

Econometría I

Workclass 03

Carlos A. Yanes Guerra

Universidad del Norte | Departamento de Economía

Contenido

Antes de empezar	1
Objetivo	1
Base de datos	1
Preguntas y solicitudes	2

Antes de empezar

Este tercer **workclass** puede ser entregado en grupo de máximo dos (2) estudiantes. En esta ocasión vamos a interactuar con una base de datos (ficticia). El tiempo para entrega es de *1 hora y 40 minutos*. El formato de recepción solo es en word o pdf en la carpeta del curso de **Brightspace**. No se admite por correo ni en formato distinto al solicitado.

Objetivo

Entender el manejo **interpretativo** de los datos. Familiarizarse con los paquetes de `tidyverse` y `moments` para hacer un análisis descriptivo de una base de datos.

Base de datos

Tiempo	Plan	Ingreso	Genero	Servicio
4312	43000	1840000	NB	Estudiante
5208	84000	1960000	NB	Premium
8311	42000	2320000	M	Estudiante
4712	42000	1810000	M	Estudiante
5645	84000	2890000	F	Premium
3765	62000	890000	F	Sencillo
2455	62000	920000	F	Sencillo
3664	62000	890000	NB	Sencillo
5535	91000	3490000	F	Premium
5212	92000	3650000	F	Premium
5245	94000	1290000	M	Premium
5141	92000	1790000	M	Premium
5425	92000	2890000	F	Premium
4312	42000	1850000	M	Estudiante
5711	42000	1810000	M	Estudiante
6642	44000	1980000	F	Estudiante
6842	44000	1000000	M	Estudiante
6712	45000	1850000	NB	Estudiante

La base de datos anterior contiene información un grupo de estudiantes de sus hábitos televisivos, se encuentran el número de horas que dedica en un mes a mirar cualquier contenido de entretenimiento, el valor del plan que paga, nivel de ingresos, género y por último el tipo de suscripción que tiene. Las últimas, son variables de tipo *binaria* y categórica que toman valores cualitativos.

Preguntas y solicitudes

De acuerdo a las respuestas situadas en la base, importe a R y/o Rstudio. *Si importa directamente y debe trabajar con este tipo de variables (cualitativas) la idea es convertirlas en un factor o etiqueta estipulada.* Esta etiqueta de cada variable cualitativa debe tomar un número o clasificación. Por ejemplo: (1=Premium, 2= Sencillo, 3= Estudiante). Las de genero puede codificarla con el orden que usted desee pero respetando la numeración de 0, 1 y 2 respectivamente para cada genero expuesto.

1. Piense por un segundo que desea agregar información de una persona de genero femenino, con un ingreso de dos millones seiscientos mil pesos m/lc, un plan sencillo, y que reporta utilizar en el mes alrededor de 6023 horas. Paga regularmente \$54900. Utilice el operador pipe %>% y adicione la fila. Cree una nueva variable (para todxs) con la opción de mutate, que nos diga que porcentaje del ingreso se destina al pago del servicio. No olvide activar el paquete de tidyverse para eso. *Muestre todo el proceso.*
2. Presente una tabla con un resumen estadístico de las variables *Plan, Ingreso y Tiempo* en formato tibble. No solo plantee el código, interprete los resultados.
3. ¿Qué porcentaje de hombres, mujeres y no binarios tienen plan premium? ¿Considera que el nivel de ingreso tiene que ver con esa selección de plan?. ¿Qué tipo de servicio por ende es el mas usado entre todos los suscritos?. ¿Qué porcentaje de estudiantes de genero binario usan el tipo de servicio de estudiante? Pista

```
objeto<-table(base$variable_cualitativa, base$variable_cuantitativa)
porcentaje_x <- prop.table(objeto)
porcentaje_x
```

4. ¿Qué tipo de distribución, curtosis y asimetría posee las horas de entretenimiento de las personas encuestadas? *Explique.*

5. Desarrolle la gráfica de dispersión y linea de ajuste con el siguiente código base:

```
grafico<- base_datos %>%
  ggplot(aes(y = Tiempo, x = Ingreso))+
  geom_point() +
  ylab("Tiempo (en horas)") + xlab("Ingreso") +
  geom_smooth(method = 'lm', col = 'red', size = 0.5)

grafico # Salida de Gráfico
```

- Interprete el gráfico que observa entre las variables
- Escriba la ecuación econométrica de regresión de lo anterior