Tarea 1

Fecha de entrega: Lunes 14 de octubre, 11:59 P.M.

Nota: Deberán subir a *Canvas* **dos archivos**. Un **primer archivo** de texto con sus respuestas. Este archivo debe ser auto-suficiente. Esto quiere decir que con revisar este archivo debe ser posible calificar su tarea en su totalidad. Para este archivo pueden utilizar el formato de su preferencia (e.g. pdf, LaTeX, Word u hojas escritas a mano y escaneadas). El **segundo archivo** es un archivo de soporte que genere todos sus resultados. Típicamente este será un archivo de programación, mismo que al ejecutarse replicaría los resultados que estás reportando en tu primer archivo. Dado que ustedes son libres de elegir el software que utilizarán, este puede ser un R-script, Do-File o similar.

BILLONARIOS

En un mundo donde la desigualdad económica es un tema de creciente preocupación, el estudio de los individuos más ricos del planeta ofrece una perspectiva única sobre la acumulación y distribución de la riqueza global. Dado que tu trabajas en en el Departamento de Análisis Económico Global de las Naciones Unidas, se te ha encargado realizar un estudio sobre la concentración de riqueza extrema con el objetivo principal de examinar cómo diversos factores socioeconómicos, como la edad, el origen geográfico, el ser "self-made", y las características económicas de los países de origen, están relacionados con la acumulación de riqueza extrema. Este análisis nos permitirá comprender mejor los patrones de riqueza global, la desigualdad interna entre los ultra-ricos, y cómo la riqueza de estos individuos se compara con indicadores económicos nacionales.

Para este estudio, contamos con la base de datos *Philanthropy_2023.csv* que incluye datos sobre los 205 multimillonarios, con información adicional sobre sus actividades filantrópicas. Al final de esta tarea pueden encontrar una tabla con una explicación de todas las variables de esta base.

1. Procesamiento inicial de datos

Como primera fase de este análisis, se te ha solicitado crear algunas variables adicionales a las existentes en las bases de datos. **Nota:** NO debes dar ninguna respuesta a esta sección, solo son instrucciones de variables que debes crear.

- Una nueva variable $worth\_pct\_gdp$, que represente el porcentaje que la riqueza de cada billonario ($finalWorth$) representa con respecto al PIB de su país ($gdp\_country$)

$$worth\_pct\_gdp = 100 * \frac{finalWorth}{gdp\_country}$$

- Una nueva variable gdp_per_capita, que represente el PIB per capita del país de origen del multimillonario.

$$gdp\_per\_capita = \frac{gdp\_country}{population\_country}$$

- Una nueva variable $worth\_v\_gdp\_pc$, que capture cuántas veces es más rico el billonario observado en relación con el PIB per cápita de su país.

$$worth\_v\_gdp\_pc = \frac{finalWorth}{gdp\_per\_capita}$$

- Variable $finalWorth\_75$: Crea una variable dummy llamada $finalWorth\_75$ que tome el valor de 1 si el patrimonio neto ($finalWorth$) del multimillonario está por encima del percentil 75 de la distribución de multimillonarios, y 0 en caso contrario. Es decir, esta dummy te permitirá identificar al 25 % de multimillonarios con más dinero dentro de esta muestra de multimillonarios.

- Crea las siguientes variables dummy que representen las principales industrias a las que pertenecen los multimillonarios. Cada dummy deberá tomar el valor de 1 si el multimillonario pertenece a dicha industria, y 0 si no pertenece:

  - $Technology$: Toma el valor de 1 si el multimillonario pertenece a la industria de Tecnología.
  - $Finance\ Investments$: Toma el valor de 1 si el multimillonario pertenece a la industria de Finanzas e Inversiones.
  - $Fashion\_Retail$: Toma el valor de 1 si el multimillonario pertenece a la industria de Moda y Retail.
  - $Manufacturing$: Toma el valor de 1 si el multimillonario pertenece a la industria de Manufactura.

