



Universidad Nacional Autónoma de México  
Facultad de Ciencias UNAM

# **Análisis de criminalidad en la CDMX en 2020 bajo un enfoque Bayesiano**

Estadística Bayesiana  
Semestre 2021-1

Presentan: Becerra, Flores, Hernández, López, Maqueda  
Enero del 2021



- ▶ La violencia que sufre actualmente la población mexicana se traduce en un problema de seguridad pública, tanto por la dimensión que ha adquirido la muerte por dichas causas, como por los efectos materiales y emocionales que ocasiona en la población.
- ▶ En el presente proyecto se busca utilizar técnicas y métodos bayesianos para poder evaluar el nivel de delincuencia en la capital del país. Para esto se trabaja con las carpetas de investigación emitidas por la Fiscalía General de Justicia (PGJ) en el 2020. Para fines de este proyecto, y debido a la gran dimensión de los datos, sólo se trabajará con la información disponible de Enero a Octubre del 2020, que contiene hasta 150,000 carpetas de investigación que se han abierto tan sólo en esa fracción del año.

# Antecedentes de Crimen en la CDMX

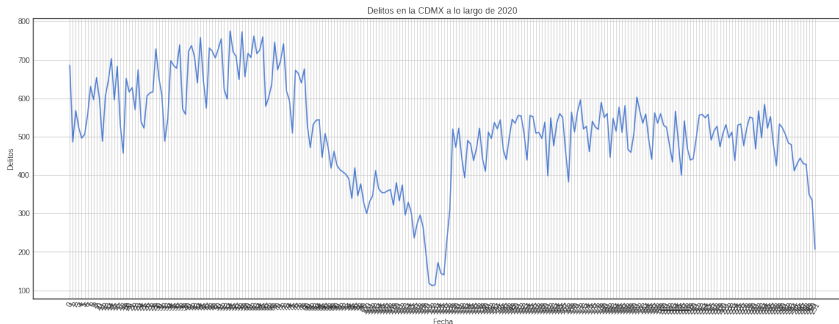


- ▶ La ciudad de México tiene un alto índice delictivo ya sea por el control de carteles del narcotráfico o asociaciones de fraude y extorsión a pequeños comerciantes en la ciudad, o simplemente factores de delitos menores. El propósito principal del cual deriva este trabajo es inferir un poco acerca del comportamiento del crimen en la Ciudad de México y el caso especial que ha representado 2020 para la historia del crimen en la CDMX.
- ▶ 2020 ha sido un año atípico para el crimen en la ciudad de México, la cuarentena ha modificado los ritmos de vida y de trabajo de los capitalinos, y la actividad de grupos criminales dentro de la ciudad no ha sido la excepción. A lo largo del trabajo veremos, analizaremos y cuantificaremos el impacto que ha tenido la pandemia durante gran parte del 2020 en la ciudad de México.

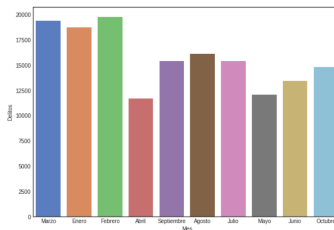


- Como parte del análisis descriptivo de nuestros datos, se pueden identificar ciertos factores importantes en cuanto a la temporalidad, localización y tipo de los incidentes delictivos en la Ciudad de México.

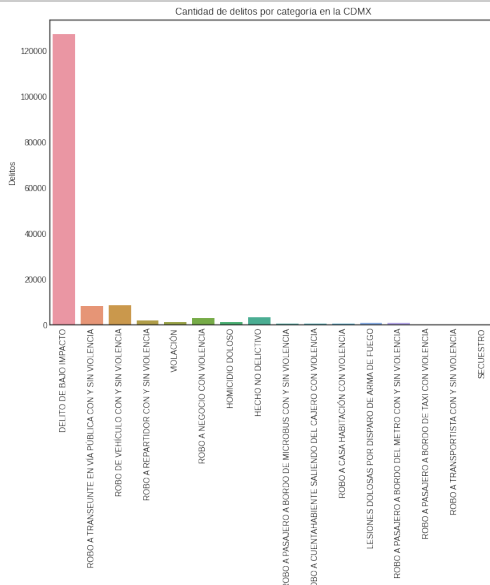
- Como se muestra a continuación, se puede distinguir fácilmente la causalidad entre los meses con más manifestaciones delictivas en la entidad, que resultan ser los primeros del año, mismos que corresponden al periodo de tiempo anterior al reconocimiento de la COVID-19 como una pandemia global que eventualmente alcanzó a la Ciudad de México



