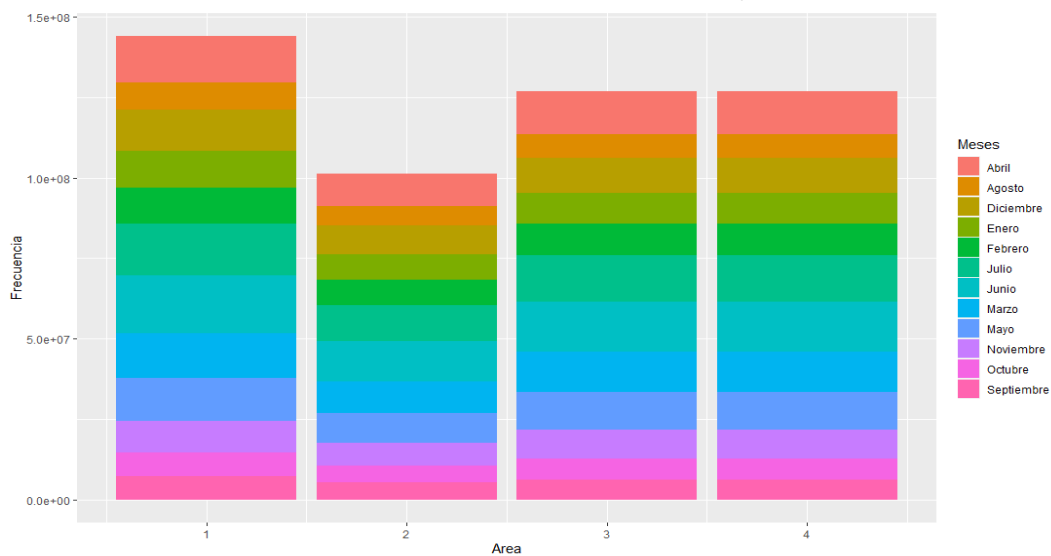


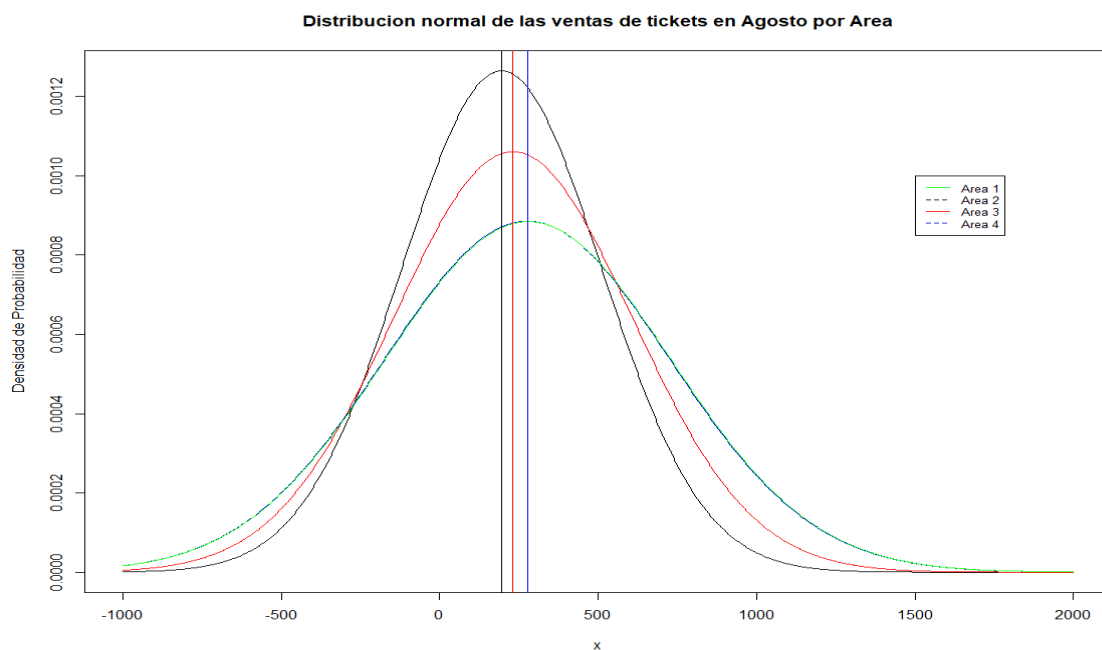


**Proyecto final - Inferencia estadística**  
**Análisis de datos para Cine**  
Por : Esteban Cárdenas 179150

## Frecuencia de venta por área y meses



La grafica de frecuencia de ventas por área y meses nos muestra que el área 3 y 4 tiene comportamientos similares como en frecuencia y cantidad de ventas por mes, por el otro lado, el area 1 se muestra como el área con mayor frecuencia de venta de tickets y el área dos con el menor numero de venta de tickets. Asimismo, con este mapa de calor podemos distinguir los periodos donde la gente compra mas boletos que sería entre febrero y Mayo.



