



En el archivo `Hunger GDP Extreme Poverty.csv` se encuentra por país y a finales de 2018, el índice Global de Hambre (GHI, del 0 (no existe hambre) al 100), el porcentaje de la población viviendo en hogares bajo la línea de pobreza extrema (PBL) y el Producto Interno Bruto a valores de Paridad de Poder Adquisitivo (GDP, en cientos de dólares estadounidenses del 2017). Va a utilizarse un modelo de regresión lineal múltiple para explicar cómo influye el porcentaje de la población viviendo en hogares bajo la línea de pobreza extrema y el Producto Interno Bruto a valores de Paridad de Poder Adquisitivo en el índice Global de Hambre.

1. Utilizando Stata:

- a) Verifica si los datos usados contienen multicolinealidad según el criterio de la inflación de la varianza.

Solución: Utilizando el comando `vif` después de ejecutar la regresión `regress ghi bpl gdp` en Stata obtuvimos los siguientes resultados

Variable	VIF	1/VIF
bpl	1.67	0.598685
gdp	1.67	0.598685
Mean VIF	1.67	

por lo que podemos concluir que este modelo no presenta el problema de multicolinealidad ya que la inflación de la varianza no es mayor a 10.

□

- b) Realiza una gráfica de los valores esperados estimados contra los residuales y concluye si gráficamente existe o no heteroscedasticidad.

Solución: Generamos los valores de los residuales con `predict rrl, residuals` y los valores estimados con `predict yh` y obtenemos la gráfica con `scatter rrl yh`

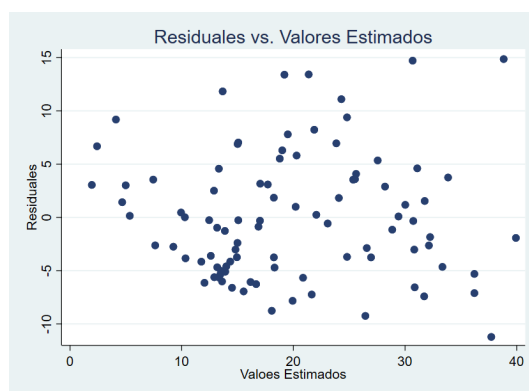


Figura 1: Gráfica valores estimados contra sus residuales

como podemos observar hay indicios de heteroscedasticidad ya que los datos están muy dispersos, lo valores esperados crecen más que el de los residuales.

□

- c) Indica si existe o no heteroscedasticidad respecto a la variable BPL usando la prueba de Goldfeld-Quandt, a un nivel de significancia de 0.05.

Solución: Primero ordenamos los datos BPL con `sort bpl ghi gdp` como tenemos más de 60 datos entonces no consideramos los 10 observaciones de en medio, y realizamos la regresiones a los grupos `regress ghi bpl gdp in 1/41` y `regress ghi bpl gdp in 52/93` y guardamos los escalares `scalar VG1=e(rmse)^2` y `scalar VG2=e(rmse)^2` al ejecutar cada regresión, para así obtener el estadísticos $EstF = VG2/VG1 = 2.0073499$ como es una prueba con nivel de significancia del 0.05 entonces calculamos los umbrales `scalar UmbralD=invFtail(40,40,0.025)` y `scalar UmbralL=invFtail(40,40,0.975)` y al usar `scalar list` obtuvimos

```
UmbralD  = 1.8751974
UmbralL  = 0.5332772
EstF     = 2.0073499
VG2      = 45.487781
VG1      = 22.660613
```

por lo tanto el estadístico F esta en el área de rechazo por lo que existe heteroscedasticidad.

□

- d) Indica si existe o no heteroscedasticidad usando la prueba de Breusch-Pagan-Godfrey, a un nivel de significancia de 0.04.