- Estas circunstancias principalmente afectan la distribución de los sucesos delictivos en los periodos siguientes, pues la movilidad de la población es evidentemente benéfica para delincuentes



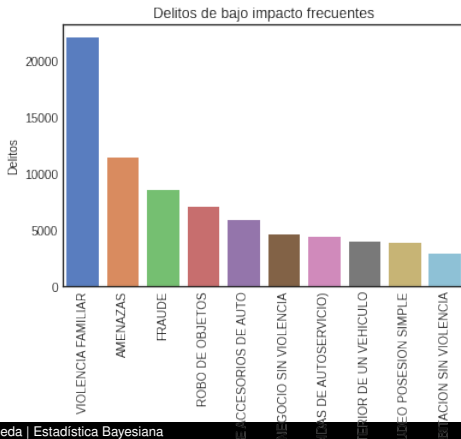
# Analisis exploratorio de datos



# Analisis exploratorio de datos



- ▶ A través de los datos trabajados se puede observar que predominan los delitos de bajo impacto en la CDMX, e identificamos los más concurrentes, y exploramos el tipo de delito que con mayor frecuencia es reportado/propiciado en la ciudad.







- ▶ Se realizaron algunas inferencias gaussianas sobre la distribución de los delitos en la Ciudad de México asumiendo que los datos siguen una distribución gaussiana, para esto se utiliza la biblioteca Pymc3 en Python, que nos permite realizar análisis Bayesiano.
- ▶ Como no conocemos la media ni la desviación estándar entonces debemos de elegir distribuciones prior no informativas para tratar de aproximar por medio de un modelo bayesiano el verdadero valor de este.

Un modelo razonable podría ser:

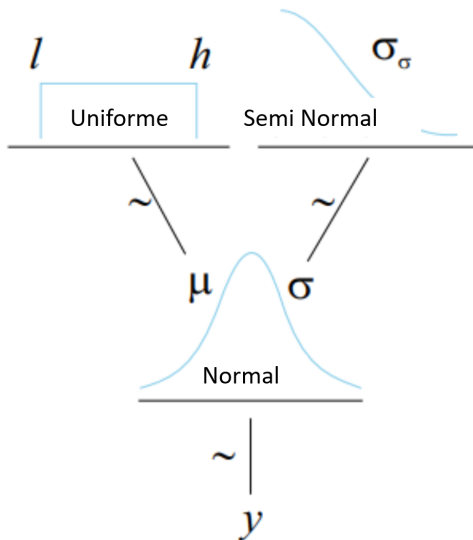
$$\mu \sim \text{Uniforme}(l, h)$$

$$\sigma \sim \text{SemiNormal}(\sigma_h)$$

$$y \sim \text{Normal}(\mu, \sigma)$$

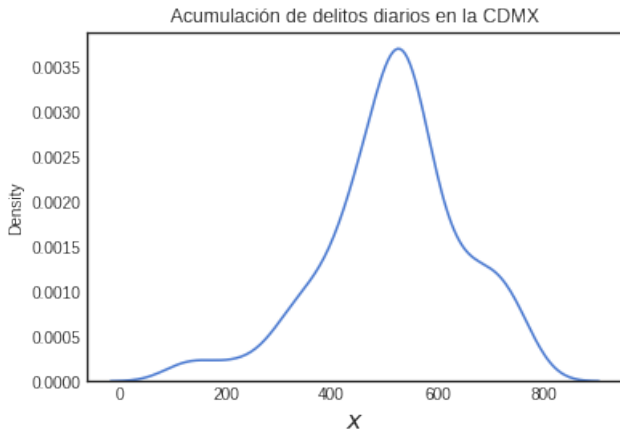
Así,  $\mu$  viene de una distribución uniforme entre  $l$  y  $h$  que son desconocidos y que elegiremos como parámetros iniciales no informativos, y  $\sigma$  viene de una distribución normal truncada o Semi Normal. El diagrama de distribución del algoritmo es el siguiente:

