

Tarea 1: Food Trucks en CDMX

Inferencia Causal

Mateo De La Roche Macías

El objetivo de este reporte presentar un análisis sobre la industria de los food truck en la Ciudad de México. Todos los análisis se realizaron con el censo *foodtruck* el cual contiene todos los food trucks registrados ante la COFEPRIS. Los datos que ocupamos en este análisis nos fueron proporcionados por *CDMX a la Carta*, de una encuesta que levantaron. La información que contiene la encuesta proviene de los foodtrucks con 20% más reviews.

Nuestros datos se ven de la siguiente manera

```
# A tibble: 6 x 9
  ...1 Number_of_Customers Menu_Price Marketing_Spend Cuisine_Type
  <dbl>          <dbl>      <dbl>      <dbl> <chr>
1     1           95       45.5       11.2 Pasta
2     2           79       32.9       16.0 Hamburger
3     3           95       45.4       19.3 Sushi
4     4           48       19.2        3.91 Tacos
5     5           30       35.3        9.33 Tacos
6     6           94       47.6        4.15 Hamburger
# i 4 more variables: Average_Customer_Spending <dbl>, Promotions <dbl>,
#   Reviews <dbl>, Monthly_Revenue <dbl>
```

Algunas estadísticas descriptivas

...	1	Number_of_Customers	Menu_Price	Marketing_Spend
Min.	: 1.00	Min. :10.00	Min. :10.05	Min. : 0.0376
1st Qu.:	50.75	1st Qu.:30.75	1st Qu.:22.30	1st Qu.: 4.2822
Median	:100.50	Median :53.00	Median :32.17	Median :10.2429
Mean	:100.50	Mean :53.84	Mean :31.59	Mean : 9.9261
3rd Qu.:	150.25	3rd Qu.:78.00	3rd Qu.:41.02	3rd Qu.:15.2547
Max.	:200.00	Max. :99.00	Max. :49.97	Max. :19.9225
Cuisine_Type		Average_Customer_Spending	Promotions	Reviews

Length:200	Min. :10.07	Min. :0.00	Min. :81.00
Class :character	1st Qu.:22.01	1st Qu.:0.00	1st Qu.:85.00
Mode :character	Median :30.61	Median :0.00	Median :90.00
	Mean :30.74	Mean :0.46	Mean :90.11
	3rd Qu.:40.07	3rd Qu.:1.00	3rd Qu.:95.00
	Max. :49.82	Max. :1.00	Max. :99.00