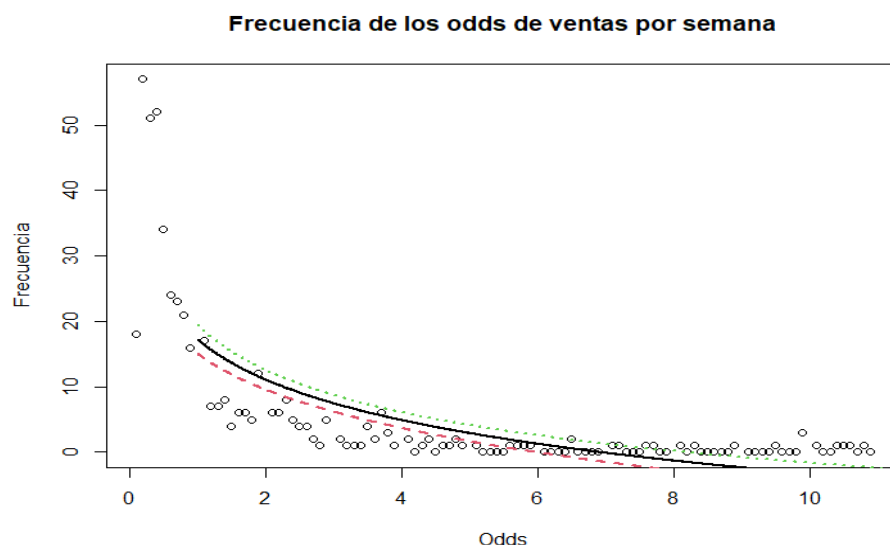
Se pregunta si el comportamiento de las ventas en agosto es similares. Analizando las distribuciones normales de la venta de boletos por área podemos observar que las medias no están tan separadas entre si por lo que hay menos error. La variable que si cambia es la desviación estándar lo cual nos dice que la dispersión entre las áreas es distinto. Si tomamos

estos dos criterios las áreas que se parecen mas en las ventas de agosto serian el area 1 y el área 4 ya que presenta muy poca diferencia entre ambos parámetros.

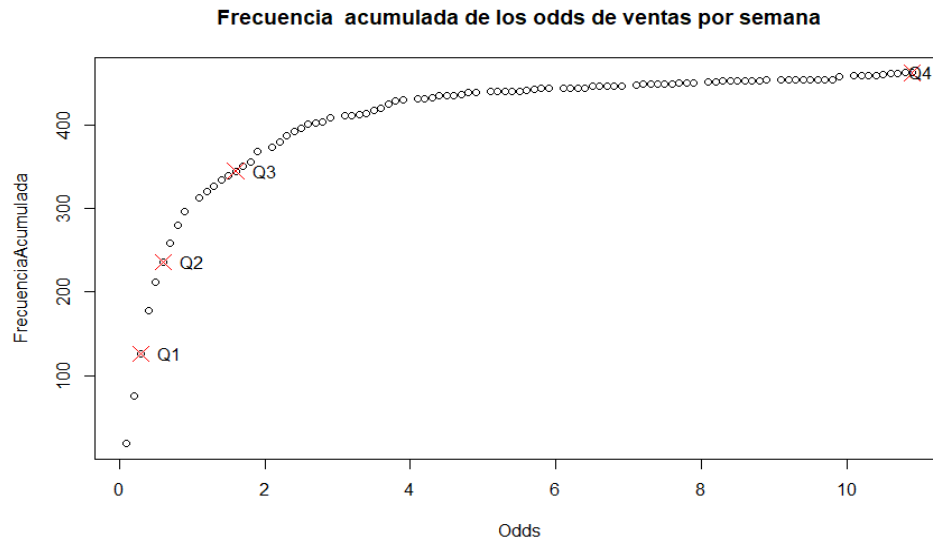
El criterio de ventas que se quería analizar para este mes era la probabilidad de que ese mes fuera un mes outstanding por lo que necesitaría superar los 4.5 millones de dólares. Con esto se analizaron todos los meses de agosto que nos proporcionaba la base de datos que es desde 2001-2031, el problema es que para ciertos años el área no contaba con datos sobre ese mes por lo que seria una frecuencia de 0 . Esto se podría arreglar con algún método de Bootstrap pero para analizar los datos de manera mas pura se saco la media y la desviación standard con estos datos y se calculó la probabilidad de que el mes supere una venta de 500000 boletos.