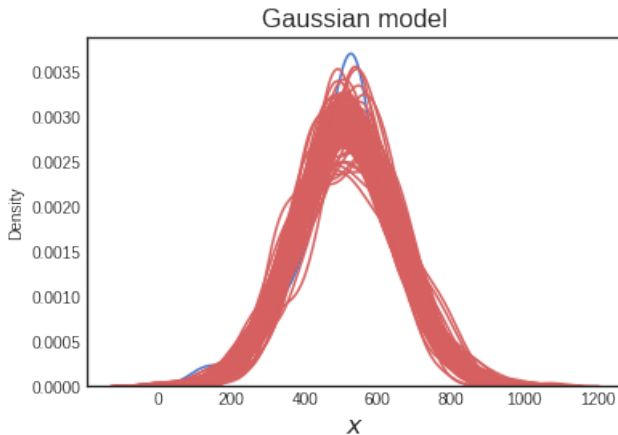
# Inferencias Gaussianas





- ▶ Este primer modelo muestra todos los posibles valores que se calcularon para delitos diarios dentro de la Ciudad de México.
- ▶ Calculamos 1600 predicciones para los delitos diarios, esto para poder obtener una distribución más confiable de los sucesos que en realidad ocurrieron en la CDMX, pues no todos los delitos que suceden en la ciudad son registrados, pues para esto necesita hacerse una denuncia e iniciar una carpeta de investigación.

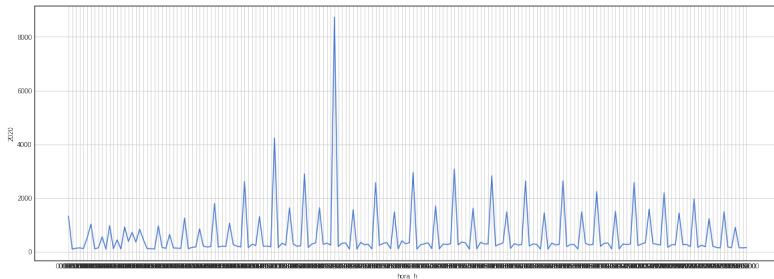






- ▶ Al realizar la predicción posteriori por un método bayesiano obtenemos una aproximación más confiable de la verdadera distribución de los delitos en la ciudad.
- ▶ Obtenemos una media de 514.77 delitos con una desviación estándar de 131.57
- ▶ De manera análoga al modelo con los delitos por día, realizamos un análisis de los delitos por hora en la Ciudad de México para poder observar aquellos intervalos de tiempo con más incidencia delictiva, obteniendo como resultado la siguiente gráfica:

# Inferencias Gaussianas



- ▶ Ya observamos que los delitos parecen tener una periodicidad más o menos homogénea, pero existe un punto que llama nuestra atención por su particularidad, se trata de los delitos que se disparan a más de 8000 en una simple hora en el transcurso del día.
- ▶ En la siguiente tabla se muestran cuáles son las horas donde ocurren más delitos en la ciudad, y contrario a lo que se esperaba (que los delitos se acumularan en la tarde-noche) la mayoría de los delitos se acumulan de hecho en la mañana y en el medio día, posiblemente tratándose de delitos de bajo impacto, que como ya se comentaba, conforman una gran porción de los sucesos delictivos y se llevarían a cabo en las horas donde hay más movilidad en las calles, comercios y transportes.

