

Análisis de variables alternativas de crecimiento mediante un modelo de regresión lineal múltiple

Diego Alonso Garcia (197551) María Fernanda Claudio González (191725)
Xaymara Anahí Pastrana Ocampo (195393)

2023-11-25

Introducción

El crecimiento económico, entendido como el incremento de la producción de bienes y servicios en una economía de un periodo a otro, es una de las incógnitas más estudiadas por los economistas. No es trivial el interés, cuando una economía crece, sus ciudadanos pueden gozar de más empleo, consumo y en general mayor calidad de vida. Así, procurar el crecimiento económico de un país es una tarea fundamental de cada uno de los gobiernos y demás agentes económicos. Por lo anterior nos preguntamos: ¿cuáles son los motores del crecimiento?

En principio, el crecimiento económico, entendido como el aumento en la producción agregada, se explicaría teóricamente por el cambio de aquellas variables que componen al PIB. En un simple Modelo Keynesiano, estas variables son consumo, inversión, gasto público, importaciones y exportaciones. No obstante, por lo observado en la coyuntura actual del mundo, creemos que existen variables alternas que podrían influenciar el crecimiento económico de un país. Por lo anterior, **el objetivo de este proyecto es identificar aquellas variables alternas que influyen en el crecimiento económico de un país.**

Dicho esto, las variables que consideramos en nuestro modelo son crecimiento de la población, tasa de fertilidad en adolescentes, tasa de fertilidad, inflación, esperanza de vida, tasa de mortalidad en niños menores a 5 años, migración neta, suscripciones de teléfono celular y crecimiento de la población urbana. Estudiamos cada una de estas variables de manera gráfica y numérica. Posteriormente, realizamos un modelo de regresión lineal múltiple.

Ahondemos un poco en nuestra base y en la selección de las variables antes de pasar a la definición de nuestro modelo.

Conjunto de Datos

La base de datos inicial era de panel, pues constenía observaciones que iban desde el año 2000 hasta el 2021 para los diferentes países. Sin embargo, decidimos elaborar el proyecto usando el enfoque del corte transversal y fijar un solo año, el cual se trata del 2021. De este modo, gozamos de 202 observaciones, las cuales corresponden a cada uno de los países. Utilizamos la siguiente fuente para obtener los datos.

1. Baladram,S. (2023). “Global Socio -Economic & Demographic Insights”. *kaggle*. Consultado por última vez el día 29 de noviembre de 2023 de: <https://www.kaggle.com/datasets/samybaladram/databank-world-development-indicators/data>

Diccionario de variables

La variable dependiente para la estimación del modelo es el crecimiento del PIB del 2021. Las siguientes variables son aquellas que nos permiten construir la regresión.

1. `GDPGrowth`. es el crecimiento del PIB. Esta es nuestra variable dependiente.
2. `PopGrowth`. es el crecimiento de la población, la cual es la primera regresora.
3. `AdolFertRate` es la tasa de fertilidad adolescente, que es la segunda regresora.
4. `FertRate` es la tasa de fertilidad. Esta es la tercera regresora.
5. `InflConsPric`. es la inflación de los precios del consumidor en porcentaje, la cual es la cuarta regresora.
6. `LifeExpBirth` es la esperanza de vida al nacer. Se trata de la quinta regresora.
7. `MobileSubs.100` son las suscripciones a telefonía por cada 100 personas
8. `MortRateU5` es la tasa de mortalidad de niños menores a cinco años, la cual es séptima regresora.

9. **NetMigr** es la migración neta (diferencia entre número de inmigrantes y emigrantes) que es la octava regresora.
10. **UrbanPopGrowth**. es la tasa de crecimiento de población urbana en porcentaje Esta es la novena regresora

Una explicación más a profundidad de las regresoras (aunque optativa) se encuentra a continuación:

1. El crecimiento de la población es el “incremento en el número de individuos pertenecientes a un determinado ámbito en un año concreto por cada 1000 habitantes” (Instituto Nacional de Estadística de Madrid, 2023). La relación es directa.
2. La tasa de fertilidad de adolescentes (TFA) es aquella que “corresponde al número de nacimientos por cada 1000 mujeres de 15 a 19 años de edad en un año determinado” (Guerrero, 2020). La relación es negativa.
3. La tasa de fertilidad es la “frecuencia con la que un nacimiento vivo completa el proceso reproductivo” (Vinacur et al., 2001, p. 3).
4. La inflación es “es un fenómeno que se observa en la economía de un país y está relacionado con el aumento desordenado de los precios de la mayor parte de los bienes y servicios que se comercian en sus mercados, por un periodo de tiempo prolongado” (BANXICO, 2023).
5. La esperanza de vida al nacer “se refiere a la cantidad de años que viviría un recién nacido si los patrones de mortalidad vigentes al momento de su nacimiento no cambian a lo largo de la vida del infante” (Banco Mundial, 2013).
6. La tasa de mortalidad de niños menores a 5 años es “es la probabilidad de que un niño nacido en un año o período específico muera antes de alcanzar la edad de 5 años” (CEPAL, 2023) y se expresa por cada 1000 niños nacidos vivos.
7. La migración neta son las “inmigraciones menos las emigraciones realizadas en un año concreto por individuos de un determinado ámbito, por cada 1.000 habitantes de ese ámbito” (Instituto Nacional de Estadística de Madrid, 2023).
8. La tasa de suscripciones a telefonía mide el número de suscripciones a servicios telefónicos por cada 100 personas.
9. La tasa de crecimiento de la población urbana, indica la razón a la cual la población urbana incrementa.

Una vez establecidas las variables, podemos comenzar con nuestro Análisis Descriptivo.

Análisis Descriptivo de los Datos

Lo primero que hacemos es instalar y abrir los paquetes que vamos a ocupar:

```
# Instalar Librerías
options(repos = "https://cran.r-project.org")
install.packages("ggplot2")

## package 'ggplot2' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\diego\AppData\Local\Temp\RtmpUzmf22\downloaded_packages
```

```
install.packages("DataExplorer")
```

```
## package 'DataExplorer' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\diego\AppData\Local\Temp\RtmpUzmf22\downloaded_packages
```

```
install.packages("readxl")
```

```
## package 'readxl' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\diego\AppData\Local\Temp\RtmpUzmf22\downloaded_packages
```

```
install.packages("reshape2")
```

```
## package 'reshape2' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\diego\AppData\Local\Temp\RtmpUzmf22\downloaded_packages
```

```
install.packages("stargazer")
```

```
## package 'stargazer' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\diego\AppData\Local\Temp\RtmpUzmf22\downloaded_packages
```

```
install.packages("car")
```

```
## package 'car' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\diego\AppData\Local\Temp\RtmpUzmf22\downloaded_packages
```

```
install.packages("stats")
install.packages("tseries")
```

```
## package 'tseries' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\diego\AppData\Local\Temp\RtmpUzmf22\downloaded_packages
```

```
install.packages("corrplot")
```

```
## package 'corrplot' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\diego\AppData\Local\Temp\RtmpUzmf22\downloaded_packages
```

```
install.packages("lmtest")
```

```
## package 'lmtest' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\diego\AppData\Local\Temp\RtmpUzmf22\downloaded_packages
```

```
install.packages("skedastic")
```

```
## package 'skedastic' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\diego\AppData\Local\Temp\RtmpUzmf22\downloaded_packages
```

```
install.packages("sandwich")
```

```
## package 'sandwich' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\diego\AppData\Local\Temp\RtmpUzmf22\downloaded_packages
```

```
install.packages("tinytex")
```

```
## package 'tinytex' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\diego\AppData\Local\Temp\RtmpUzmf22\downloaded_packages
```

```
# Activar Librerías
library(DataExplorer)
library(ggplot2)
library(readxl)
library(reshape2)
library(stargazer)
library(car)
library(stats)
library(tseries)
library(corrplot)
library(lmtest)
library(skedastic)
library(sandwich)
library(tinytex)
```

Luego hacemos la lectura de los datos

```
archivoCrecimiento<- "C:\\Users\\diego\\Documents\\Econometría\\ProyectoEconometria\\world_development_d
datosCrecimiento<-read.csv(archivoCrecimiento)
print(head(datosCrecimiento,n=3))
```

```
##   Year      Country Region      SubRegion SurfAreaSqKm PopTotal      PopDens
## 1 2000 Afghanistan Asia      Southern Asia      652860 19542982 29.96333
## 2 2000      Malta Europe Southern Europe      320 390087 1219.02187
## 3 2000    Belgium Europe Western Europe      30530 10251250 338.54855
##   PopGrowth.      GDP GDPGrowth. AdolFertRate AgriValAdd.GDP Exports.GDP
## 1 1.4438030 18012478281 -5.206288 152.572 33.096776 13.31525
## 2 0.6452672 4323339212 19.681791 19.869 1.871997 120.24748
## 3 0.2425180 237000000000 3.716679 11.915 1.176005 72.54740
##   FertRate  FDI NetBoP GNI.CapAtlas      GNIAtlas Imports.GDP IndValAdd.GDP
## 1 7.534 170000 916.9233 17786690838 41.31263 17.17877
## 2 1.680 743185325 10950.0000 4273279544 128.37458 26.93999
## 3 1.670 88738714000 25890.0000 265000000000 69.68274 24.94383
##   InflConsPric. LifeExpBirth MerchTrade.GDP MobileSubs.100 MortRateU5 NetMigr
## 1 37.611028 55.29800 52.77705 0.00000 129.3 -1007135
## 2 -4.512396 78.34878 135.68216 28.66747 7.6 1799
## 3 2.014617 77.72195 154.51590 54.84034 5.9 32262
##   UrbanPopGrowth.
## 1 1.8613773
## 2 0.9522992
## 3 0.3084313
```