Área	Media	Desviación std	P(X>500,000)
Área 1	278884.9	451033.1	0.3941464
Área 2	195604.3	315446.9	0.2585067
Área 3	231307.6	376070.9	0.3268726
Área 4	276230.1	450468.6	0.3917510

Posterior a esto se calculo el momio de odds para cada semana de venta de boletos y se hizo un análisis sobre la distribución de este parámetro.



Esta frecuencia nos indica que la distribución de la muestra es exponencial por lo que entre mayor sea el odd de ocurrencia, la frecuencia de este será menor, y mientras menor sea el odd tendremos una mayor frecuencia.



Aquí tenemos la frecuencia acumulada de dicha muestra, y como podemos ver Q1, Q2 y Q3 están muy cerca entre si por lo que la mayor probabilidad de ocurrencia están entre los odds 0 y 2. Con esto deseamos sacar un máximo estimador de verosimilitud el cual nos permitirá analizar el mayor número de datos posibles:

Sea  $x_1, \dots, x_n$  una m.a de la dist.  $\text{Exp}(\theta)$

La función de verosimilitud es

$$\begin{aligned} L(\theta) &= f(x_1; \theta) \dots f(x_n; \theta) \\ &= \theta \cdot e^{-\theta x_1} \dots \theta \cdot e^{-\theta x_n} \\ &= \theta^n \cdot e^{-\theta n\bar{x}} \end{aligned}$$

Observación:  $L(\theta)$  es máxima en el mismo punto en donde  $\ln L(\theta)$  lo es

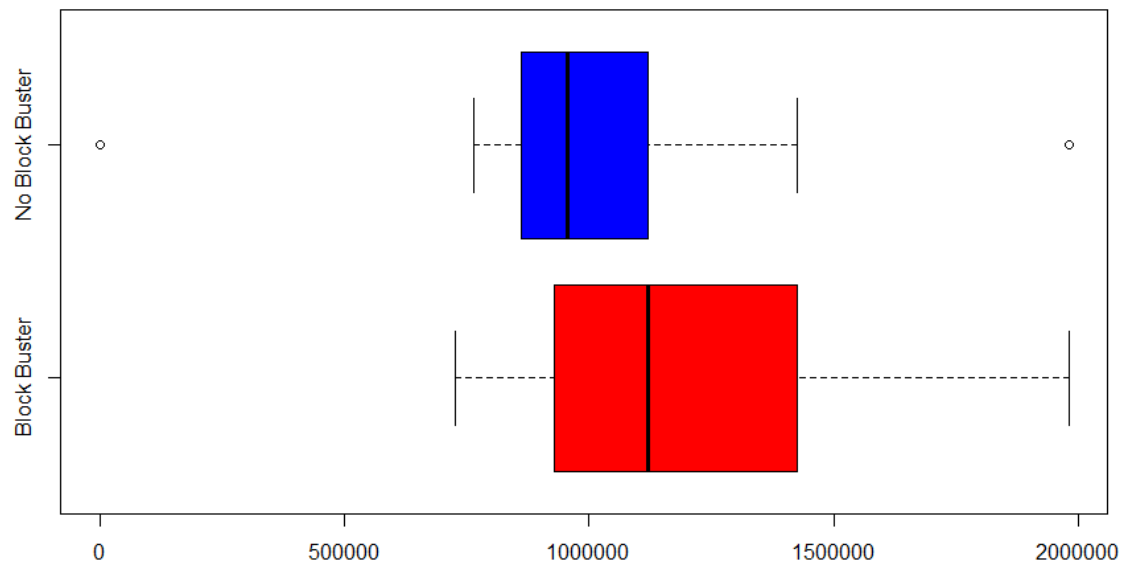
$$\begin{aligned} \Rightarrow \ln L(\theta) &= n \ln \theta - \theta n\bar{x} \\ \therefore \frac{d}{d\theta} \ln L(\theta) &= \frac{n}{\theta} - n\bar{x} \\ \frac{d}{d\theta} &= 0 \Leftrightarrow \theta = \frac{1}{\bar{x}} \end{aligned}$$

