mes_12))  
regresion_todos = glm(Max_Precipitacion_MM ~.,  
                      family = Gamma(link='log'),  
                      data = plaza_model)  
summary(regresion_todos)
```

Resultados

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.6197 -0.9759  0.0491  0.3113  1.9019

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.4908309   0.4212068   -1.165   0.2488
Media_Precipitacion  0.2282158   0.0492140   4.637 2.11e-05 ***
Tendencia     -0.0009423   0.0055234   -0.171   0.8651
mes_1         0.9570530   0.5446748    1.757   0.0843 .
mes_2         0.5927439   0.5444550    1.089   0.2809
mes_3         0.2819681   0.5445674    0.518   0.6066
mes_4        -1.5331108   0.5447754   -2.814   0.0067 **
mes_5         0.4929857   0.5458276    0.903   0.3702
mes_6         1.2105318   0.6724313    1.800   0.0771 .
mes_7         0.9601819   0.7203748    1.333   0.1879
mes_8         0.6857190   0.7507112    0.913   0.3649
mes_9         0.4912944   0.8787785    0.559   0.5783
mes_10        1.2002803   0.5856196    2.050   0.0450 *
mes_11        1.4991581   0.5788246    2.590   0.0122 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.8886784)

Null deviance: 260.60  on 70  degrees of freedom
Residual deviance: 131.47  on 57  degrees of freedom
AIC: 309.61

Number of Fisher Scoring iterations: 16

```

Figura: Resultado de ajustar el modelo para la variable Max.Precipitación y las variables de regresión. Se observa que la variable de tendencia no resulta significativa, mientras que la variable Media Mensual de la Precipitación resulta muy significativa. Los meses significativos fueron Abril, Octubre y Noviembre.

Ejemplo GML Gamma

Ejemplo GML

```
regresion_sim = glm( Max_Precipitacion_MM ~  
                    Media_Precipitacion +  
                    mes_1 + mes_2 + mes_3 +  
                    mes_4 + mes_5 + mes_6 + mes_7 +  
                    mes_8 + mes_9 + mes_10 + mes_11 ,  
                    family = Gamma(link='log') ,  
                    data = plaza_model )  
  
summary( regresion_sim )
```

Ejemplo GML Gamma

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.6277 -0.9946  0.0647  0.2941  1.8860

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.53211    0.38100  -1.397  0.16786
Media_Precipitacion 0.22897    0.04868   4.703 1.63e-05 ***
mes_1         0.97531    0.53904   1.809 0.07558 .
mes_2         0.59402    0.53874   1.103 0.27475
mes_3         0.30419    0.53872   0.565 0.57449
mes_4        -1.53314    0.53871  -2.846 0.00611 **
mes_5         0.49917    0.53956   0.925 0.35873
mes_6         1.21342    0.66520   1.824 0.07328 .
mes_7         0.95466    0.71258   1.340 0.18556
mes_8         0.68038    0.74248   0.916 0.36327
mes_9         0.49136    0.86930   0.565 0.57409
mes_10        1.20362    0.57760   2.084 0.04159 *
mes_11        1.50697    0.57234   2.633 0.01083 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.8704667)

    Null deviance: 260.60  on 70  degrees of freedom
Residual deviance: 131.49  on 58  degrees of freedom
AIC: 307.62

Number of Fisher Scoring iterations: 11

```

Figura: Resultado de ajustar el modelo para la variable Max_Precipitación y las variables de regresión.

Proyección de la Media

Mediante Regresión

```
# La proyeccion se realizo mediante regresion
# nota: lm construye el modelo de regresion
# implicitamente con los casos completos
```