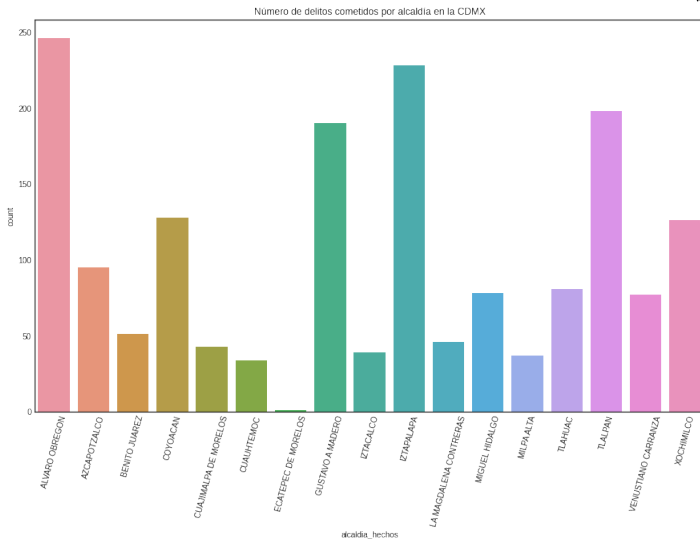
Número de delitos por hora en la Ciudad de México										
Hora	12:00	10:00	15:00	14:00	11:00	16:00	18:00	17:00	09:00	20:00
# Delitos	8733	4233	3080	2946	2904	2829	2638	2633	2616	2578





El problema al que nos enfrentamos ahora es el de observar las diferencias entre distintos tipos de grupos, específicamente por alcaldías, pues queremos hacer inferencias sobre estos parámetros para poder observar si es que hay alguna diferencia entre los niveles de crimen entre las alcaldías Alvaro Obregón y Gustavo A. Madero, por citar un ejemplo. y tomemos Para esto primero observemos que los niveles de crimen en las diferentes alcaldías de la ciudad de México no es el mismo como se puede comprobar en la siguiente gráfica:

# Modelos de comparacion de grupos

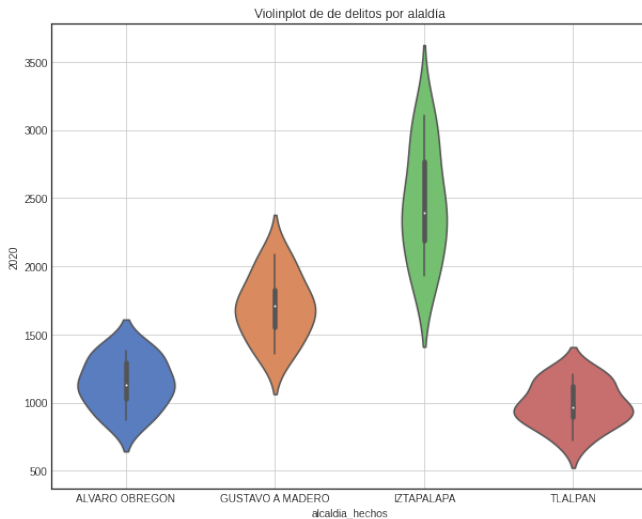




De igual forma nos apoyaremos de un modelo bayesiano gaussiano, pero ahora, contrastaremos al mismo tiempo 4 diferentes tipos de muestras (que corresponden a las alcaldías con mayores incidencia criminal):

- ▶ Alvaro Obregón
- ▶ Gustavo A. Madero
- ▶ Iztapalapa
- ▶ Tlalpan

# Modelo de comparación de grupos



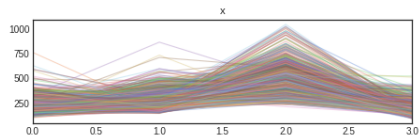
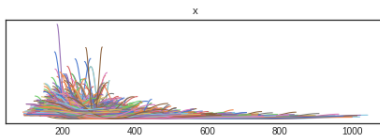
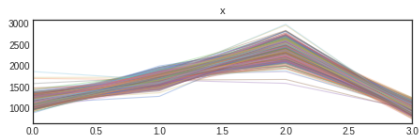
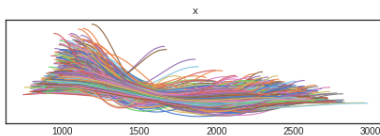
# Modelo de comparacion de grupos



Delitos en registrados en las alcaldías en 2020										
Alcaldía	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Sept	Octubre
ALVARO OBREGON	1324	1371	1380	906	871	1002	1121	1202	1131	1097
GUSTAVO A MADERO	1862	2083	1981	1356	1415	1542	1728	1713	1706	1593
IZTAPALAPA	2852	2953	3106	1935	1926	2133	2426	2526	2362	2349
TLALPAN	1160	1176	1207	838	721	910	992	1013	941	888

