UNIVERSIDAD DE SANTIAGO DE CHILE
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA INFORMÁTICA

Laboratorio 1 - Análisis Estadístico

Integrantes:	Felipe Cornejo
	Bastián Loyola
Curso:	Análisis de Datos
Sección	A-1
Profesor:	Max Chacón Pacheco

24 de Septiembre de 2022

Tabla de contenidos

1. Introducción

Dentro del campo de los bancos, específicamente en el área de mercadotecnia, donde hay distintas estrategias con distintos grados de efectividad, uno de ellos es el telemercadeo, el cual consiste en realizar llamadas telefónicas hacia potenciales clientes con la intención de presentar y entregar un servicio de depósitos a largo plazo, generando dos casos que se contrate el servicio o no, por lo tanto surge un problema interesante sobre tratar de predecir si la llamada será exitosa (Se contrata el servicio) o no lo será.

En este documento se presentará un estudio estadístico e inferencial de los datos y variables dado una base de datos con 41188 registros reales proporcionada por Moro et al. (2014), mediante el entorno de software que proporciona R (v. 4.2.1), se buscará realizar un estudio para determinar la importancia y relevancia de ciertas variables relacionadas con los posibles clientes, como su edad, trabajo, contactos previos realizados, entre otros.

Se presentará y desarrollará la problemática a plantear para el estudio con un detallado cada variable el cual contiene el conjunto de datos, para luego analizar de manera estadística e inferencial estas observaciones, finalmente converger a una conclusión sobre los resultados obtenidos.

1.1. Objetivos

1. Interiorizar la base de datos del problema con el fin de entender e interpretar cada una de las variables y sus tipos de datos.

2. Realizar estudio estadístico de datos y variables de la base de datos.

3. Realizar estudio inferencial de datos y variables de la base de datos.

4. Encontrar variables predictoras y su relevancia.

2. Descripción del problema

Un banco portugués realizó una campaña de telemarketing llamando a posibles clientes que realicen un depósito a largo a plazo, recopilando la información de dichas llamadas y clientes, a veces requiriendo más de una llamada para lograr determinar si se realizaría una subscripción en un depósito a largo plazo. Los datos recopilados permiten realizar un estudio a la efectividad de las campañas de telemercadeo y lograr predecir con ciertos datos si una llamada será exitosa (consiguiendo una subscripción).

2.1. Descripción de la base de datos

La base de datos utilizada para el análisis del problema fue la proporcionada por el estudio con enfoque en los datos en la predicción del éxito de una campaña de telemarketing bancario de un banco portugués, Moro et al. (2014). Esta base de datos contaba con, 41188 registros de llamada con un total de 20 variables de entrada y 1 de salida. Dentro de la base de datos se encontraron 12 datos duplicados, los cuales corresponden a un porcentaje de los datos muy pequeño, permitiendo eliminarlos para realizar un mejor estudio y análisis de los datos. Del estudio de telemarketing se pueden extraer 4 distintas bases de datos, diferenciándose en la cantidad de datos que se contienen y las variables de entradas utilizadas. Siendo las siguientes:

1. 'bank-additional-full' y su versión con el 10 % de los datos con 20 variables de entrada.
2. 'bank-full' y su versión con el 10 % de los datos con 16 variables de entrada.

A la base de datos a utilizar también se deberá realizar una limpieza de datos, según corresponda, para realizar estudios y análisis de las distintas variables, eliminando o sustituyendo variables desconocidas o nulas.

2.2. Descripción de clases y variables

Las variables de un cliente que fue contactado por una llamada en esta campaña de telemarketing corresponde a una registro de la base de datos mencionada.

2.2.1. Variable de Salida

Y: Es una variable de tipo binario que indicará la decisión del cliente al cual se llamó para ofrecer la subscripción a un depósito a largo plazo, los posibles valores de esta variable son 'yes' en caso de que el cliente se subscribió y 'no' en el caso contrario, se trata del objetivo deseado en este problema y una variable con gran relevancia para la campaña.

2.2.2. Variables Categóricas

1. **Job:** Tipo de trabajo el cual el cliente realiza, con los siguientes valores: 'admin.' (Administrativo), 'blue-collar' (Obreril), 'entrepreneur' (Emprendedor), 'housemaid' (Empleado Domestico), 'management' (Gerencia), 'retired' (Jubilado), 'self-employed' (Independiente), 'services' (Servicios), 'student' (Estudiante), 'technician' (Técnico), 'unemployed' (Desempleado) y 'unknown' (Desconocido).

2. **Marital:** Estado civil del cliente, con los siguientes valores: 'divorced' (Divorciado), 'married' (Casado), 'single' (Soltero) y 'unknown' (Desconocido).

3. **Education:** Último grado académico alcanzado del cliente, con los siguientes valores: 'basic.4y' (Básica 4 años),'basic.6y'(Básica 6 años),'basic.9y' (Basica 9 años), 'high.school' (Secundaria), 'illiterate' (Analfabeta), 'professional.course' (Curso Profesional), 'university.degree' (Título Universitario), 'unknown' (Desconocido).

4. **Default:** Si el cliente tiene un incumplimiento en el pago de su cuenta, los valores, es decir, tiene una deuda o está en mora, con los siguientes valores: 'no' (No), 'yes' (Sí), 'unknown' (Desconocido).

5. **Housing:** Si el cliente tiene una deuda de hipoteca o préstamo de vivienda con el bango, con los siguientes valores: 'no' (No), 'yes' (Sí), 'unknown' (Desconocido).