```
lm_media <- lm (Media_Precipitacion ~
                mes_1+mes_2+mes_3+
                mes_4+mes_5+mes_6+
                mes_7+mes_8+mes_9+
                mes_10+mes_11 , predi_reg)

summary(lm_media)
new_mes_12 <- new_mes_12[12:25 ,]
pred.1 <- predict (lm_media , new_mes_12)
```


Proyección de la Media

Media_Precipitacion	Tendencia	mes_1	mes_2	mes_3	mes_4	mes_5	mes_6	mes_7	mes_8	mes_9	mes_10	mes_11	mes_12
0.244787849	74	1	0	0	0	0	0	0	0	0	0	0	0
0.347882327	75	0	1	0	0	0	0	0	0	0	0	0	0
0.139590964	76	0	0	1	0	0	0	0	0	0	0	0	0
0.004249088	77	0	0	0	1	0	0	0	0	0	0	0	0
0.031190102	78	0	0	0	0	1	0	0	0	0	0	0	0
0.618933147	79	0	0	0	0	0	1	0	0	0	0	0	0
1.746537182	80	0	0	0	0	0	0	1	0	0	0	0	0
3.304952670	81	0	0	0	0	0	0	0	1	0	0	0	0
3.511007297	82	0	0	0	0	0	0	0	0	1	0	0	0
1.218307384	83	0	0	0	0	0	0	0	0	0	1	0	0
0.978424905	84	0	0	0	0	0	0	0	0	0	0	1	0
0.339959296	85	0	0	0	0	0	0	0	0	0	0	0	1

Cuadro: Visualización de las proyecciones de la Media, la tendencia y las variables dummy de estacionalidad para la Estimación del Modelo GML Gamma.

Estimación del intervalo de confianza al 99 %

Los intervalos de incertidumbre (intervalos de confianza o de predicción) para valores ajustados son bastante fáciles si solo se trabaja con modelos lineales, pero cuando los modelos se vuelven más complejos, por ejemplo, modelos lineales generalizados, las facilidades para cuantificar la incertidumbre para las predicciones son más complicadas o inexistente.

Para calcular los intervalos de confianza se hace un intervalo en la escala del predictor lineal y luego aplicando la función de liga inversa g^{-1} del ajuste del modelo para transformar los intervalos de confianza del nivel lineal en el nivel de respuesta. Donde los intervalos están dados por:

$$g^{-1} \left(x' \hat{\beta} \pm z_{1-\alpha/2} \sqrt{\hat{\sigma} x' (X'X)^{-1} x} \right)$$

Proyección del Max. Precipitación por año

Proyección e Intervalo C.

```
# Hacemos la predicción con las var.
# Media + estacionalidad

prediccion <- predict(regresion_sim, predi_reg_t,
                      type="response")

mmaximos <- maximos$prediccion
maximo_ <- max(mmaximos)
Data_P <- data.frame(maximos)
# Usamos ciTools