- ▶ La interpretación de la gráfica de violín parece complicada pero en realidad no lo es, básicamente estamos tratando con la forma de distribución de los incidentes en cada alcaldía seleccionada, y podría compararse con un diagrama de caja.
- ▶ Ahora, volvemos a implementar un modelo gaussiano para comparar las 4 alcaldías, pero con la diferencia de que en esta ocasión en lugar de calcular un sólo valor para  $\mu$  y  $\sigma$ , como en el primer modelo, tenemos que calcular los mismos valores para 4 grupos con sus respectivas medias y desviaciones.

# Modelo de comparación de grupos





- ▶ Y a continuación se expone la tabla resumen final de las medias y desviaciones de los 4 diferentes tipos de grupo calculados mediante un modelo gaussiano vectorizado e iterado; devolvemos para la cadena los valores después de las primeras 100 iteraciones una vez pasado el periodo de calentamiento.
- ▶ Tomemos en cuenta la siguiente notación para interpretar los resultados siguientes y veamos el valor promedio de sus medias y desviaciones:

$$- \mu_0: \mu_{(ALVARO\ OBREGON)} = 1142,34$$

$$- \sigma_0 = 209,70$$

$$- \mu_1: \mu_{(GUSTAVOAMADERO)} = 1696,61$$

$$- \sigma_1 = 266,45$$

$$- \mu_2: \mu_{(IZTAPALAPA)} = 2403,66$$

$$- \sigma_2 = 466,86$$

$$- \mu_3: \mu_{(TLALPAN)} = 989,75$$

$$- \sigma_3 = 183,73$$



Una forma de calcular la diferencia que existe entre el tamaño de 2 grupos es el coeficiente D de Cohen:

$$d = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}}$$

Este coeficiente mide el efecto del tamaño de la diferencia de las medias con respecto a la desviación estándar de ambos grupos. En las gráficas siguientes calculamos precisamente este coeficiente de la media y desviación estimadas, así que podemos calcularlo como una distribución, y no sólo un simple valor, pues los valores obtenidos de la vectorización gaussiana nos permite precisamente esto. A continuación mostraremos la tabla de interpretación del coeficiente:

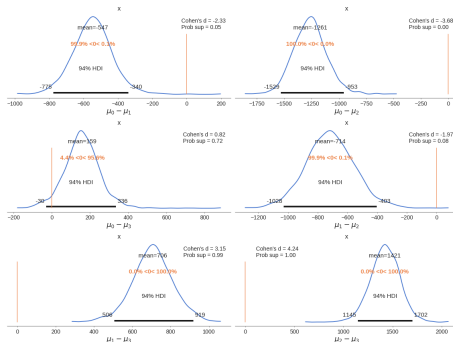


# Modelo de comparación de grupos



Rango Valor	Interpretación
$d \leq 0.2$	Pequeño Tamaño
$0.2 \leq d \leq 0.5$	Magnitud Media
$0.5 \leq d \leq 0.8$	Alta Magnitud

# Modelo de comparación de grupos

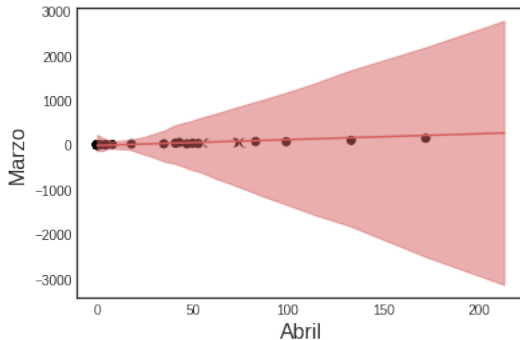


- Con lo anterior podemos concluir que existen alcaldías que en sus delitos tiene poca relación de impacto, pues estos no presentan la misma media. Ahora podemos afirmar que la alcaldía Iztapalapa es la más peligrosa por su relación del coeficiente d de Cohen