6. **Loan:** Si el cliente tiene una deuda personal con el banco, con los siguientes valores: 'no' (No), 'yes' (Sí), 'unknown' (Desconocido).

7. **Contact:** La forma de comunicación con el cliente en su último contacto, con los siguientes valores: 'cellular' (Celular), 'telephone' (Teléfono).

8. **Month:** Mes del último contacto, con los valores de cada mes reducido a tres letras características en inglés.

9. **Day_of_Week:** Día de la semana del último contacto, con los valores de cada día reducido a tres letras características en inglés.

10. **Poutcome:** El resultado de la campaña de marketing anterior, con los siguientes valores: 'failure' (Fracaso), 'nonexisting' (No existente), 'success' (Éxito).

2.2.3. Variables Numericas

1. **Age:** Edad del cliente al cual se realizó la llamada. Rango: [17-99] años

2. **Duration:** Duración de la ultima llamada en segundos, esta variable solo es conocida luego del contacto, afectando en gran medida si el cliente se subscribió a un depósito a plazo, si es 0 indica que el cliente no se subscribió. Esta variable solo debe ser considerada con propósitos de referencia, ya que para un modelo predictivo esta variable no se tendrá y deberá descatarse.

3. **Campaign:** Número de llamadas realizadas hacia este cliente en esta campaña

4. **pdays:** Días desde la última llamada realizada hacia este cliente en la última campaña, siendo el caso de 999 para aquellos clientes a los que no se contactó previamente.

5. **previous:** Número de llamadas realizadas hacia este cliente previo a la campaña actual.

6. **emp.var.rate:** Tasa de variación de empleo, es un indicador trimestral, siendo una variable de contexto económico y social, representa el cociente entre la cantidad de empleados sobre la población con edad y libertad suficiente para trabajar o buscar trabajo, Fontánez (1999)

7. **cons.price.idx:** Índice del precio al consumo es un indicador económico mensual que permite ver el aumento en el costo de vida, canasta y servicios basicos, es un indicador para la inflación, Bryan and Cecchetti (1993)

8. **cons.conf.idx:** Índice de confianza del consumidor, es un indicador social mensual que determina el grado de optimismo y confianza sobre la economía y su estado actual, Islam and Mumtaz (2016)

9. **euribor3m:** tasa Euribor a 3 mese, corresponde a una tasa calculada y publicada por la Federación Bancaria de Europa que indica el tipo de interés para préstamos entre bancos de la zona euro, Aranda (2011).

10. **nr.employed:** Número de empleados del banco portugués, indicador trimestral siendo una variable de contexto económico y social.

3. Análisis Estadístico e inferencial

Se realizó un análisis de los valores dentro de la base de datos completa que contaba con, 41176 registros que recopila la información de la campaña de telemarketing realizada por el barco portugués, sin duplicados. Como se mencionaba la variable de salida 'y' es una de alta importancia para el estudio de tele-mercadeo, ya que esta nos indica directamente los resultados de la campaña, los resultados obtenidos para la variable 'y' son 'yes' con 4639 clientes, el 11.27 % aproximadamente y para 36537 clientes que dijeron que respondieron 'no' representando al 88.73 %, tal como indica la Figura 1.

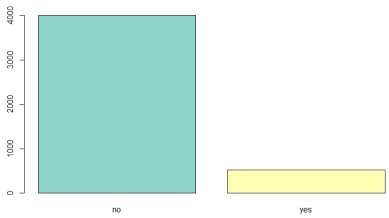


Figura 1: Frecuencia para la variable 'y'

3.1. variables categóricas

Empezando con las variables categóricas cabe destacar que se inició con un limpiado en la base de datos, donde se empezó revisando la cantidad de filas duplicadas, lo cual se encontraron 12 y se borraron.

El resumen de las variables, con la función 'summary()' de R, indica la moda de cada una de las variables respectivas, mostrando hasta cinco de los valores posibles para la variable, como se puede ver en la Figura 2.

Además se analizó la proporción del dato 'unknown' en cada una de las variables categoricas y al observar que en 'job' $(0,8\,\%)$, 'marital' $(0,19\,\%)$, 'education' $(4,2\,\%)$, 'housing' $(2,4\,\%)$ y 'loan' $(2,4\,\%)$ tenia el dato una proporción menor a $5\,\%$, se descartaron las observaciones las cuales existía 'unknown' en cada una de las variables descritas.

En resumen todos los datos a analizar son con los cambios realizados, es decir, sin duplicados ni datos NA que no superaran el $5\,\%$ en su variable respectiva.

