

Universidad EAFIT

DEPARTAMENTO DE CIENCIAS MATEMÁTICAS

PRIMER TRABAJO

Estadística General

Simón Marín Giraldo
David Calle González
Agosto 2020

Índice

1. Introducción	2
2. Descripción del problema	2
2.1. Variables de la encuesta	2
2.2. Objetivo	2
3. Análisis de datos	2
3.1. Carga del dataset	2
3.2. Hombres y mujeres	3
3.3. Grupos etarios	4
3.4. Medidas de tendencia central	5
4. Nuestras variables	6
4.1. Aumento anual teatro	7
4.2. GPSM	7
4.3. Porcentaje para el teatro	8
5. Conclusiones	9
5.1. Mujeres mayores (54-64)	9
5.2. Mujeres jóvenes (18-28)	10
6. Video	10

1. Introducción

La **estadística descriptiva** es la rama de la estadística que recolecta, analiza y caracteriza un conjunto de datos con el objetivo de **describir** las características y comportamientos de este conjunto mediante **medidas de resumen, tablas o gráficos**.

En el curso de **Estadística General** se propone, como parte de la evaluación de lo aprendido, realizar un **análisis** a un determinado problema haciendo uso de técnicas de **estadística descriptiva**, para identificar posibles estrategias para dar solución al problema, teniendo un entendimiento del negocio y sus objetivos.

2. Descripción del problema

Se tiene una base de datos `theatre.csv`, que fue una encuesta realizada a 699 personas que tienen la tarjeta de cliente frecuente del teatro Schauspielhaus Zürich.

2.1. Variables de la encuesta

Los participantes registraron la siguiente información en la base de datos:

- **Age**: Edad del encuestado.
- **Sex**: Género del encuestado.
- **Income**: Ingreso anual del encuestado (en miles de Francos Suizos).
- **Culture**: Gasto mensual en cultura (en Francos Suizos).
- **Theatre**: Gasto mensual en visitas al teatro (en Francos Suizos).
- **Theatre_ly**: Gasto mensual estimado en visitas al teatro en el año anterior (en Francos Suizos).

2.2. Objetivo

El objetivo es analizar los datos de la encuesta y determinar en qué **grupos poblacionales** se deberían enfocar las estrategias para incrementar los ingresos del teatro Schauspielhaus Zürich.

3. Análisis de datos

3.1. Carga del dataset

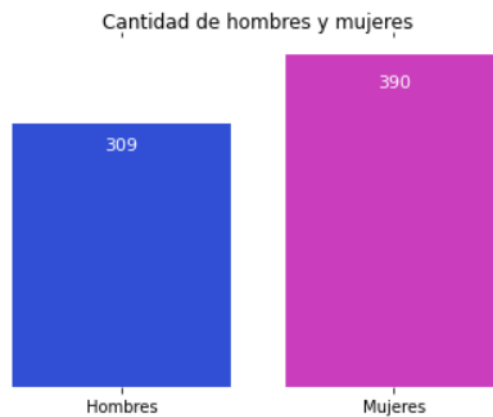
Inicialmente se cargó el **dataset** para poder visualizar la estructura de los datos.

	Age	Sex	Income	Culture	Theatre	Theatre_ly
Index						
1	31	1	90.5	181	104	150
2	54	0	73.0	234	116	140
3	56	1	74.3	289	276	125
4	36	1	73.6	185	75	130
5	24	1	109.0	191	172	140

Luego procedimos a hacer diferentes análisis a partir de gráficos y medidas estadísticas.

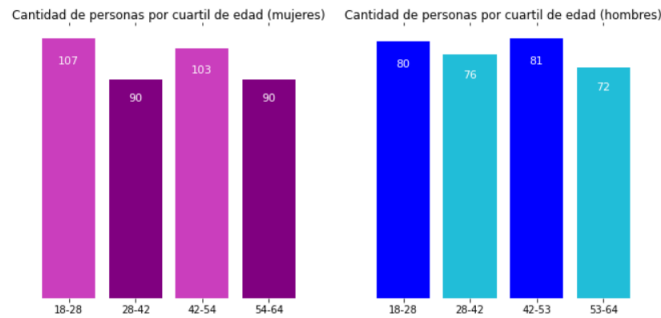
3.2. Hombres y mujeres

El primer paso fue analizar la cantidad de mujeres y hombres para hallar el sexo predominante a través de un simple diagrama de barras, en el cual, se puede evidenciar que la población de mujeres es mayor que la de hombres por **81** personas, lo cual es una diferencia significativa.