Monthly_Revenue

Min. :-28.98

1st Qu.:205.00

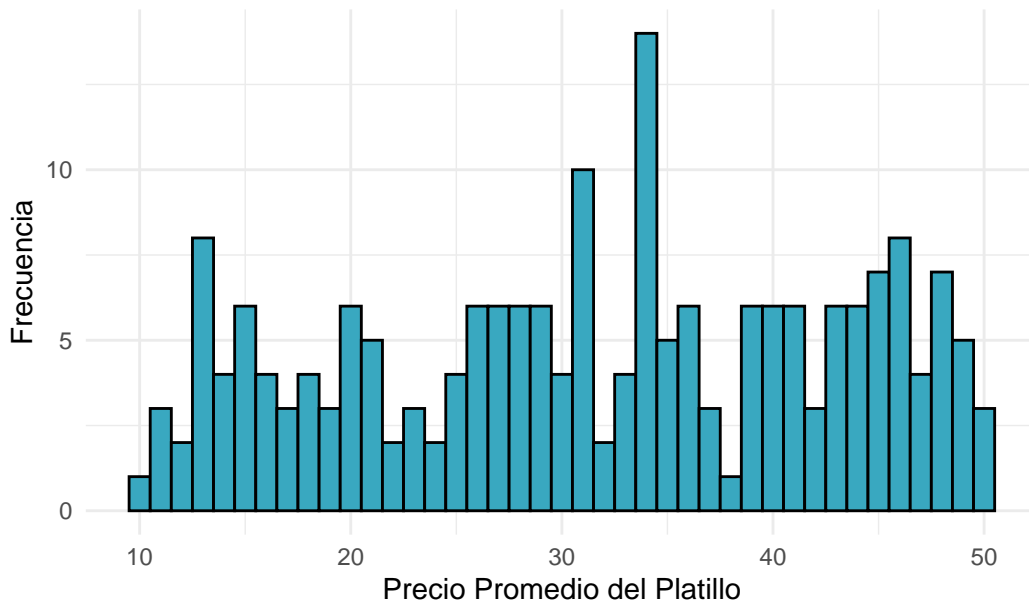
Median :274.03

Mean :270.72

3rd Qu.:337.47

Max. :522.77

Histograma de Precio Promedio del Platillo



1. Intervalo de confianza del 95% de la media de *Menu_Price*

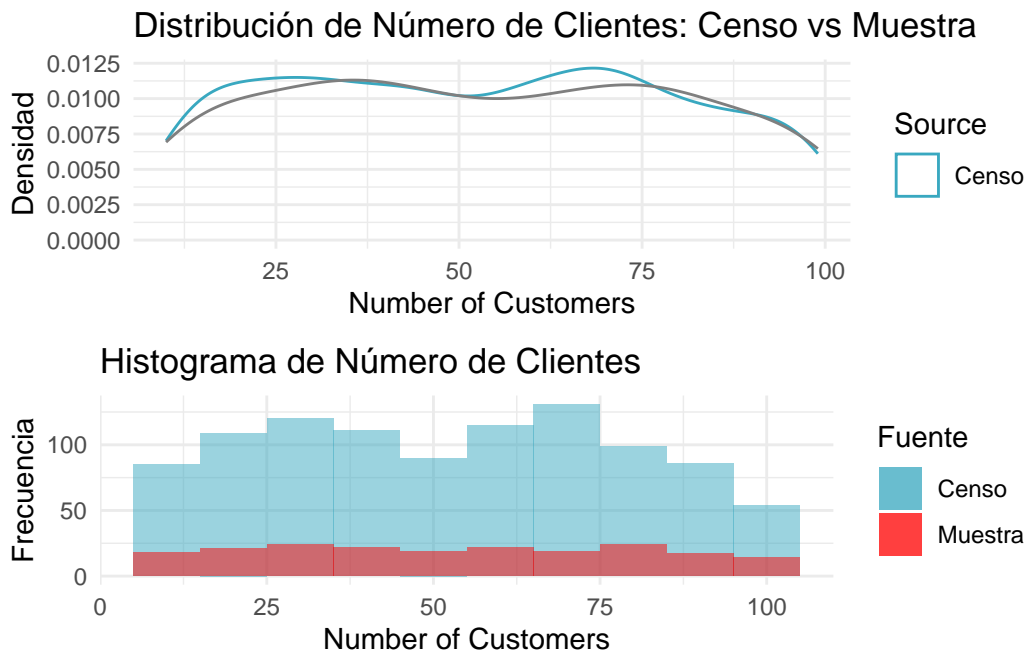
Podemos afirmar con un 95% de confianza que el precio promedio del platillo se encuentra entre \$ 30.00795 y \$ 33.17191 MXM

2. Representatividad

Debido a las sospechas acerca de la representatividad de los datos ya que estos vienen de la revista *CDMX a la Carta* y la forma en la que los recuperaron fue basandose únicamente en el 20% de los food trucks con más reviews. Es muy probable que la muestra se encuentre sesgada ya que posiblemente los food trucks con más reviews sean también los que tienen mayor demanda y esto podría hacer que sus precios sean mayores a los de los food trucks con menos demanda. También se podría dar el caso que los food truck con más cantidad de reviews sean aquellos más baratos y esto sesgaría a nuestra variable de interés para abajo. Sea cual sea el caso, podemos afirmar que nuestra muestra no es representativa de la población.