- *Food_Beverage*: Toma el valor de 1 si el multimillonario pertenece a la industria de Alimentos y Bebidas.
- *Others*: Crea una dummy adicional que tome el valor de 1 si el multimillonario pertenece a una industria diferente a las mencionadas anteriormente o si no reporta su industria.

2. Análisis descriptivo

Una vez creadas las variables anteriores, debes contestar las siguientes preguntas.

1. Reporta un panorama general de las variables en tu base de datos. Para esto, debes crear una tabla con estadísticas descriptivas básicas (número de observaciones, media, desviación estándar, valores mínimos y máximos). Tu debes decidir qué variables vas a incluir [Tip: si la variable será utilizada en el análisis de la Tarea es señal de que deberías incluirla en esta tabla]. Si alguna de las variables es de tipo dummy, señálalo de alguna forma en la tabla [Tip: por ejemplo, puedes poner un asterisco en el nombre de las variables que sean dummies y en el pie de la tabla indicar qué quiere decir dicho asterisco]. Elige 4 variables que consideres interesantes y di algo BREVE acerca de la estadística descriptiva de dichas variables. [R tip: Para hacer la tabla puedes usar el comando *stargazer*].

2. Realiza una gráfica que muestre la distribución de *worth_pct_gdp* y proporciona una breve descripción de lo que observas con esta gráfica. Fíjate cómo se modifica la distribución cuando utilizas el logaritmo de *worth_pct_gdp*.

3. Imagina que tu objetivo principal es estimar la desviación estándar de la variable *worth_pct_gdp*. Definamos a este parámetro como σ_w. Utilizando *bootstrap*, y 500 simulaciones con submuestras de tamaño 205 (misma N que la muestra), crea un histograma para la estimación de σ_w.

4. Utilizando el resultado de la pregunta anterior y sin asumir nada acerca de la distribución de σ_w: (i) calcula la varianza de σ_w y repórtala; (ii) calcula un intervalo de confianza de 95 % para σ_w.

5. (Puntos extra) ¿Qué porcentaje del PIB de México representa el patrimonio neto de Carlos Slim Helú & Family?

3. Análisis de riqueza y factores socioeconómicos

De acuerdo con observaciones preliminares y estudios económicos previos, existe la intuición de que diversos factores como la edad, el origen geográfico y el ser *self-made* podrían influir significativamente en el nivel de riqueza de los multimillonarios. Este análisis nos permitirá comprender mejor cómo estos factores se relacionan con la acumulación de riqueza extrema.

1. Haz dos gráficas:

   a) Ilustra la relación entre la edad (X) y el logaritmo del patrimonio neto (Y) distinguiendo entre multimillonarios *self-made* y los que no lo son [R tip: Puedes usar $color = selfMade$ en ggplot para incluir esta distinción].

   b) Ilustra la relación entre el PIB del país de origen y el patrimonio neto, también distinguiendo entre *self-made* y no *self-made*. Utiliza logarítmos de ambas variables para visualizar mejor la relación.

   c) Agrega en las gráficas anteriores la estimación de las siguientes especificaciones:

$$\begin{aligned} \log(\text{finalWorth}_i) &= \beta_0 + \beta_1 \text{age}_i + \cdots + U_i \\ \log(\text{finalWorth}_i) &= \beta_0 + \beta_1 \log(\text{gdp\_country}_i) + U_i \end{aligned} \tag{1}$$

   En el caso de la primera estimación ($finalWorth$ vs age) tu deberás de decidir y justificar el grado de polinomio a incluir.

   En la respuesta de los incisos anteriores, pon la gráfica que sí incluya estas estimaciones, no hay necesidad de duplicar la gráfica. Como respuesta a este inciso, reporta el resultado de las estimaciones en formato de ecuación e incluye una interpretación de los valores estimados relevantes para cada ecuación. Justifica si el valor estimado de ambas especificaciones es grande o pequeño. [Nota: al final de la tarea viene un ejemplo de como reportar los resultados con formato de ecuación].