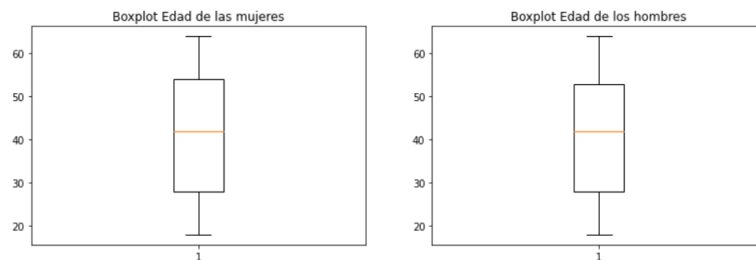


3.3. Grupos etarios

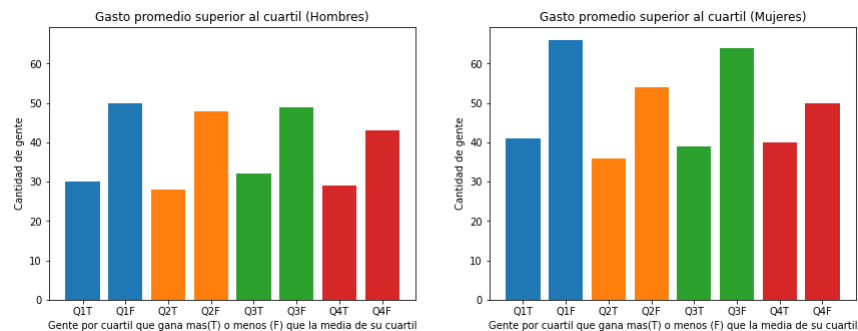
Posteriormente se definieron cuartiles para dividir a la población de cada género por rangos de edad.



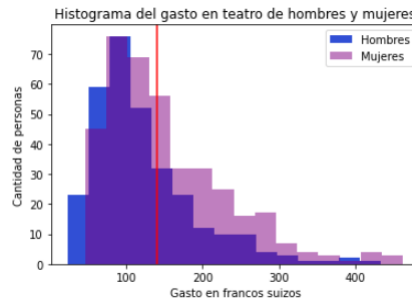
Esta visualización nos permite evidenciar que entre los cuartiles no hay mayores diferencias en la cantidad de personas. En cierta forma están balanceados. Aun así, se hicieron gráficos **boxplot** para poder analizar a mayor profundidad los cuartiles.



Con estos se pudo confirmar una distribución equitativa de los datos. Adicional a esto se utilizó esta información para generar una nueva variable cualitativa llamada **GPSC** (gasto promedio superior en cuartil), que indica si el gasto en teatro del individuo es superior al promedio del gasto en teatro del cuartil al que pertenece con verdadero o falso, y se ilustra en la siguiente gráfica:



Donde se evidencia también que, independientemente del cuartil o del género, el gasto de la mayoría de la gente es menor al del promedio, como se comprueba con este histograma:



Donde la línea roja indica el promedio.

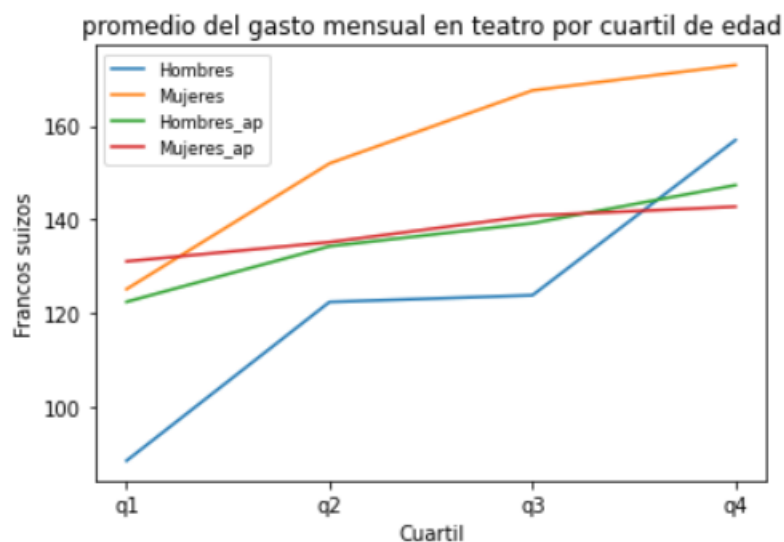
3.4. Medidas de tendencia central

Para analizar el gasto en teatro y cultura, el gasto en teatro del año pasado, y un poco más a profundidad la edad, hicimos uso de la media y la mediana en ambos grupos.

Hombres				
	Age	Culture	Theatre	Theatre_ly
mean	41.197411	222.711974	122.132686	135.598706
median	42.000000	219.000000	103.000000	130.000000
Mujeres				
	Age	Culture	Theatre	Theatre_ly
promedio	41.330769	217.592308	153.582051	137.333333
mediana	42.000000	214.000000	132.000000	130.000000

Con estos datos se aprecia claramente que las mujeres superan a los hombres en gastos en teatro este año, y que los hombres superan también, aunque por muy poco, a las mujeres en cuanto a los gastos en cultura. Pero la apreciación más importante es la diferencia del gasto en teatro con respecto al año anterior; ambos generos tenían gastos similares el año pasado, sin embargo, el gasto en teatro por parte de los hombres decrementó, mientras que el de las mujeres aumentó, lo cual es un factor importante a tener en cuenta para la decisión de hacia cuál género enfocar la campaña.