Solución: Volvemos a cargar los datos para hacer la regresión sin orden, entonces ejecutando `regress ghi bpl gdp` y después usando `estat hettest bpl gdp` obtuvimos que el p -valor es igual a $p = 0.1439$ y por lo tanto como $p > 0.04$ se acepta la hipótesis nula y se concluye que no existe heteroscedasticidad.

□

- e) Indica si existe o no heteroscedasticidad usando la prueba de White, a un nivel de significancia de 0.04.

Solución: Para la prueba de White ejecutamos el siguiente comando `estat imtest, white` y obtuvimos que el p -valor es igual a $p = 0.1143$ y por lo tanto como $p > 0.04$ se acepta la hipótesis nula y se concluye que no existe heteroscedasticidad.

□

- f) Asume que existe heteroscedasticidad. Encuentra el valor esperando del Índice Global de Hambre para un país donde el 3.85% de la población vive por debajo de la línea de la pobreza con un GDP-PPA de 13.15, ya habiendo removido la heteroscedasticidad y asumiendo que su forma funcional es $\sigma_i^2 = 3BPL_i + GDP_i$.

Solución: Generamos la forma funcional de la varianza con el comando `generate FFV=3*bpl+gdp` y los pesos con `generate W=1/FFV` entonces ejecutamos la regresión `regress ghi bpl gdp [aweight=W]` y finalmente para obtener el valor esperando del Índice Global de Hambre para un país donde el 3.85% de la población vive por debajo de la línea de la pobreza con un GDP-PPA de 13.15 utilizamos `mfxx, at(3.85,13.15)` donde obtenemos

$$y = 13.052484$$

□

- g) Asume que existe heteroscedasticidad. Muestra el estimador S_0 de White usado para estimar la matriz varianza-covarianza.

Solución: **NOTA:** Este inciso esta al final del Archivo log de STATA.

Calculamos los residuales de la regresión con `predict u, residuals` y generamos los datos `generate r=u^2` y creamos el vector `mkmat r` y finalmente la matriz diagonal con los residuales al cuadrado `matrix D=diag(r)` y ahora, para la matriz X basta con generar una columna de 1 y esto lo hacemos con `generate o=1` para así obtener X con `mkmat o bpl gdp, matrix(X)` y finalmente obtenemos la matriz S_0 con `matrix S=1/93*X'*D*X` con lo que obtenemos

$$S_0 = \begin{pmatrix} 33.645456 & & \\ 747.02864 & 34720.768 & \\ 246.18768 & 2264.7268 & 3416.3611 \end{pmatrix}$$

□

h) Indica si existe o no autocorrelación usando la prueba de las rachas, a un nivel de significancia de 0.035.

Solución: Ordenamos los datos BPL con `sort bpl ghi gdp` y realizamos la regresión con `regress ghi bpl gdp` y procedemos a obtener los residuales con `predict u, residuals` y realizamos la prueba de rachas con `runtest u, threshold(0)` con lo que obtuvimos un p -valor igual a $p = 0.54$, por lo tanto a un nivel de significancia de 0.035 tenemos que $p > 0.035$ por lo tanto se acepta la hipótesis nula y existe autocorrelación en la variable BPL.

□

i) Indica si existe o no autocorrelación usando la prueba de Durbin - Watson, a un nivel de significancia de 0.01.

Solución: Siguiendo con el mismo modelo del inciso anterior basta con generar los datos `generate num=_n`, definiendo `tsset num` para finalmente ejecutar la prueba de Durbin - Watson con `estat dwatson` tenemos que $d = 1.975953$ entonces de la tabla de Durbin - Watson tenemos que $k = 2$ y $n = 93$ por lo tanto

$$dL = 1.489 \text{ y } dU = 1.573 \Rightarrow 4 - dL = 2.511 \text{ y } 4 - dU = 2.427$$

por lo tanto como $dU < d < 4 - dU$ tenemos que no hay autocorrelación.

□

j) Asume que existe autocorrelación. Encuentra el valor esperado del Índice Global de Hambre para un país donde el 4.45 % de la población vive por debajo de la línea de la pobreza con un GDP-PPA de 12.54, ya habiendo removido la autocorrelación.