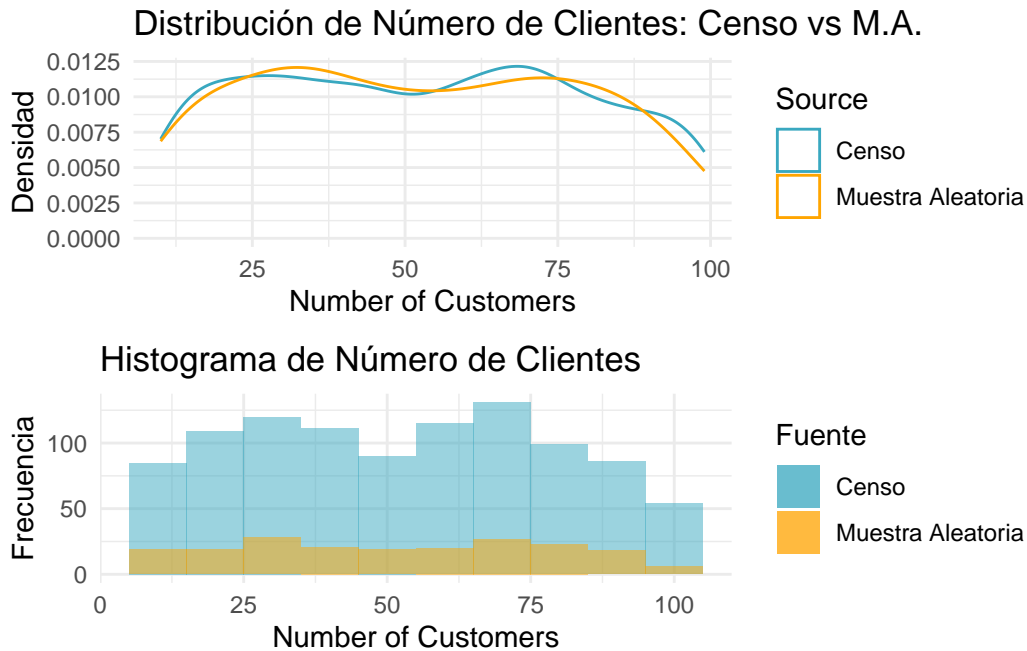
Para mostrar que no es representativa, podemos hacer uso del censo *foodtruck* el cual contiene información de 1000 foodtrucks. Sin embargo, en el censo solo tenemos acceso al número de clientes, tipo de cocina y al ingreso mensual. Debido a que no tenemos a nuestra variable objetivo en el censo, podemos hacer uso de las preguntas ancla. Nos apoyaremos de la variable número de clientes.

Podemos observar el histograma y la gráfica de distribución de la variable.



Podemos ver claramente como no es la misma distribución, los datos del censo y la muestra tienen diferentes histogramas.

Para poder realizar una muestra representativa y sin sesgos se debió haber realizado un muestreo aleatorio. Para demostrar la diferencia, abajo se muestra la gráfica de distribución y el histograma de los datos obtenidos con muestreo aleatorio del censo *foodtruck*.

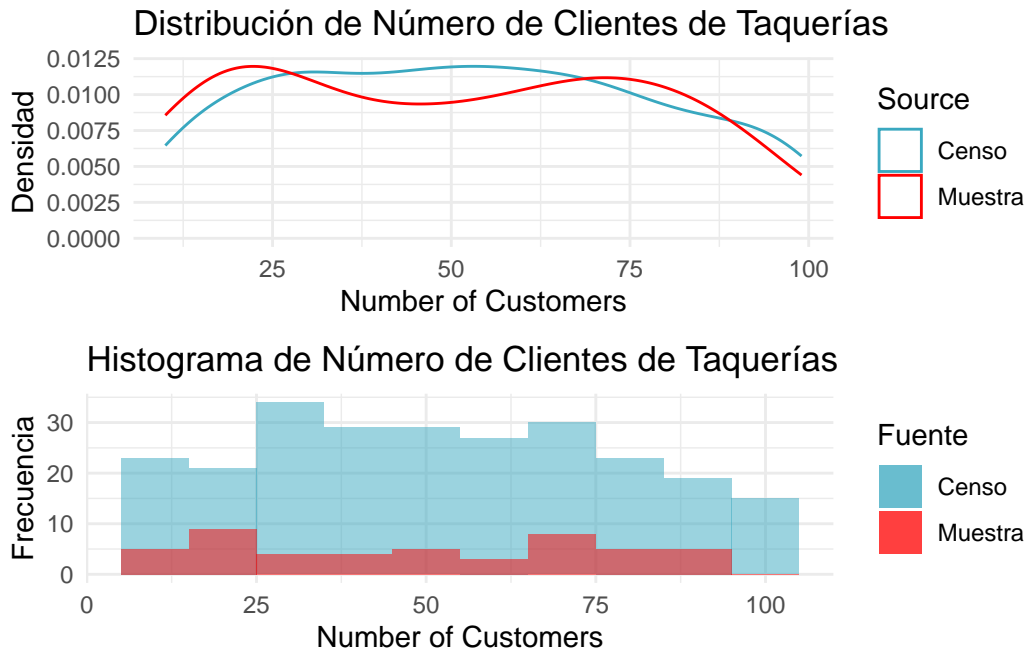


A diferencia de los datos obtenidos con la muestra de *CDMX a la Carta*, podemos ver como con la muestra aleatoria, muestra y datos tienen la misma distribución.