Para un análisis más completo e ilustrativo, se graficó el gasto en teatro tanto de hombres como mujeres por cuartiles en ambos años.



A través de este gráfico confirmamos la similitud del gasto en teatro por parte de hombres y mujeres en el año pasado (Marcados en el gráfico como *Hombres_{ap}* y *Mujeres_{ap}*) y la enorme brecha que se formó este año entre ambos generos. Y a la vez descubrimos nueva información, el gasto en teatro tiende a ser mayor en personas de mayor edad.

4. Nuestras variables

Creamos **3** nuevas variables para analizar mas a fondo la información del teatro las cuales se pueden observar en esta pequeña muestra:

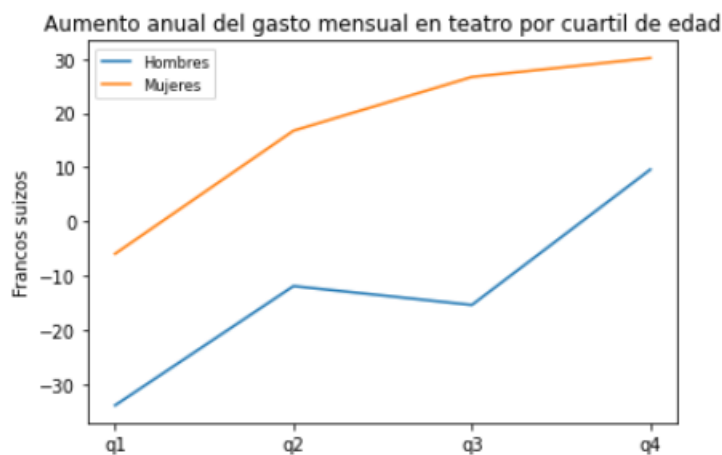
percentage theatre aumento anual teatro GPSPM

1.379006	-46	False
1.906849	-24	False
4.457604	151	True
1.222826	-55	False
1.893578	32	True

procederemos a explicar cada una.

4.1. Aumento anual teatro

Esta variable indica la diferencia entre el gasto mensual del año pasado y el de este año en teatro para cada individuo.



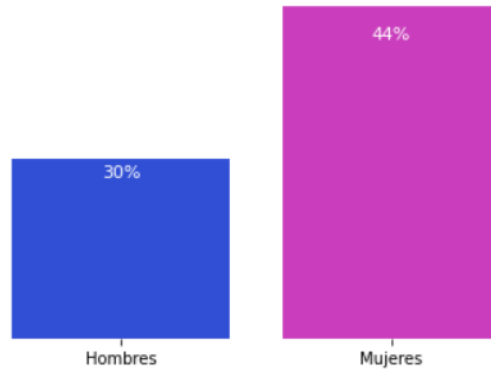
Graficamos la variable según el promedio de los cuartiles y para ambos géneros con lo que logramos enfatizar que las mujeres muestran un aumento considerablemente mayor que los hombres en el gasto mensual en teatro, y a pesar de que el primer cuartil de mujeres demuestra un decremento, no decremента tanto como los primeros 3 cuartiles de los hombres.

También descubrimos que las personas de mayor edad tienden a aumentar mas el gasto mensual en teatro anualmente.

4.2. GPSM

GPSM significa gasto promedio superior a la media, y es similar a la variable **GPSC** que vimos anteriormente, pero esta indica si la persona gasta más que el promedio general del gasto mensual en teatro, en vez del gasto promedio de cada cuartil.

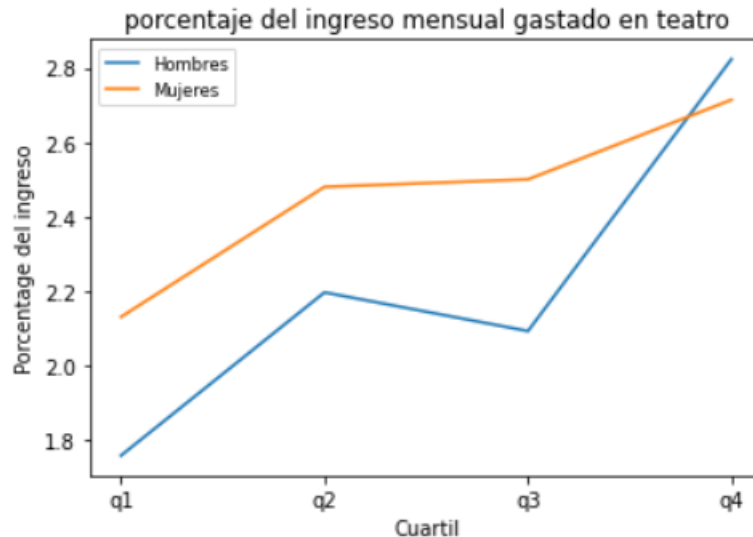
porcentaje de hombres y mujeres que gastan mas del promedio en teatro