2. Basado en lo que observas en las gráficas de la pregunta anterior, justifica qué errores estándar deberías de utilizar: *homocedásticos* o *heterocedásticos*. En adelante, deberás utilizar dichos errores en las estimaciones del resto de la tarea.

3. Para continuar con el análisis deberás llevar a cabo las siguientes estimaciones:

   a) $finalWorth$ vs age, selfMade, female, log ($gdp\_country$), total_tax_rate_country, PhilantropyScore, education, MaritalStatus

   b) log ($finalWorth$) vs age, selfMade, female, log(gdp_country), total_tax_rate_country, PhilantropyScore, education, MaritalStatus

c) $worth\_v\_gdp\_pc$ vs age, selfMade, female, log(gdp_country), total_tax_rate_country, PhilantropyScore, education, MaritalStatus

d) $finalWorth\_75$: vs age, selfMade, female, log(gdp_country), total_tax_rate_country, PhilantropyScore, education, MaritalStatus

Tu deberás definir cuál es la mejor forma de agregar las variables explicativas correspondientes (por ejemplo, cuando digo *MaritalStatus* sería bueno que al mneos incluyas una dummy para Married). Reporta los resultados en una tabla de regresiones con el formato que hemos visto en clase. Los coeficientes estimados deben venir acompañados de asteriscos para indicar su nivel de significancia estadística, de la siguiente manera: * para el 10 %, ** para el 5 %, y *** para el 1 %. Se emplearán errores heterocedásticos para las estimaciones. [R Tip: Para correr una regresión, utiliza el comando *lm*. Para producir las tablas, usa el paquete *stargazer*].

4. Sin importar la significancia estadística, da una interpretación lo más específica posible para el valor estimado de los coeficientes asociados a las siguientes variables. Si hay más de un coeficiente asociado a dicha variable, elige uno (por ejemplo: si para MaritalStatus creaste tres dummies solo necesitas interpretar el coeficiente de una de esas tres variables).

   a) $selfMade$ en la especificación (3a)
   b) $\log(gdp\_country)$ en la especificación (3a)
   c) $total\_tax\_rate\_country$ en la especificación (3b)
   d) $MaritalStatus$ en la especificación (3b)
   e) age en la especificación (3c)
   f) $education$ en la especificación (3c)
   g) $PhilatropyScore$ en la especificación (3d)
   h) $\log(gdp\_country)$ en la especificación (3d)

5. Tu jefe está interesado en saber si las industrias tienen un impacto en conjunto sobre la riqueza de los multimillonarios.

   a) Propón cómo extendrías la especificación (3b) para poder dar evidencia acerca de la pregunta que te plantea tu jefe. Reporta en una columna adicional de la tabla que creaste el resultado que obtuviste de tu estimación.
   b) Plantea claramente la prueba de hipótesis que necesitarías evaluar para responder la pregunta de interés.
   c) ¿Rechazas o no la hipótesis nula a un nivel de significancia del 5 %? ¿Qué implica esto sobre el impacto de las industrias en el patrimonio neto de los multimillonarios?

d) Comenta sobre los resultados y los impactos de las distintas industrias sobre el patrimonio de los multimillonarios.

e) Compara el coeficiente de female con y sin las variables de industria (i.e. la especificación (3b) y la que creaste en esta pregunta (5a). ¿Qué te dice el cambio del coeficiente acerca de la relación de female con las variables de industria?

6. Extiende tu especificación de la pregunta anterior (5a) para poder ver con una sola regresión la diferencia del patrimonio entre multimillonarios hombres y mujeres (por nacimiento) para todas las industrias (utilizando la agrupación de industrias que creamos en la primera sección.

   a) Indica la ecuación de tu nueva especificación resaltando los coeficientes de interés que capturarán la respuesta a esta pregunta. Solo indica la ecuación, no el resultado de la estimación.

   b) Reporta en una gráfica el intervalo de confianza de 90 % para la diferencia de todas las industrias. Reporta los valores exactos para la industria de *Food and Beverage*.