Mientras realizábamos el análisis, hubo un analista creativo al que durante el reporte nos referiremos como C. C decidió tomar una muestra eligiendo 20 food trucks de cada tipo de comida, para sugerir una muestra. A pesar que su intención no era mala, el realizar una muestra de esta forma introduce un sesgo de selección a los datos ya que está forzando a que los food trucks de todos los alimentos tengan el mismo número de observaciones lo cual no es representativo de la realidad, cabe agregar que su estrategia no es un muestreo aleatorio.

Sin embargo, si tuviéramos que trabajar con la muestra de C, es posible encontrar una solución, ya que en la muestra de C se tiene la misma cantidad de cada food truck por comida y en el censo no, lo que se puede hacer es ver la proporción de cada food truck por tipo de comida en el censo y asignarle una ponderación a ese tipo de comida según su proporción en el censo.

Volviendo a nuestro análisis inicial, si solo nos interesa evaluar el precio promedio por platillo de un tipo de comida en específico, por ejemplo en las taquerías, lo que tendríamos que hacer es filtrar los datos del censo por taquería y ver si nuestra muestra de taquerías es representativa.



Podemos ver que las taquerías tampoco son una muestra representativa, si deseáramos sacar una muestra representativa de las taquerías, deberíamos de realizar una muestra aleatoria de todas las taquerías.

3. Pruebas de Hipótesis.

Debido al particular interés del gobierno en los food trucks de tipo taquería, realizaremos un análisis detallado de la diferencia entre la media del precio promedio por platillo de las taquerías menos la media del precio promedio de los demás tipos de cocina. Para esta sección asumiremos que la muestra que tenemos es representativa y basándonos en ese supuesto realizaremos nuestro análisis.

Por ahora nos interesa conocer si la media de precio de las taquerías es igual que la media de precio de los otros tipos de comida.

Planteamos la siguiente prueba de hipótesis.

Hipótesis Nula (H_0): $\mu_{tacos} - \mu_{otros} = 0$ Hipótesis Alternativa (H_1): $\mu_{tacos} - \mu_{otros} \neq 0$

Nos apoyaremos del siguiente estadístico de prueba t.

$$t = \frac{\bar{X}_{tacos} - \bar{X}_{otros}}{\sqrt{\frac{S_{tacos}^2}{n_{tacos}} + \frac{S_{otros}^2}{n_{otros}}}}$$

donde:

- \bar{X}_{otros} y \bar{X}_{tacos} son los estimadores de las medias respectivas.
- S^2_{otros} y S^2_{tacos} son los estimadores de las varianzas de los grupos.
- n_{otros} y n_{tacos} son los tamaños de las muestras.

La prueba de hipótesis nos da un P-value de 0.1029526 por lo que no tenemos evidencia suficiente para rechazar la hipótesis nula. Es decir, no podemos decir que la media de precios de los tacos sea diferente a la media de los precios de los otros foodtrucks.

Nota Ejecutiva para el gobierno:

En el reciente estudio estadístico realizado por nuestra consultoría, nos enfocamos en estudiar, analizar y comparar los precios promedio ofrecidos por los food trucks especializados en tacos con aquellos especializados en otros tipos de cocina. El objetivo era identificar si existía alguna diferencia significativa en los costos de las taquerías con respecto a otros tipos de comida, ya que esto podría impactar al mercado.

Cuando realizamos nuestro análisis estadístico, no encontramos evidencia que sugiera diferencias significativas en los precios promedios entre los food trucks de tacos y sus contrapartes que ofrecen otras variedades de comida. Esto implica que si nos basamos únicamente en nuestros datos no podemos concluir que los tacos son o más caros o ni más baratos que el resto de opciones de comida ofrecida en los food trucks.