# Modelo de regresión logística múltiple para delitos



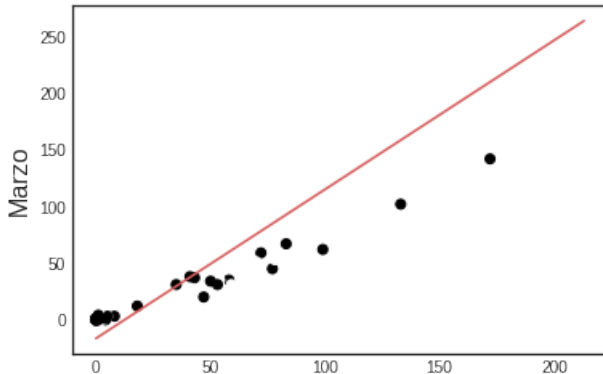
- Se realizó un regresión logística múltiple y posteriormente se realiza una comparación 2 diferentes meses (Marzo y Abril), se seleccionaron estos meses para poder analizar si hubo una repercusión significativa en el número de delitos en el inicio de la pandemia en México. Podemos observar que los crímenes no variaron mucho y se mantuvieron de igual forma en estos periodos.



# Modelo de regresión logística múltiple para delitos



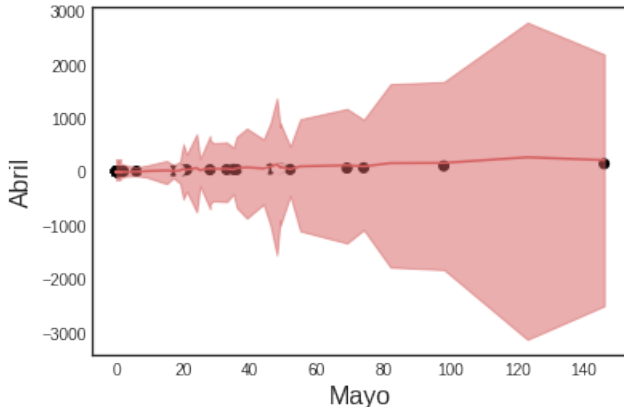
- Posteriormente se realizó una regresión logística para los meses Marzo Y Mayo, se puede observar que los crímenes se agudizaron más en Mayo (recordar que en Marzo inicio el confinamiento y en principio se respetaban las regulaciones para reducir la movilidad) por la relajación"de medidas del confinamiento.



# Modelo de regresión logística múltiple para delitos



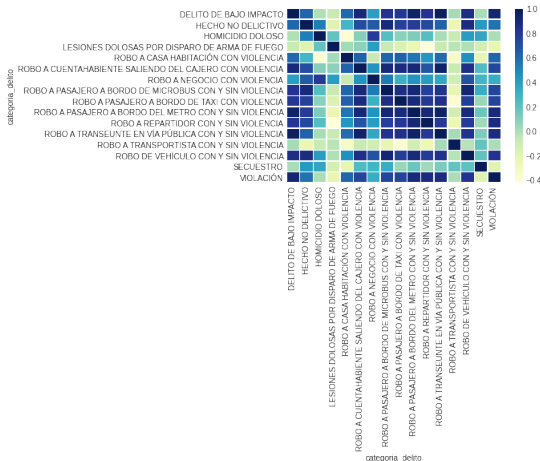
- Posteriormente se realizó una comparación de la regresión logística para Abril y Mayo, a diferencia de lo que pasa con Marzo y Abril, aquí la tendencia ya no es tan lineal y se empieza a ver en algunos puntos la tendencia que es observable en las bandas.





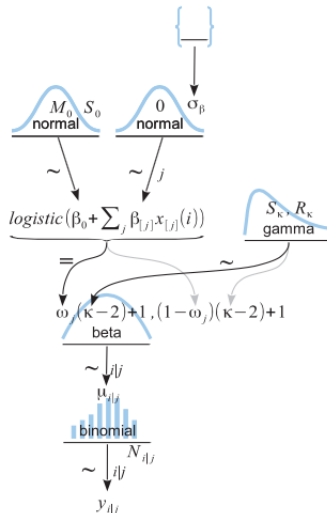
- ▶ Se realiza un mapa de calor para poder determinar la correlación que existe entre diferentes tipos de delitos en la Ciudad de México, observemos que hay algunos tipos de delitos que están altamente correlacionados y otros que no lo están.
- ▶ En general los delitos de robo tienen una correlación similar, no todos los tipos de delitos tienen la misma correlación por lo que no podemos trabajarlos como eventos independientes unos de otros.