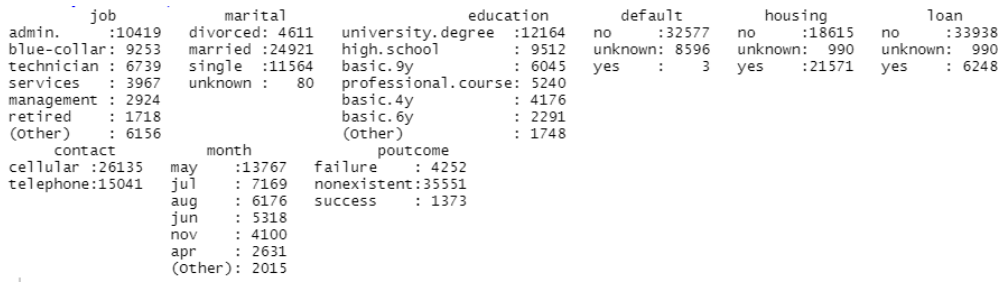


Figura 2: Resumen de la información estadística categórica

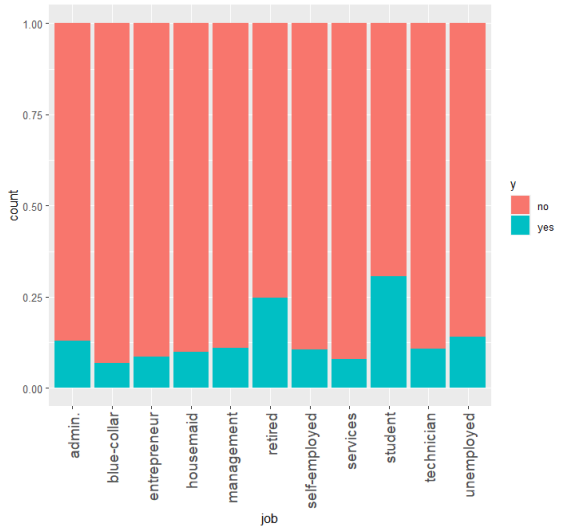
En relación sobre la correlación entre las variables categóricas en frente a la variable a predecir 'y', antes de realizar el modelo, se revisó cada una de las variables con pruebas de Chi Cuadrado de Homogeneidad para observar que variables tienen las mismas proporciones en frente a la variable resultante. Estos resultados se compararon con un nivel de significancia de 0,05 y se observó que 'housing' y 'loan' son homogeneos en frente a 'y', por tanto no se usaron para realizar el modelo para predecir el resultado 'y'.

3.1.1. Job

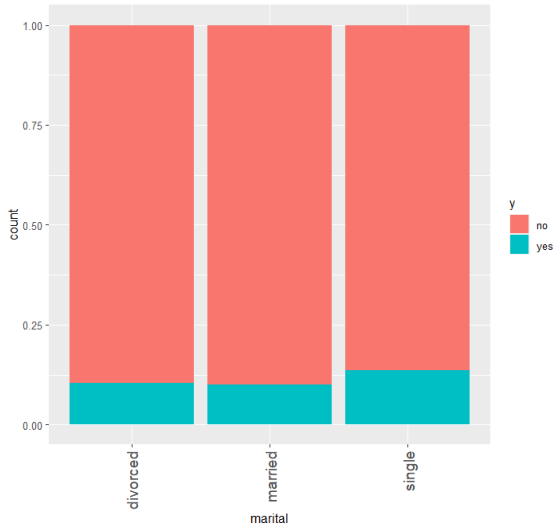
Esta variable da a conocer los tipos de clientes respecto a su trabajo. De esto se obtuvieron las proporciones respecto a cada trabajo sobre si la decisión de tomar el crédito a plazo fue 'yes' o 'no' como se muestra en la Figura 3a. Cabe notar que en siguiendo esta razón, los estudiantes son los que más dicen que sí enfrente a esta respuesta, seguido por los retirados y desempleados.

3.1.2. Marital

Luego, esta variable da a conocer la razón de clientes que se suscriben al depósito a plazo según su estado civil. El gráfico de la Figura 3b da a conocer las proporciones respecto a la respuesta por cada valor de la variable de esta sección. Se puede observar que entre las tres opciones las proporciones no varían mucho, mas la prueba de chi cuadrado de homogeneidad otorga un p-valor cercano a 0 al realizarlo sobre 'marital' e 'y'.



(a) Proporciones para cada trabajo por cada respuesta



(b) Proporciones para cada estado civil por cada respuesta

Figura 3: Distribución de proporciones para 'y'

3.1.3. Default

Al analizar 'default' se puede observar en la Figura 4 que personas que tienen algún tipo de incumplimiento con el pago de su cuenta su razón es cercana a 0 (existen valores de 'default' que son 'yes' y la respuesta 'y' también). Por otro lado, no se removieron los datos 'unknown' debido a que son más del 20 % de las filas de la base de datos, por lo que eliminarlos sería contraproducente para el análisis de los datos. No obstante, debido a la razón anterior acerca de su proporción, no aporta un valor significativo a la construcción de un modelo, por tanto se puede retirar esta variable.

3.1.4. Housing y Loan

Las variables 'housing' junto con 'loan', han otorgado valores de estudios similares, ambas variables tienen proporciones idénticas en sus posibles valores, obtenidos por medio de pruebas de chi cuadrado de homogeneidad con un nivel de confianza del 95 %, con p-valores de 0,05664 y 0,05772 respectivamente, además se corroboró con los gráficos de barras apiladas correspondientes a cada variable denotados en la Figura 5a para 'housing' y 5b para 'loan'.