4. Bootstrap

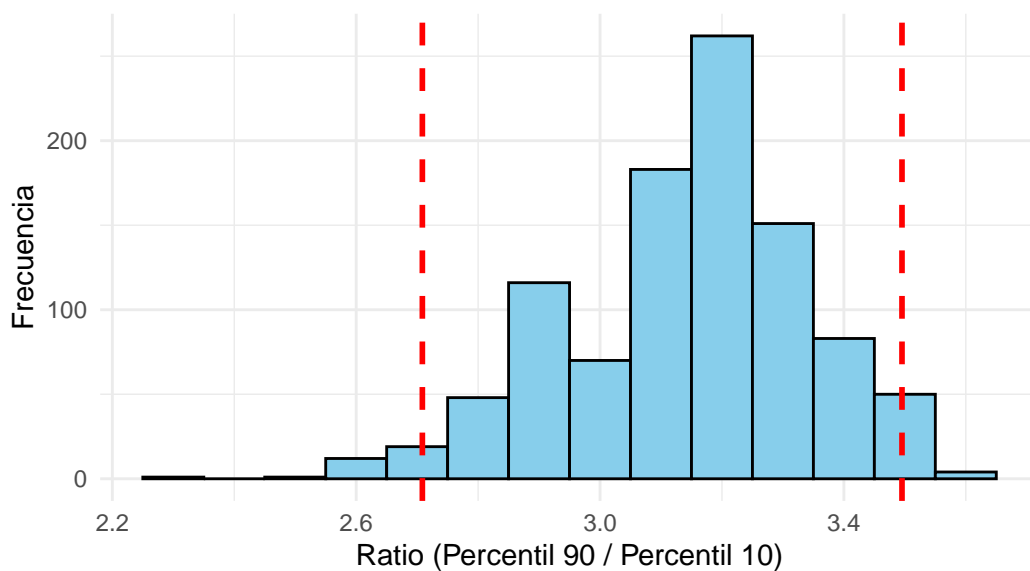
En el mundo de la cocina urbana existe una amplia gama de experiencias gastronómicas. En un extremo se encuentran los food trucks gourmet y en otro los food trucks que atienden el paladar cotidiano. Debido al interés del gobierno en la diferencia de precios vamos a realizar un análisis de la desigualdad de los precios. Para esto utilizaremos el estadístico de interés siguiente.

$$ratio = \frac{\text{percentil } 90}{\text{percentil } 10}$$

Bootstrap con muestreo de 200 observaciones

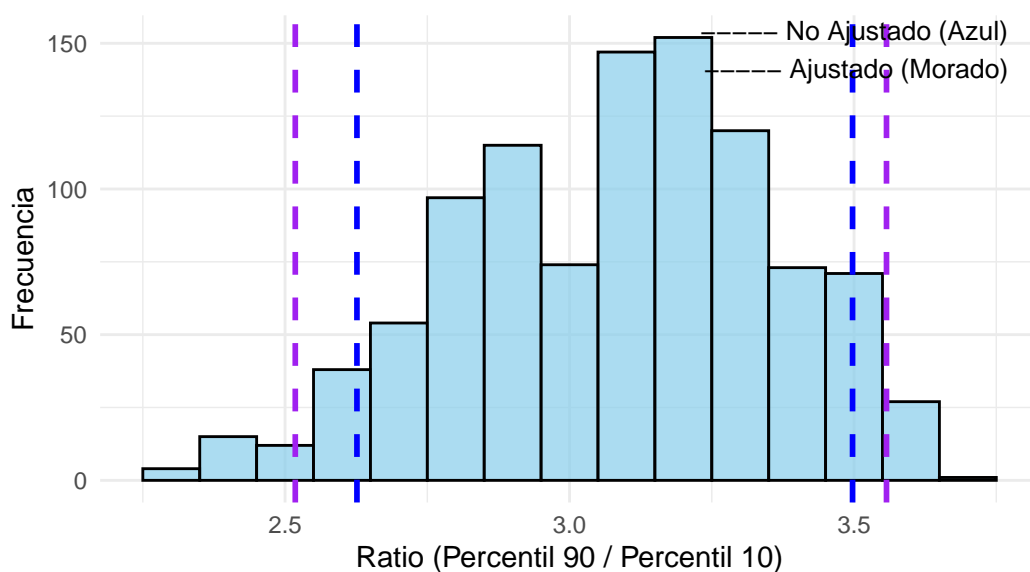
Estimación puntual del ratio: 3.145229 con un intervalo de confianza del 95% entre 2.708524 3.49534 para el ratio con muestras de 200 obs.

Histograma de Ratios de Bootstrap (muestras de 200)
con Intervalo de Confianza del 95%



Bootstrap con muestreo de 100 observaciones

Histograma de Ratios de Bootstrap (muestras de 100)
con Intervalo de Confianza del 90%



Estimación puntual del ratio: 3.083906 con un intervalo de confianza del 95% entre 2.51794 3.557032 para el ratio con muestras de 100 obs.