Analizando el porcentaje de hombres y mujeres que gastan mas que la media se evidencia que aparte de que el gasto por parte de las mujeres en el teatro es mayor que el de los hombres, en cuanto a proporciones las mujeres también muestran una ventaja en gastar mas del promedio.

4.3. Porcentaje para el teatro

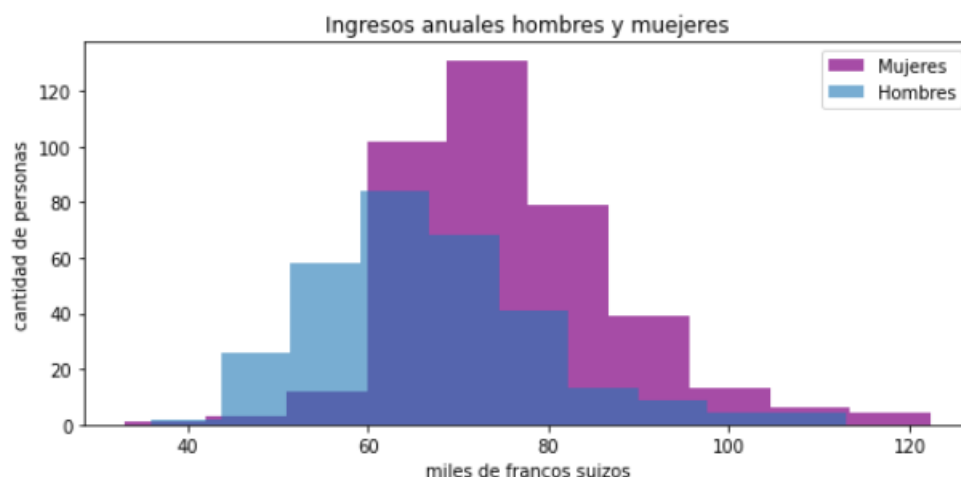
Esta variable (representada como `percentage theatre`) estaba planteada como ejemplo en el taller y nos pareció interesante analizarla también, e indica que porcentaje de los ingresos mensuales, los cuales se obtuvieron a partir de los ingresos anuales, son destinados por los individuos al teatro.



Esta variable también fue analizada tanto por promedio de cuartiles como por género y, como era de esperarse, las mujeres vuelven a ganar con un mayor

porcentaje de sus ingresos gastados en teatro, a excepción del cuarto cuartil donde los hombres sorprendentemente muestran una inversión mayor con respecto a sus salarios.

Un factor importante a tener en cuenta es que las mujeres en promedio ganan más que los hombres, teniendo las mujeres un salario anual promedio de 75.500 Francos Suizos al año, mientras que el de los hombres es de 66.900 Francos Suizos. Esta información se corrobora con el siguiente gráfico:



5. Conclusiones

Teniendo en cuenta que la cantidad de mujeres es mayor que la de hombres, que las mujeres muestran un mayor gasto en teatro tanto en cantidades como en proporciones y el hecho de que la mayoría de las mujeres muestran una tendencia a aumentar el gasto en teatro cada año, mientras que los hombres tienden a gastar menos cada año, en cuanto al género esta más que claro que la campaña debe de ser dirigida hacia las mujeres. Sin embargo, a nuestro parecer, hay **2** grupos etarios de mujeres hacia los cuales se puede dirigir la campaña dependiendo del enfoque.

5.1. Mujeres mayores (54-64)

Esta parte de la población es la que gasta mayores cantidades en el teatro y la que tiende a aumentar más los gastos cada año. La campaña sería para mantener su asistencia y promover que aumenten cada vez más el gasto al ser el que se podría decir el grupo mas importante de los clientes frecuentes del teatro.

5.2. Mujeres jóvenes (18-28)

Este es el único grupo entre las mujeres que disminuyó los gastos en teatro con respecto al año pasado y en general, al ver los datos de los más jóvenes, estos tienden a ser los que menos invierten en teatro en todos los sentidos, por lo que consideramos que sería una buena estrategia formar una campaña orientada a la motivación de las mujeres más jóvenes para que gasten mas en el teatro y se pueda evitar que esta población joven empiece a ser cada vez mas escasa.

6. Video

Finalmente, se hace un video en donde se explica el análisis realizado. El video se puede encontrar en el siguiente [enlace](#).