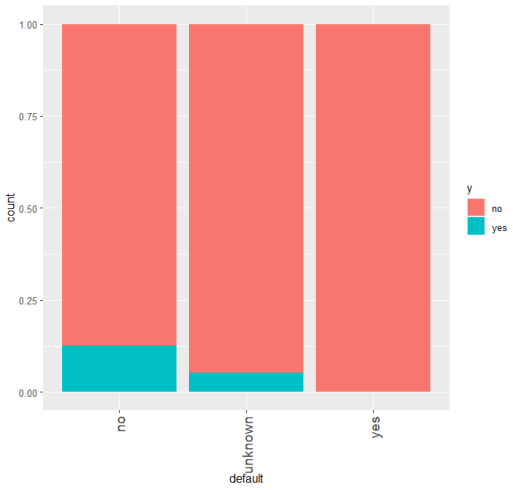
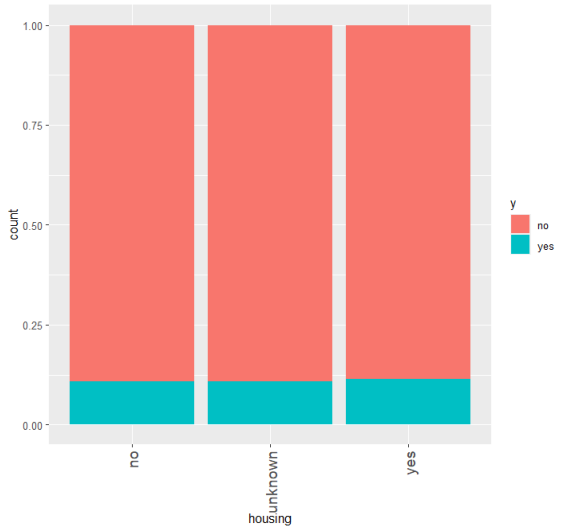
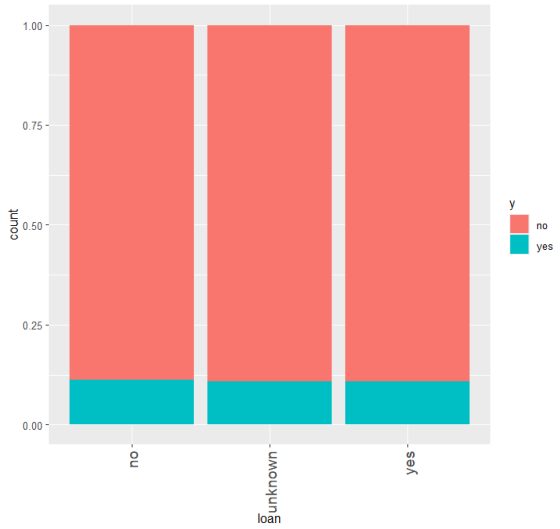


Figura 4: Proporción para cada valor de 'default' por cada respuesta

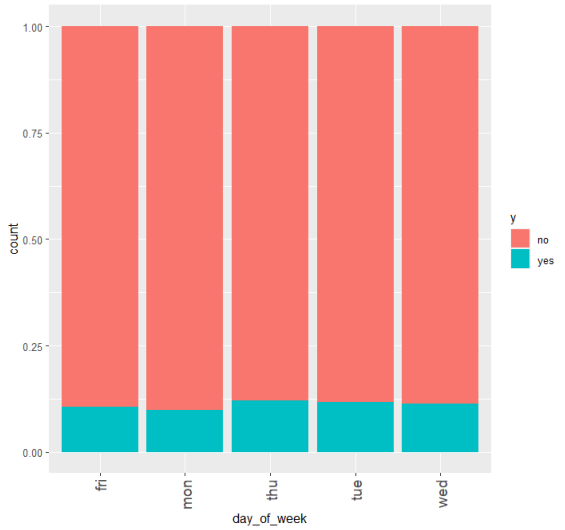


(a) Proporciones para cada valor de 'housing' por cada respuesta

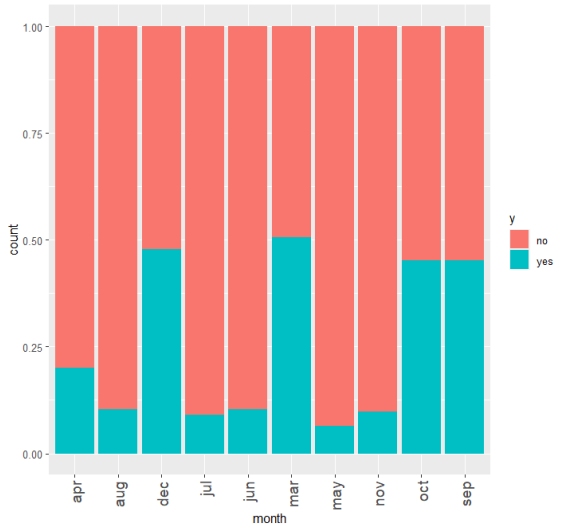


(b) Proporciones para cada valor de 'loan' por cada respuesta

Figura 5: Distribución de proporciones para 'y'



(a) Proporciones para cada 'day_of_week' por cada respuesta



(b) Proporciones para cada 'month' por cada respuesta

Figura 6: Distribución de proporciones para 'y'

3.1.5. Day_of_week y Month

Para empezar en esta sub sección, es necesario revisar no solo la proporción de ambas variables, sino que también la cantidad desde la Figura 2. Empezando por 'day_of_week', como se nota en la Figura 6a, todos los días de la semana en los cuales se realiza la llamada tienen casi la misma proporción, rompiendo así con la condición de existencia de una relación lineal entre los predictores y la respuesta. Esto es a pesar de que al realizar la prueba de chi cuadrado, otorga un p-valor por debajo y cercano al nivel de significación. Por otro lado, la variable 'month' en los gráficos de barras apiladas de la Figura 6b se puede ver una diferencia en las razones de cada mes, pero no solo eso, sino que la cantidad de llamadas hechas en Mayo es mucho mayor a cualquier otro mes (doblando la cantidad del segundo mes más llamado), por tanto esta variable podría ser un buen predictor.