conf <- add_ci(Data_P, regresion_sim,
               alpha = 0.1,
               names = c("lwr", "upr"))
```

Section 4

- 1 Descripción del proyecto.
 - Objetivos del proyecto.
 - Información utilizada.
- 2 Modelo GML, caso: Gamma.
- 3 Implementación.
- 4 Resultados.
- 5 Conclusiones.

MTY
Modelando
el siglo XXI

Resultados

Plaza	2015	2016	2017	2018	2019	2020	2021	2022	Intervalo C.	
Colima	0	20.6278	18.9683	24.7124	20.3172	17.431	45.2791	24.8179	13.1273	46.9018
Cuernavaca	0.2167	10.7571	16.0286	14.3289	15.9934	16.2607	12.0088	14.4555	8.5375	24.4625
Culiacan	0	11.7569	11.7435	17.7607	10.1676	13.9238	18.9024	26.3554	11.1664	62.1342
Oaxaca	0.0993	19.756	34.3143	19.0423	13.4268	23.7375	30.3191	17.5918	9.6294	32.1384
San Luis Potosi	0.206	6.2714	9.3265	12.0613	6.0824	7.9719	10.8801	7.6612	5.8311	10.0657
Puebla	0.8114	11.0413	10.8808	8.5227	14.1686	12.2633	17.1162	13.123	9.5791	17.9781
Queretaro	0.7008	10.5099	8.8434	13.6853	10.1803	9.7124	13.4652	10.4578	8.2395	13.2735
Reynosa	11.1771	8.9709	6.9045	12.1516	11.2895	29.2805	29.5235	11.8642	6.2751	22.4312
Saltillo	4.2	7.3857	6.2857	8.2	3.9844	26.2817	17.7667	5.1867	3.5714	7.5327
San Luis Potosi	0.206	6.2714	9.3265	12.0613	6.0824	7.9719	10.8801	7.6612	5.8311	10.0657
Zacatecas	0.4269	9.5436	15.4575	25.6081	10.8178	13.965	22.0742	10.7068	5.581	20.5403

Cuadro: Historico de la precipitación de 2015-21 así como la estimación para 2022 y e intervalo de confianza.

Resultados

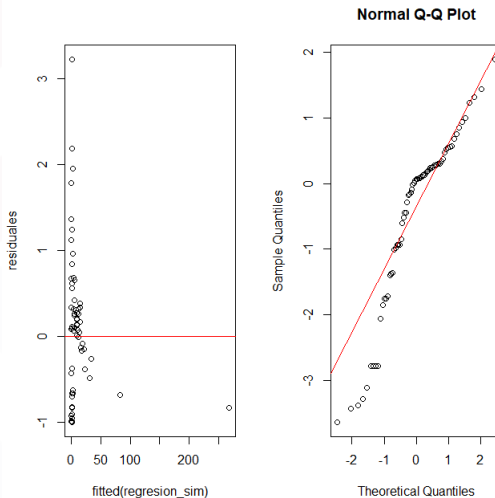