Solución: Volveremos a cargar los datos originales y ejecutaremos la regresión con los datos originales `regress ghi bpl gdp` y ejecutaremos la prueba de D-W asumiendo que existe autocorrelación, con lo que obtuvimos $d = 2.263144$ y entonces generamos los siguientes datos `generate ghiauto=ghi-(1- 2.263144/2)*ghi_01`, `generate bplauto=bpl-(1- 2.263144/2)*bpl_01` y `generate gdpauto=gdp-(1- 2.263144/2)*gdp_01`, finalmente ejecutamos la regresión `regress ghiauto bplauto gdpauto in 2/93` y evaluamos con `mfxx, at(4.45,12.54)` con lo que obtuvimos

$$y = 17.219622$$

□

2. Repetir el inciso 1) utilizando R.

- a) Verifica si los datos usados contienen multicolinealidad según el criterio de la inflación de la varianza.

Solución: Ejecutamos la regresión lineal con el comando `rl<-lm(GHI BPL+GDP,data = datos)` y `vif(rl)` en R con lo que obtuvimos

BPL	GDP
1.670327	1.670327

por lo que podemos concluir que este modelo no presenta el problema de multicolinealidad ya que la inflación de la varianza no es mayor a 10.

□

- b) Realiza una gráfica de los valores esperados estimados contra los residuales y concluye si gráficamente existe o no heteroscedasticidad.

Solución: Para obtener la gráfica de los valores esperados contra los residuales simplemente ejecutamos el comando

```
plot(predict(rl),rl$residuals,ylab = "Residuales", xlab = "Valores estimados",  
      title(main = "Residuales vs. Valores estimados"),pch=19,col="darkorchid1")
```

donde obtuvimos la siguiente gráfica:

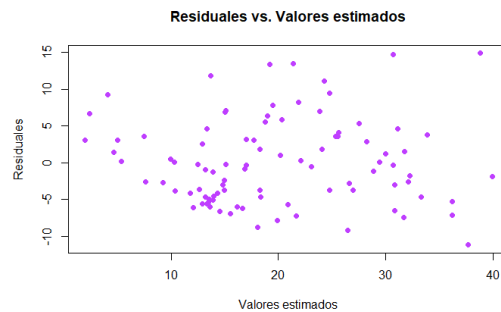


Figura 2: Gráfica valores estimados contra sus residuales

como podemos observar hay indicios de heteroscedasticidad ya que los datos están muy dispersos, los valores esperados crecen más que los de los residuales.

□

- c) Indica si existe o no heteroscedasticidad respecto a la variable BPL usando la prueba de Goldfeld-Quandt, a un nivel de significancia de 0.05.

Solución: Primero ordenamos los datos con el comando `S=datos[order(datos$BPL,datos$GHI,datos$GDP),]`, ejecutamos la siguiente regresión `MR<-lm(GHI BPL+GDP,S)` y ejecutamos `gqtest(MR,fraction = 10)` donde obtenemos un p -valor igual a $p = 0.01691$ como $0.05 > p$ la hipótesis nula se rechaza por lo que existe heteroscedasticidad.

□

- d) Indica si existe o no heteroscedasticidad usando la prueba de Breusch-Pagan-Godfrey, a un nivel de significancia de 0.04.

Solución: Para la prueba de BPG basta con ejecutar el comando `bptest(rl,studentize = FALSE)` con lo que obtuvimos un p -valor igual a $p = 0.1439$ y por lo tanto como $p > 0.04$ se acepta la hipótesis nula y se concluye que no existe heteroscedasticidad.

□

- e) Indica si existe o no heteroscedasticidad usando la prueba de White, a un nivel de significancia de 0.04.

Solución: Para la prueba de White ejecutamos el siguiente comando `white_lm(r1,interactions=TRUE)` y obtuvimos que el p -valor es igual a $p = 0.114$ y por lo tanto como $p > 0.04$ se acepta la hipótesis nula y se concluye que no existe heteroscedasticidad.