3.1.6. Poutcome

Para finalizar esta sección de variables categóricas, 'poutcome' ha demostrado ser muy variable respecto a cada uno de sus valores. Al ver su gráfico en la Figura 7 se puede apreciar que el valor de 'success' al cruzarlo con 'yes' de la variable de respuesta tiene una

proporción de más de un $62,5\,\%$, haciendo esta una de las mayores proporciones cargadas hacia 'yes' en la variable 'y' en toda la base de datos, por tanto su participación como una variable predictora podría ser bastante útil para un modelo de regresión logística y así recabar los predictores útiles vistos en esta sección y en la siguiente.

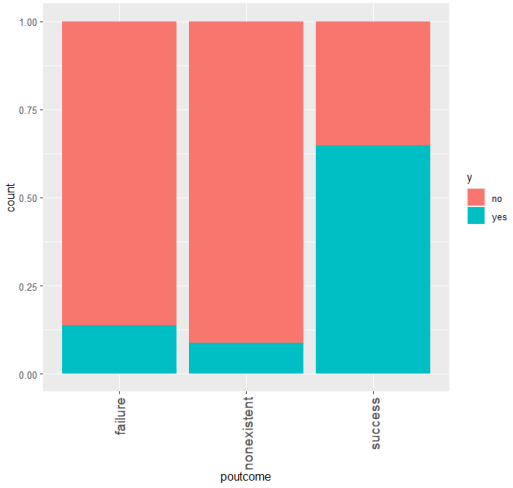


Figura 7: Proporción para cada valor de 'poutcome' por cada respuesta

3.2. variables numéricas

Con respecto a las variables numéricas se extrajo un resumen de la información estadística a través de la función 'summary()' de R, la cual entrego la información de la Figura 8. La información entregada por el resumen indica los siguientes: el valor mínimo, el primer cuartil, la mediana, promedio, tercer cuartil y el valor máximo. La base de datos utilizada contenía, 41176 contactos, que son 12 menos que la base de datos original, ya que estos eran duplicados.

```
      age            duration        campaign          pdays          previous       emp.var.rate      cons.price.idx
 Min.   :17.00   Min.   :   0.0   Min.   : 1.000   Min.   :  0.0   Min.   :0.000   Min.   :-3.40000   Min.   :92.20
 1st Qu.:32.00   1st Qu.: 102.0   1st Qu.: 1.000   1st Qu.:999.0   1st Qu.:0.000   1st Qu.:-1.80000   1st Qu.:93.08
 Median :38.00   Median : 180.0   Median : 2.000   Median :999.0   Median :0.000   Median : 1.10000   Median :93.75
 Mean   :40.02   Mean   : 258.3   Mean   : 2.568   Mean   :962.5   Mean   :0.173   Mean   : 0.08189   Mean   :93.58
 3rd Qu.:47.00   3rd Qu.: 319.0   3rd Qu.: 3.000   3rd Qu.:999.0   3rd Qu.:0.000   3rd Qu.: 1.40000   3rd Qu.:93.99
 Max.   :98.00   Max.   :4918.0   Max.   :56.000   Max.   :999.0   Max.   :7.000   Max.   : 1.40000   Max.   :94.77
 cons.conf.idx     euribor3m      nr.employed
 Min.   :-50.8   Min.   :0.634   Min.   :4964
 1st Qu.:-42.7   1st Qu.:1.344   1st Qu.:5099
 Median :-41.8   Median :4.857   Median :5191
 Mean   :-40.5   Mean   :3.621   Mean   :5167
 3rd Qu.:-36.4   3rd Qu.:4.961   3rd Qu.:5228
 Max.   :-26.9   Max.   :5.045   Max.   :5228
```

Figura 8: Resumen de la información estadística numérica

Otra información relevante para las variables numéricas es la correlación entre dichas variables, la cual indica la relación lineal y proporción entre dos variables estadísticas, dichos valores fueron obtenidos por medio de R, el resultado obtenido se encuentra en la Figura 9.

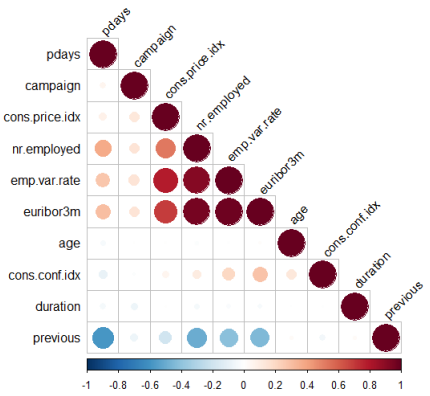


Figura 9: Correlación entre las variables numéricas

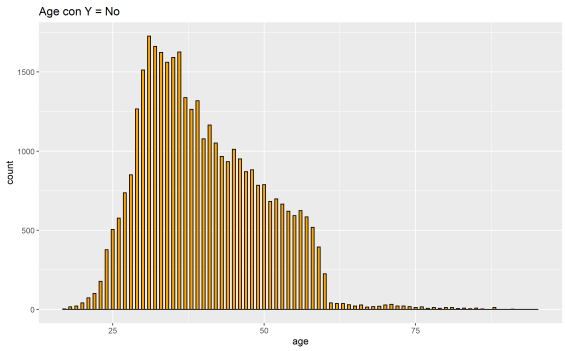
Observando la información de sobre las correlaciones realizadas a través de R por el método de correlaciones de Pearson, se puede apreciar qué gran parte de las variables presentan una correlación pequeña o nula, principalmente se destacaría las siguientes correlaciones.

1. emp.var.rate y nr.employed: esta correlación tiene sentido, ya que la primera variable describe la tasa de variación de empleo, si esta aumenta la cantidad de empleados también lo harán.