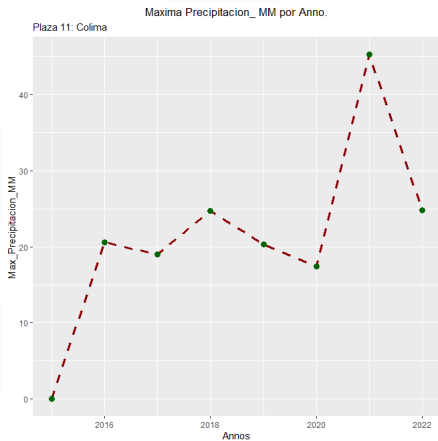
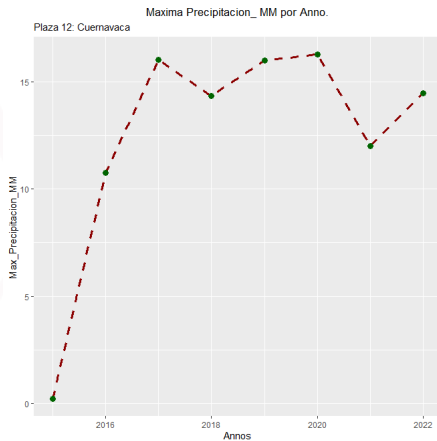


Figura: Residuales del modelo GML Gamma para la plaza 11 Colima.

Resultados

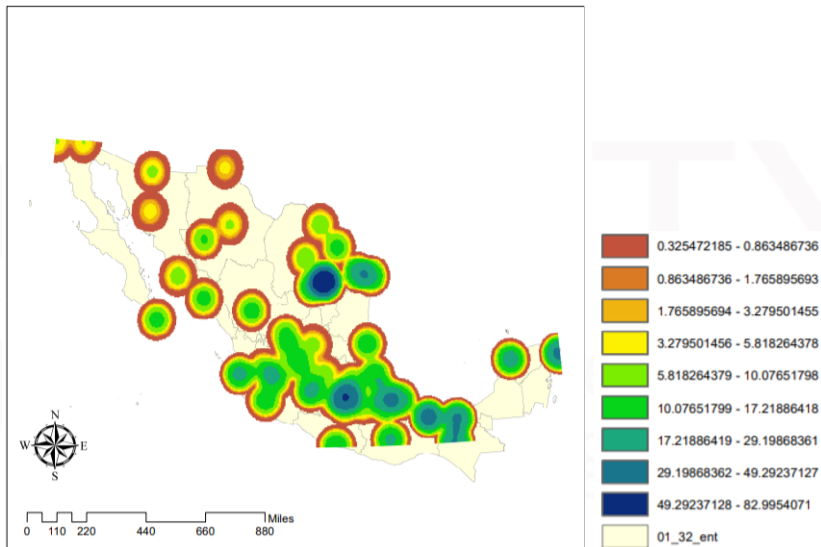


(a) Plaza 11: Colima



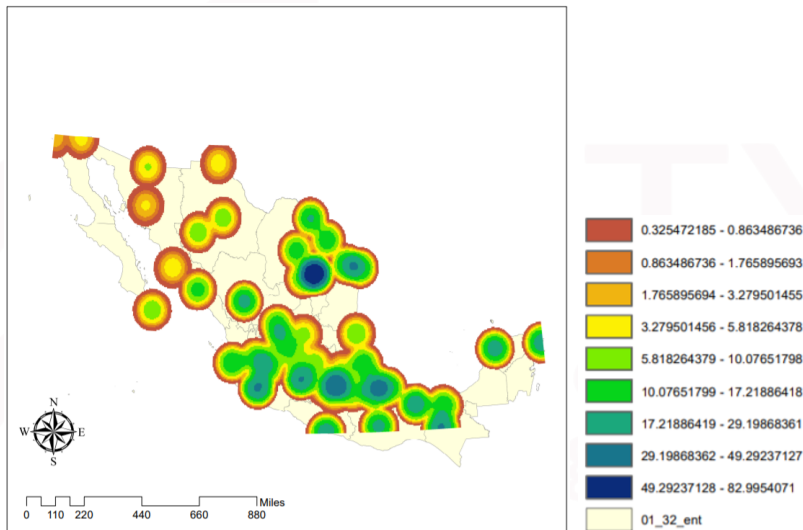
(b) Plaza 12: Cuernavaca

Figura: Grafico de los Valores Máximos de Precipitación en los años 2015-2021 y el estimado para 2022.



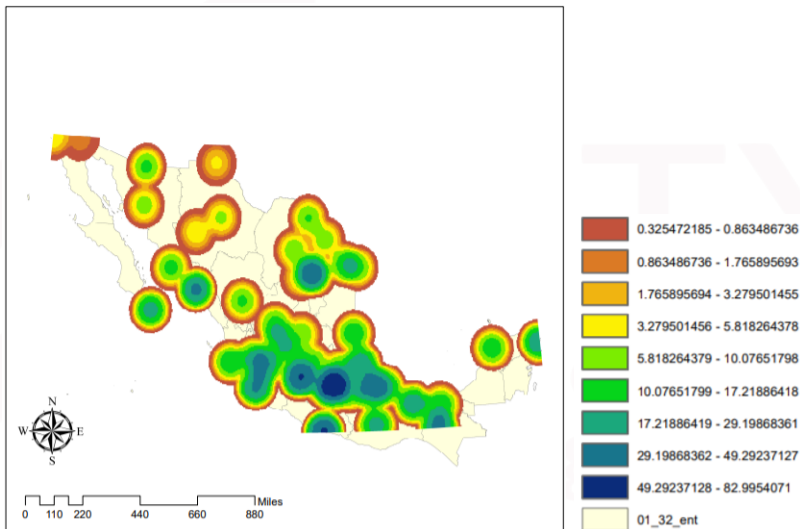
(a) 2020

Figura: Precipitación Máx Historico para 2020.



(a) 2021

Figura: Precipitación Máx Historico para 2021.



(a) 2022

Figura: Precipitación Máx Estimada para 2022.

Contraste

Plaza	Estimación	Intervalos	Registro
Ciudad Juarez	0.71	(0.4791-1.03)	0.49
Zacatecas	1.36	(1.12-6.26)	1.5

Section 5

- 1 Descripción del proyecto.
 - Objetivos del proyecto.
 - Información utilizada.
- 2 Modelo GML, caso: Gamma.
- 3 Implementación.
- 4 Resultados.
- 5 Conclusiones.

MTY
Modelando
el siglo XXI

Conclusiones.

Respecto a la estimación del valor máximo de precipitación anual para las estaciones meteorológicas,

- La información de la media de la precipitación mensual resultó una variable significativa en la mayoría de los modelos.
- Mientras que en casi todos los casos, aunque haya un coeficiente de tendencia negativo o positivo, no resultó realmente significativo aportando menor información.
- Así, sí la precipitación se une a la Media y a la estacionalidad para lograr una estimación del máximo mensual por año se obtuvieron mejores resultados

Gracias! :)