# Correlación de las categorías de delitos



- De lo anterior, viene algo que ya nos esperábamos, algunos delitos están correlacionados y muy parecidos, como los robos que están muy correlacionados entre sí.

# Predicción nominal de valores para múltiples grupos



[27]L[.55mm].55



# Predicción nominal de valores para múltiples grupos



- ▶ ANOVA logística
- ▶ En algunas situaciones es necesario realizar análisis con variables dicotómicas, pero las predicciones son de tipo nominal, para poder realizar esto lo que podemos hacer es usar una adaptación de la regresión logística adaptado a un modelo ANOVA para variables dicotómicas.
- ▶ Iniciamos con un modelo ANOVA que genera una predicción para una tendencia para cada una de las condiciones, después ese modelo es pasado por una función sigmoide lo cual lo convierte la predicción nominal a uno de tipo  $y = 1$  para  $k$  valores de diferentes grupos de manera simultánea como lo muestra el diagrama del modelo.

# Predicción nominal de valores para múltiples grupos



- Desarrollando la idea a más detalle, para los valores nominales  $M_0$  se pueden crear comparaciones para diferentes parámetros  $\beta$  en un modelo ANOVA. Estos  $\beta$  parámetros no son probabilidades, sin embargo pueden ser mapeados en una medida de probabilidad usando la regresión logística del tipo:

$$\text{Logistic}(\beta_0 + \sigma_j \beta_{[j]} x_{[j]}(i))$$

# Predicción nominal de valores para múltiples grupos



- Que después, por medio de otra aplicación con una distribución gamma, podemos agrupar la información en  $K$  grupos:

$$w_j(k - 2) + 1$$

$$(1 - w_j)(k - 2) + 1$$

y con esto trabajar con la distribución binomial para aproximar los valores  $y_{ij}$  de una predicción nominal para variables dicotómicas conocida como un modelo ‘ANOVA Logístico’

# Predicción nominal de valores para múltiples grupos



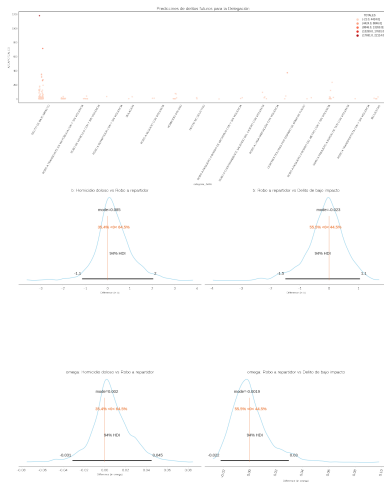
- ▶ En pocas palabras este es un modelo de clasificación predictivo. Cosa que de manera independiente no podríamos hacer al trabajar con variables dicotómicas por separado.
- ▶ En nuestro caso particular nos gustaría clasificar y ver la distribución de los delitos que hubo por cada categoría en la Ciudad de México en el 2020 y con base en esto comparar la proporción de tipos de delitos.
- ▶ Se muestra una descripción gráfica del ajuste de los modelos que se están realizando

# Predicción nominal de valores para múltiples grupos



- Una vez implementado y ejecutado el modelo, como podemos observar en las gráficas, no sólo se logró clasificar cada uno de los diferentes tipos de delitos, sino cuantificar intervalos para los delitos totales registrados en la Ciudad de México.
- Como era esperado, los delitos de bajo impacto siguen siendo los predominantes con aumentos, pero también hay otros tipos de delitos que parece que tenderán una alza, como los robos a transeúntes o los robos a vehículos.