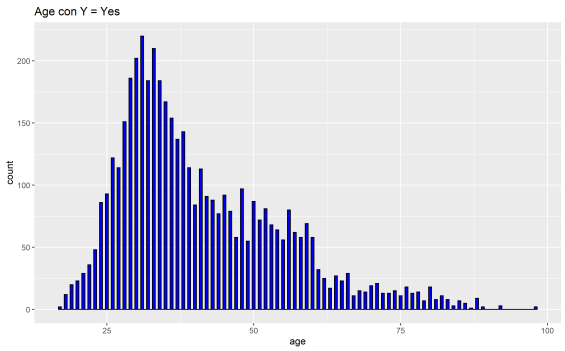
2. euribor3m con nr.employed y emp.var.rate: si euribor3m tiene una correlación grande con nr.employed también la tendrá con emp.var.rate, ya que, los ultimos también estaban correlacionados.

3.2.1. Age

Con respecto a la edad de las clientes de la campaña se puede apreciar en la Figura 10 que existe una mayor frecuencia de datos alrededor de la media de 40 años (Figura 8), además se puede apreciar en las Figuras 10b y 10b que gran parte de los clientes con una edad superior a 60 años contratan un depósito a largo plazo.

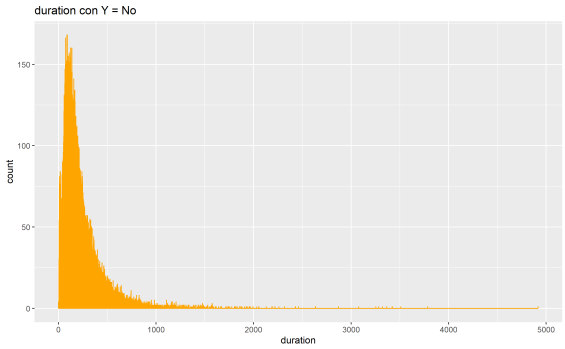


(a) Frecuencias para Age con 'y' = no

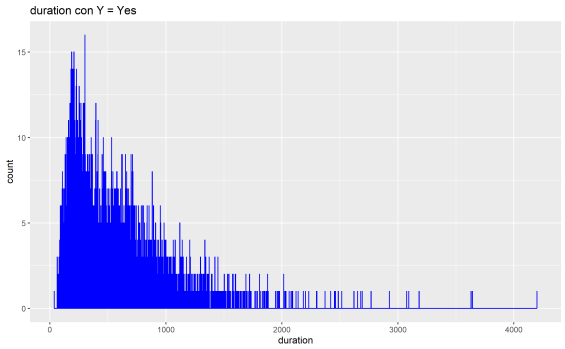


(b) Frecuencias para Age con 'y' = yes

Figura 10: Distribución de frecuencias para Age



(a) Frecuencias para Duration con 'y' = no



(b) Frecuencias para Duration con 'y' = yes

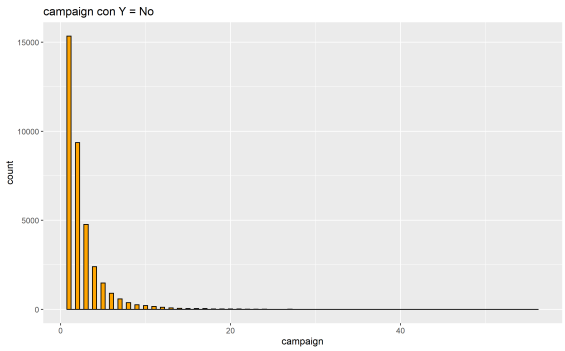
Figura 11: Distribución de frecuencias para Duration

3.2.2. Duration

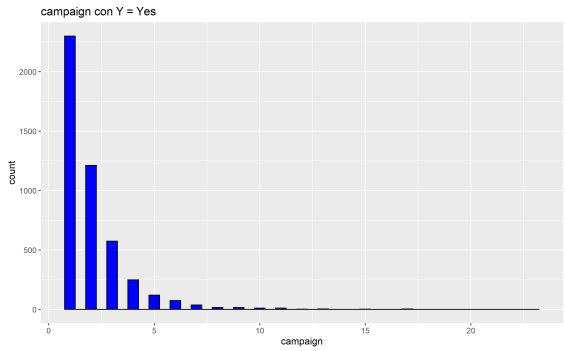
Con respecto a la duración de la llamada, como se había mencionado, esta variable debería ser descartada en un modelo predictivo, pero para un estudio estadístico nos permite revisar que donde 'y' es igual a 'no' (Figura 11a) las llamadas en gran parte no superan los 100 segundos, en caso contrarios (Figura 11b) para aquellos contactos exitosos las llamadas si se llegan a prolongar más del tiempo mencionado.

3.2.3. campaign

Como se puede apreciar tanto en el caso de llamadas exitosas como en el que no, el valor con mayor frecuencia fue de 1, lo que representa que a las personadas contactadas para esta campaña principalmente fueron contactadas solo una vez, además según las Figuras



(a) Frecuencias para campaign con 'y' = no



(b) Frecuencias para campaign con 'y' = yes

Figura 12: Distribución de frecuencias para campaign

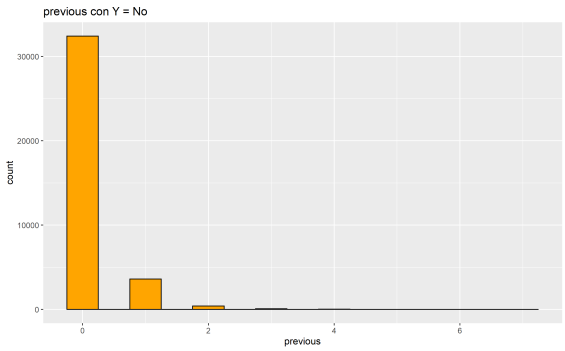
12a y 12b, la cantidad de contactos realizados no tiene relación con el valor de 'y'.