Tabla 1: Descripción de variables *Philanthropy_2023*.

Variable	Descripción
rank	Posición en la lista de multimillonarios.
finalWorth	Valor neto final en millones de dólares.
personName	Nombre del multimillonario.
age	Edad del multimillonario.
country	País de origen del multimillonario.
industries	Industria a la que pertenece el multimillonario.
selfMade	Variable que indica si la persona es self-made (1 si lo es, 0 si no lo es).
Self.Made.Score	Puntuación según Forbes (1-10), donde 1 es fortuna heredada sin gestionarla, y 10 es completamente self-made superando obstáculos significativos.
Philanthropy.Score	Puntuación de filantropía según Forbes (1-5), donde 1 es menos del 1 % de la riqueza donada, y 5 es más del 20 %.
Marital.Status	Estado civil del multimillonario (Single, Married, Divorced, Widowed).
Bachelor	Dummy para indicar si tiene un título de Bachelor (1 si lo tiene, 0 si no).
Master	Dummy para indicar si tiene un título de Master (1 si lo tiene, 0 si no).
Doctorate	Dummy para indicar si tiene un título de Doctorado (1 si lo tiene, 0 si no).
Drop.Out	Dummy para indicar si la persona dejó sus estudios sin terminarlos (1 si es un drop-out, 0 si no).
gdp_country	Producto Interno Bruto (GDP) del país en mil millones.
gross_tertiary_education_enrollment	Tasa de inscripción en educación terciaria en el país.
tax_revenue_country	Ingreso tributario del país en millones.
total_tax_rate_country	Tasa de impuestos total en el país.
population_country	Población del país en millones.
female	Variable indicadora de género (1 si es mujer, 0 si es hombre).

Ejemplo de reportando el resultado de una estimación con un diagrama de dispersión:

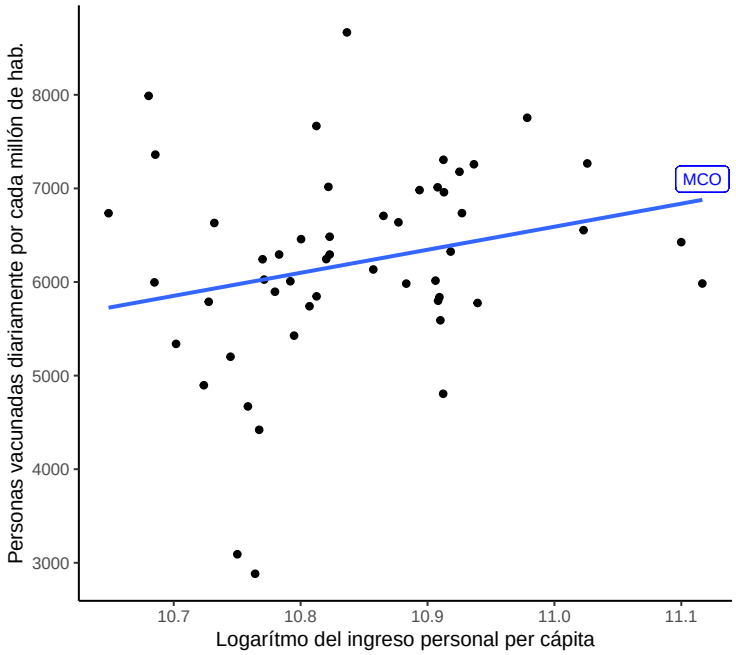

Figura 1: Scatterplot entre $dvaxx\_per\_mill$ y $\ln inc\_pc$

Ejemplo de reportando el resultado de una estimación con formato de ecuación:

$$\ln(ing\_trim_i) = \underset{(0.0044)}{9.601} - \underset{(0.0069)}{0.411}\, rural$$