$$\Rightarrow \frac{d^2}{d\theta^2} \ln L(\theta) = -\frac{n}{\theta^2} < 0$$

$\therefore \hat{\theta} = \frac{1}{\bar{x}}$  es la estimación para  $\theta$

$\therefore \hat{\theta} = \frac{1}{\bar{x}}$  es el estimador para  $\theta$

**Diagrama caja brazos de ventas de boletos entre películas blockbuster y no blockbuster**

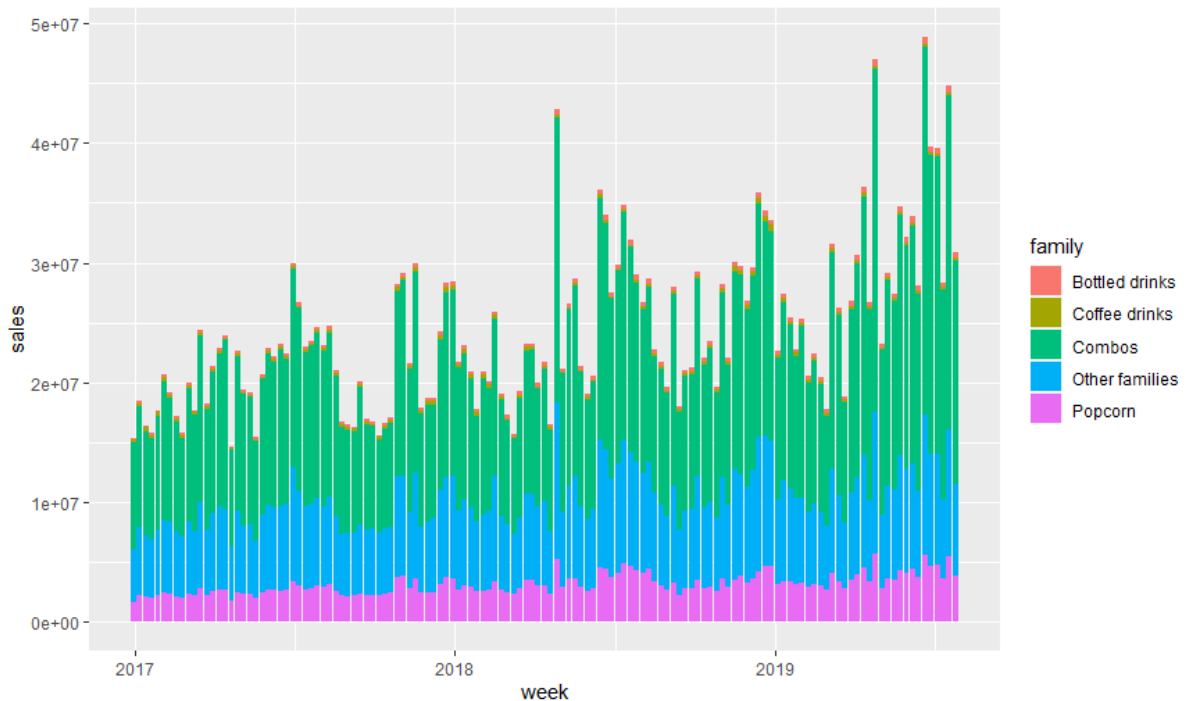


El diagrama caja brazos nos muestra la frecuencia de compra de boletos cuando hay un block buster en cartelera y cuando no hay. Dicho esto, el diagrama caja brazos nos indica que efectivamente hay un mayor número de venta de boletos cuando hay un block buster pero esta tiene una mayor desviación estándar que cuando no hay una función Block Buster en cartelera.