3.2.4. pdays

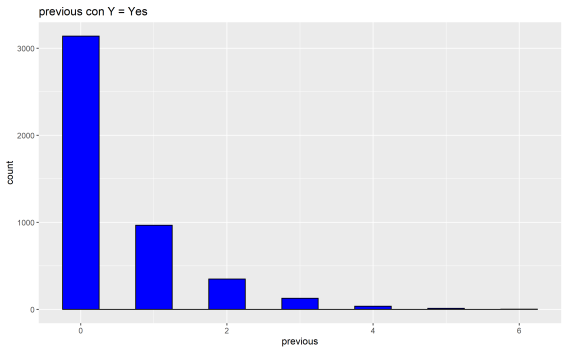
La variable numérica 'pdays' presentaba un valor especial '999' que representaba el caso en que el cliente no hubiese sido contactado en campañas previas, el cual modifica considerablemente la media, como se puede ver en la Figura 8, ya que este valor especial representa un gran porcentaje de los casos, con 39661 llamadas siendo un 93.32 % aproximadamente, lo cual impide un análisis de como afecta la cantidad de días previos al valor de 'y'

3.2.5. previous

Como se mencionó en la sección 3.2.4, el 93.32 % de los clientes no fueron contactados previamente, por lo tanto, se esperaría que el valor 0, fuera aquel que presentará mayor frecuencia, además este valor afecta considerablemente en la media de la variable, siendo muy cercano a 0 tal como se ve en la Figura 8, además se puede apreciar que los contactos realizados en campañas previas no exceden los 6 casos y no afectan en el valor de 'y' como se puede apreciar en las Figuras 13a y 13b

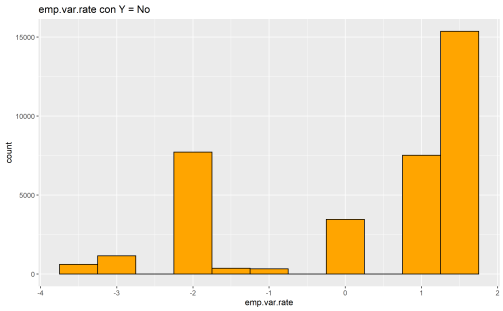


(a) Frecuencias para previous con 'y' = no

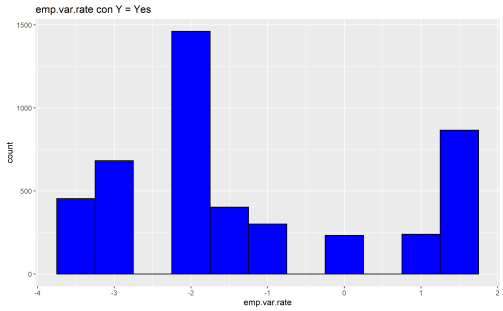


(b) Frecuencias para previous con 'y' = yes

Figura 13: Distribución de frecuencias para previous



(a) Frecuencias para emp.var.rate con 'y'=no



(b) Frecuencias para emp.var.rate con 'y'=yes

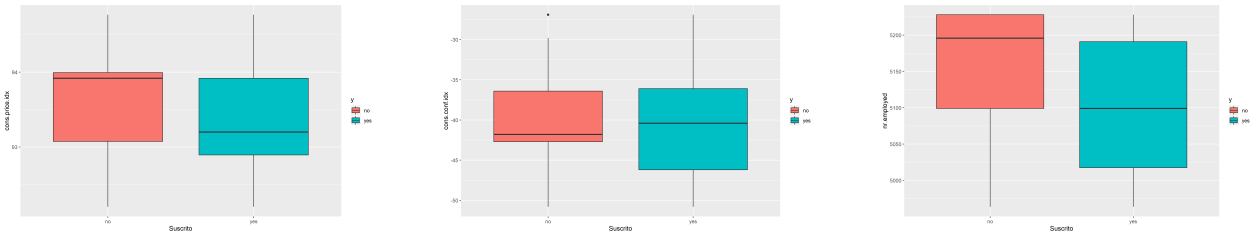
Figura 14: Distribución de frecuencias para previous

3.2.6. emp.var.rate

Aquellos casos de mayor frecuencia, donde 'y' es igual a 'no', en la Figura 14a se logra ver que es cuando la tasa de empleo es mayor, por el lado contrario en la Figura 14b se ve una mayor frecuencia cuando la variable es menor.