# Predicción nominal de valores para múltiples grupos



# Predicción nominal de valores para múltiples grupos



Por último, se realizó una observación de como es que se están desarrollando algunos cambios en contraste de las  $b \sim \text{Deterministic}$  para diferentes tipos de categorías, donde recordemos que:

$$b = m - b_0$$

$$b = (a_0 + a) - b_0$$

donde:

$$a_0 \sim \text{Normal}$$

$$a \sim \text{Normal}$$

$$b_0 \sim \frac{1}{n} \sum m$$

# Predicción nominal de valores para múltiples grupos



- ▶ En donde, entonces, de acuerdo al parámetro el contraste de tipos de categorías es muy similar.
- ▶ El valor *mode* es el valor óptimo del rango de posibles valores para nuestra distribución  $M_{i|j}$
- ▶ Observemos, por ejemplo, que la relación de los delitos de bajo impacto con los robos a repartidor sí es más alta.



# Mapa de homicidios en la CDMX



En esta última sección se muestra geográficamente los homicidios que han ocurrido en la ciudad, aprovechando que tenemos la información de estos.

El siguiente mapa presenta la siguiente información:

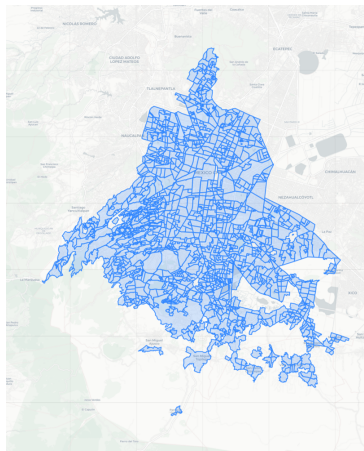
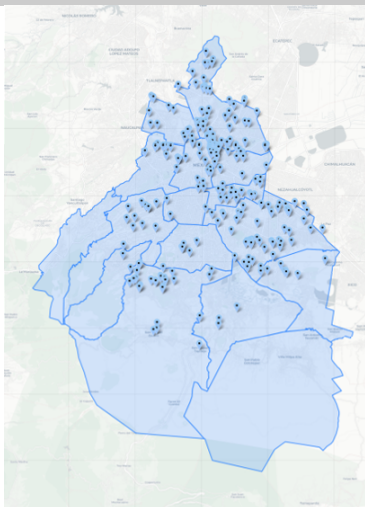
- ▶ - Con puntos se grafican de forma específica cuándo y dónde sucedió un homicidio.
- ▶ - Se muestra la delimitación delegacional.

# Mapa de homicidios en la CDMX



Como comentamos al inicio de este reporte escrito, uno de los crímenes que más han creció en México ha sido la violencia y en especial los homicidios en la capital del país, como comentamos, la siguiente gráfica muestra todos los homicidios que logramos registrar en la Ciudad de México a lo largo de 2020 y como podemos observar en la imagen, los delitos son muy numerosos y variados a lo largo del tiempo, pero principalmente la mayoría son del mes de enero y febrero y disminuyen estrepitosamente conforme avanza el tiempo, pero esta tendencia no duro mucho pues para junio y julio cuando las restricciones y las medidas de confinamiento se relajaron, los crímenes en la ciudad se volvieron a disparar. lo cual es un problema, pues los crímenes siguen en aumento.

# Mapa de homicidios en la CDMX



En este segundo mapa se muestran las medidas de seguridad que se han implementado por alcaldía en la CDMX.



- ▶ La incidencia delictiva al medio día es más recurrente que a altas hora de la noche, un hecho que resulta contraintuitivo de acuerdo a la idea colectiva
- ▶ Los crímenes disminuyeron en el mes de marzo a causa del confinamiento pero no fue duradero, pues hubo un incremento a puntos muy parecidos con meses antes de pandemia. Particularmente desde Agosto.
- ▶ Se encontró un gran caso de reportes de violencia familiar, cosa que va muy relacionada con el hecho de la medida de confinamiento que se tomó en 2020.
- ▶ Debido al confinamiento hubo pocos casos de delitos como: robo a casa habitación o robo de objetos a vehículos.
- ▶ El estar en un vehículo propio no representa un factor importante para no sufrir un hecho delictivo, es casi tan probable como sufrir uno siendo un transeunte.
- ▶ La pandemia no fue un caso determinante para la disminución de delitos después del mes de Marzo. Pero no fue significativo para mantenerlos en un rango bajo.