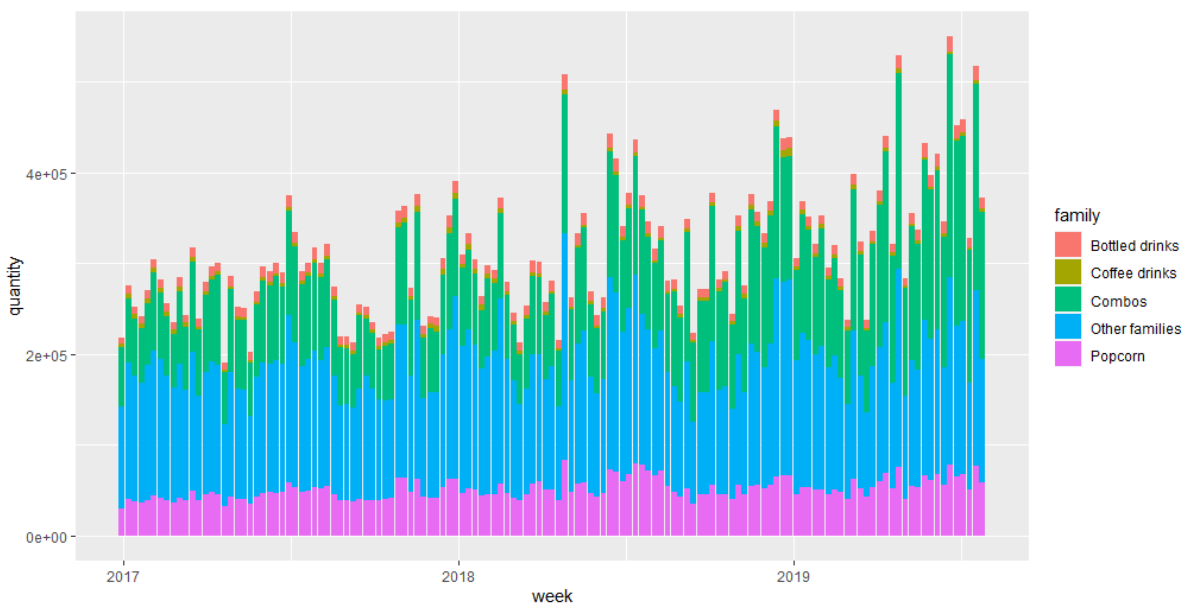
Para comprobar esto se sacó un intervalo de confianza del 80 % para ambas categorías de venta de boletos, esta comparación nos indica que el intervalo de la venta de boletos con Blockbuster es mayor que cuando no hay, y por lo tanto existe una mayor venta de boletos.

Intervalos de confianza de 80% de las ventas de boletos con y sin BlockBuster		
	-Z	Z
Block Buster	1,166,658.7 \$	1,343,093 \$
no Block Buster	999,755.3 \$	1,058,483 \$

Así mismo se analizó las ventas de la dulcería a lo largo del tiempo, en esta primera gráfica tenemos el total de ventas por semana dividido en el tipo de venta que fue. Con este análisis podemos darnos cuenta que el mayor ingreso de la dulcería deriva de los combos, seguido de "other families" y palomitas. Así mismo podemos ver como a lo largo del tiempo, las ventas en dulcería han aumentado.



Otra variable que se pudo analizar fue el de la cantidad de artículos vendidos por categoría, el cual nos muestra un interesante análisis ya que aun que “other families” no representa el mayor número de ingresos si representa el mayor numero de artículos vendidos. Las palomitas así mismo cuentan con un similar numero de ventas que los combos aun que los combos representen el mayor número de ingresos. Este análisis serviría mas en un entorno de administración de inventario.



Se analizo la decisión de un ejecutivo el cual hizo cambio en los precio de la dulcería, para comprobar si esta decisión trajo mas ingresos o no se hizo un intervalo de confianza sobre la frecuencia de una muestra de 30 datos antes y después de que implementara el cambio,

Por lo que se pudo analizar si hubo un cambio en los ingresos pero mas que nada la cota inferior fue la que cambio en mayor proporción.

Intervalos de confianza de 80% de las ventas de dulceria antes y después de 2017-09-30		
	-Z	Z
Antes	3,739,501 \$	5,094,729 \$
Después	5,252,779 \$	6,608,007 \$

Al poder tomar todas estas pautas de análisis podemos darnos cuenta que la extracción y manipulación de los datos históricos de una empresa nos pueden ayudar a entender y modelar el comportamiento de los clientes, y de esta forma poder maximizar nuestras utilidades así como tener un mejor manejo y gestión del negocio. Con estos datos podemos anticipar ciertos eventos que pudimos inferir gracias a la estadística. Con esto es importante recalcar el uso de la estadística en los negocios ya que con ella podemos ver cosas que antes no estaban a la simple vista.