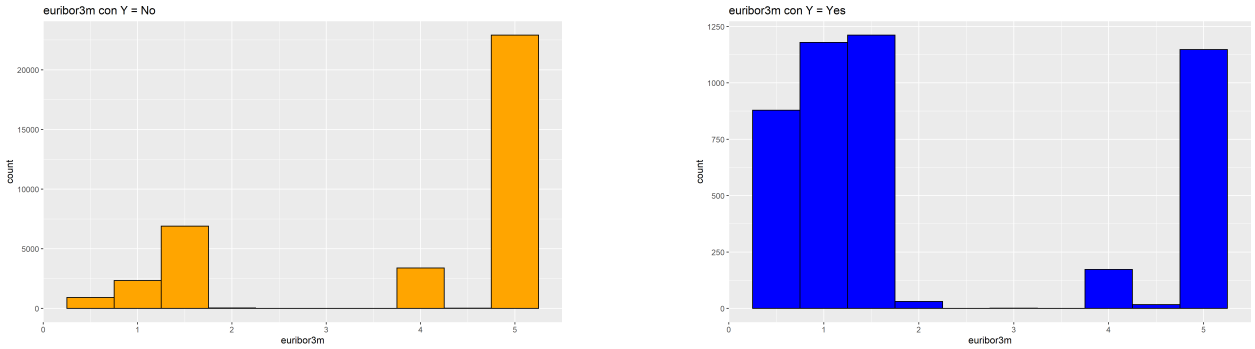
3.2.7. cons.price.idx, cons.conf.idx y nr.employed

Estas 3 variables presentan un nivel de dispersión muy grande tal como indican las Figuras 15a, 15b y 15c respectivamente, por lo que no se puede obtener mucha más información además de la vista en la Figura 8, de esta forma no nos entregan una información clara para el caso exitoso ni para el que no.



(a) Boxplot cons.price.idx (b) Boxplot cons.conf.idx (c) Boxplot nr.employed

Figura 15: Boxplots de variables sociales y economicas



(a) Frecuencias para euribor3m con 'y' = no (b) Frecuencias para euribor3m con 'y' = yes

Figura 16: Distribución de frecuencias para euribor3m

3.2.8. euribor3m

Acerca de la variable económica 'euribor3m' que regula la tasa de interés de préstamos interbancarios, se puede apreciar en la Figura 16a que cuando el valor es más alto, la mayoría de los clientes decide rechazar la subscripción al depósito, para el caso en que el contacto genere una subscripción se ve que estos casos se centran cuando el euribor es principalmente bajo, tal como se ve en la Figura 16b.

4. Conclusión

Llegada la última sección de esta experiencia, se ha realizado ya un análisis de cada variable en la base de datos del banco, los cuales se estudiaron separando las variables categóricas de las variables numéricas y así no interferir en pruebas estadísticas que sean excluyentes de una a la otra. Por el lado de las variables categóricas, se enfocó el estudio en la frecuencia (moda) y homogeneidad de las variables respecto a la variable resultante 'y', con el objetivo de obtener un conjunto de variables categóricas para la construcción de un modelo para la realización de una regresión logística a futuro. Luego, las variables numéricas se ahondaron en ellas, estudiando la correlación entre cada una de ellas, la distribución de los datos respecto a la opción escogida por cada uno de los clientes, los cuales se observó qué tipos de personas piden este crédito luego de la llamada respecto a datos los cuales se entregaron como numéricos.

Respecto a los datos borrados, ha sido necesario eliminar, primero, datos duplicados, los cuales existían en la base de datos, eran solo 12 en 41188 datos, para poder realizar los estudios sin problemas. Luego de aquello, se estudiaron las variables las cuales permitían valores problemáticos: 'job', 'marital', 'education', 'default', 'housing' y 'loan' para las variables categóricas y para las variables numéricas: 'pdays', la cual en sus valores, los cuales contenían 999 no permitían un buen análisis. Por tanto, para las variables que tenían 'unknown' en menos del 5 % de sus observaciones, solo se borraron las observaciones. Y las variables como 'pdays' y 'default' las cuales más del 93,32 % y 63,5 % respectivamente tenían ese valor, por tanto, y otras razones desarrolladas con anterioridad, no se consideran significativas para el estudio en general, al igual que 'day_of_month', 'housing' y 'loan' que sus tablas cruzadas tenían las mismas proporciones para la variable 'y'.

No obstante, hay puntos en los que fue complicado realizar los estudios, como a la hora de escoger las pruebas correctas para conocer los estadísticos necesarios para definir que tan significante es una variable respecto a la respuesta 'y'. La prueba que causo más problemas fue una que se indagó en literaturas de documentación de la tecnología R; V de Cramer, la cual no funcionó desde ninguna librería a descargar desde R Studio, ni siquiera copiando el código de fuente de la función. Por tanto, se escogió solamente probar con las

pruebas de Homogeneidad de Chi Cuadrado.

Finalmente, se cumplieron satisfactoriamente los objetivos propuestos, con ello definir las variables significativas para la realización de un modelo, el cual pueda predecir el valor de 'y'. Para esto, además se adquirió una inmersión del contexto y la base de datos, las cuales permitió afrontar cada una de las definiciones de variables e interpretar efectivamente los estadísticos resultantes. Y por último, se proponen nuevos objetivos, que se podrían lograr con mejores técnicas para el análisis de datos, los cuales permitan construir un modelo de regresión, el cual pueda predecir la variable 'y' sin depender de la variable 'duration'.

Anexos

Bibliografía

Aranda, J. F. G. (2011). Euríbor. *eXtoikos*, (3):121–122.

Bryan, M. F. and Cecchetti, S. G. (1993). The consumer price index as a measure of inflation.

Fontánez, J. L. (1999). *Análisis gráfico de la relación entre la tasa de participación, la tasa de empleo y la tasa de desempleo*. Unidad de Investigaciones Económicas, Departamento de Economía, Universidad ....

Islam, T. U. and Mumtaz, M. N. (2016). Consumer confidence index and economic growth: An empirical analysis of eu countries. *EuroEconomica*, 35(2).

Moro, S., Cortez, P., and Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, 62:22–31.

Anexos