□

- f) Asume que existe heteroscedasticidad. Encuentra el valor esperado del Índice Global de Hambre para un país donde el 3.85 % de la población vive por debajo de la línea de la pobreza con un GDP-PPA de 13.15, ya habiendo removido la heteroscedasticidad y asumiendo que su forma funcional es $\sigma_i^2 = 3BPL_i + GDP_i$.

Solución: Escribimos la forma funcional de la varianza con `FFV=3*datos$BPL+datos$GDP` y los pesos `W=1/FFV` para ejecutar la regresión con los pesos `r11<-lm(GHI ~ BPL+GDP,weights = W,data = datos)` y entonces el valor esperado del Índice Global de Hambre para un país donde el 3.85 % de la población vive por debajo de la línea de la pobreza con un GDP-PPA de 13.15 lo encontramos con `predict(r11, newdata=data.frame(BPL=3.85,GDP=13.15))` donde obtenemos

$$y = 13.05248$$

□

- g) Asume que existe heteroscedasticidad. Muestra el estimador S_0 de White usado para estimar la matriz varianza-covarianza.

Solución: El estimador S_0 lo obtenemos con el siguiente comando `vcovHC(r1,type='HC')` que es parte de la librería `sandwich` y por lo tanto obtenemos:

$$S_0 = \begin{pmatrix} 3.19780196 & -0.064995054 & -0.175688998 \\ -0.06499505 & 0.002081450 & 0.003286048 \\ -0.17568900 & 0.003286048 & 0.011576947 \end{pmatrix}$$

□

- h) Indica si existe o no autocorrelación usando la prueba de las rachas, a un nivel de significancia de 0.035.

Solución: Volveremos a usar el modelo MR que usamos anteriormente, el cual era una regresión lineal pero consideraba los valores BPL ordenados entonces ejecutando el comando `runs.test(MR$residuals,threshold = 0)` obtuvimos un p -valor igual a $p = 0.5366$ por lo tanto a un nivel de significancia de 0.035 tenemos que $p > 0.035$, y se acepta la hipótesis nula y existe autocorrelación en la variable BPL.

□

- i) Indica si existe o no autocorrelación usando la prueba de Durbin - Watson, a un nivel de significancia de 0.01.

Solución: Siguiendo con el modelo MR que es un modelo de regresión con los valores BPL ordenados ejecutamos la prueba de D-W con `dwtest(MR)` con lo que obtenemos $DW = 1.976$ por lo tanto como $k = 2$ y $n = 93$

$$dL = 1.489 \text{ y } dU = 1.573 \Rightarrow 4 - dL = 2.511 \text{ y } 4 - dU = 2.427$$

por lo tanto como $dU < DW < 4 - dU$ tenemos que no hay autocorrelación.

□

- j) Asume que existe autocorrelación. Encuentra el valor esperado del Índice Global de Hambre para un país donde el 4.45 % de la población vive por debajo de la línea de la pobreza con un GDP-PPA de 12.54, ya habiendo removido la autocorrelación.

Solución: Asumiendo que hay autocorrelación en los datos originales, ejecutamos `dwtest(r1)` para obtener que $d = 2.2631$ y creamos los datos R de la siguiente manera

```
R<-data.frame()

for (i in 1:93) {
  for (j in 1:3) {
    R[i,j]<-datos[i+1,j+1]-(1-2.2631/2)*datos[i,j+1]
  }
}
```

para ejecutar la regresión sin autocorrelación `MRA<-lm(V1 ~V2+V3,R)` y finalmente el valor esperado del Índice Global de Hambre para un país donde el 4.45 % de la población vive por debajo de la línea de la pobreza con un GDP-PPA de 12.54 lo obtenemos con `predict(MRA, newdata=data.frame(V2=4.45,V3=12.54))` lo cual arroja

$$y = 17.21917$$

□