Y también podemos observar el tipo de variable de cada una de ellas:

```
tipos_de_variables <- sapply(datosCrecimiento, class)
print(tipos_de_variables)
```

```
##           Year      Country      Region      SubRegion      SurfAreaSqKm
##      "integer" "character" "character" "character"      "numeric"
##      PopTotal      PopDens      PopGrowth.      GDP      GDPGrowth.
##      "integer"      "numeric"      "numeric"      "numeric"      "numeric"
##      AdolFertRate AgriValAdd.GDP Exports.GDP      FertRate      FDI NetBoP
##      "numeric"      "numeric"      "numeric"      "numeric"      "numeric"
##      GNI.CapAtlas      GNIAtlas      Imports.GDP      IndValAdd.GDP      InflConsPric.
##      "numeric"      "numeric"      "numeric"      "numeric"      "numeric"
##      LifeExpBirth MerchTrade.GDP MobileSubs.100      MortRateU5      NetMigr
##      "numeric"      "numeric"      "numeric"      "numeric"      "integer"
## UrbanPopGrowth.
##      "numeric"
```

Preparemos los datos que vamos a ocupar:

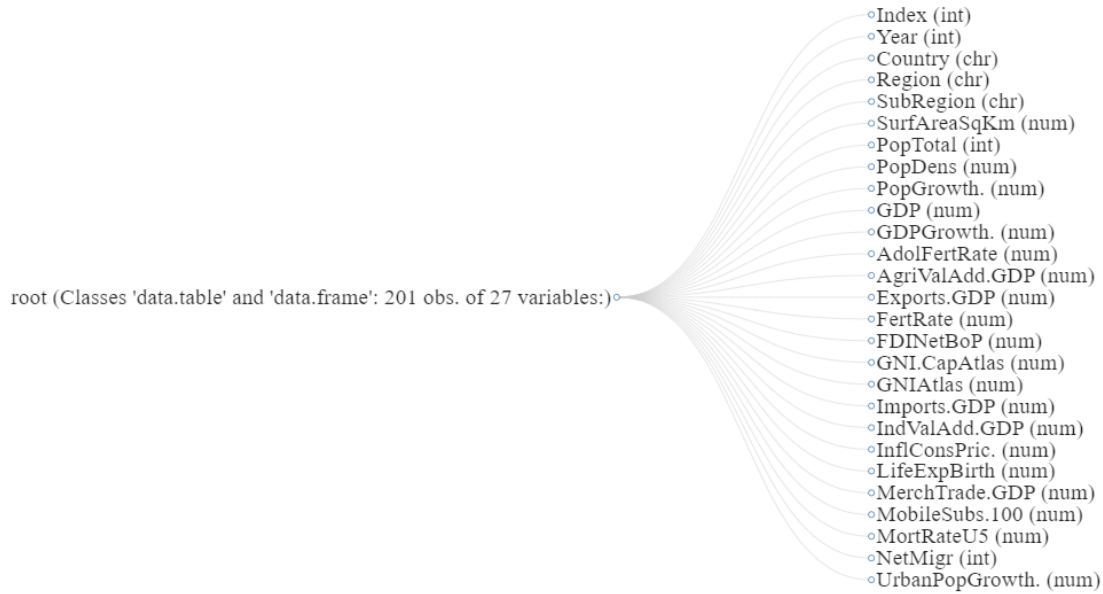
```
datosCrecimiento2021 <- subset(datosCrecimiento, Year == 2021)
datosCrecimiento2021 <- cbind(Index = seq_len(nrow(datosCrecimiento2021)), datosCrecimiento2021)

datosCrecimiento2021 <- datosCrecimiento2021[datosCrecimiento2021$Country != "Cuba", ]

#Excluimos Cuba porque era un super atípico que descomponía todas las pruebas
```

De manera rápida podemos con `DataExplorer()` realizar un reporte de nuestra base de datos.

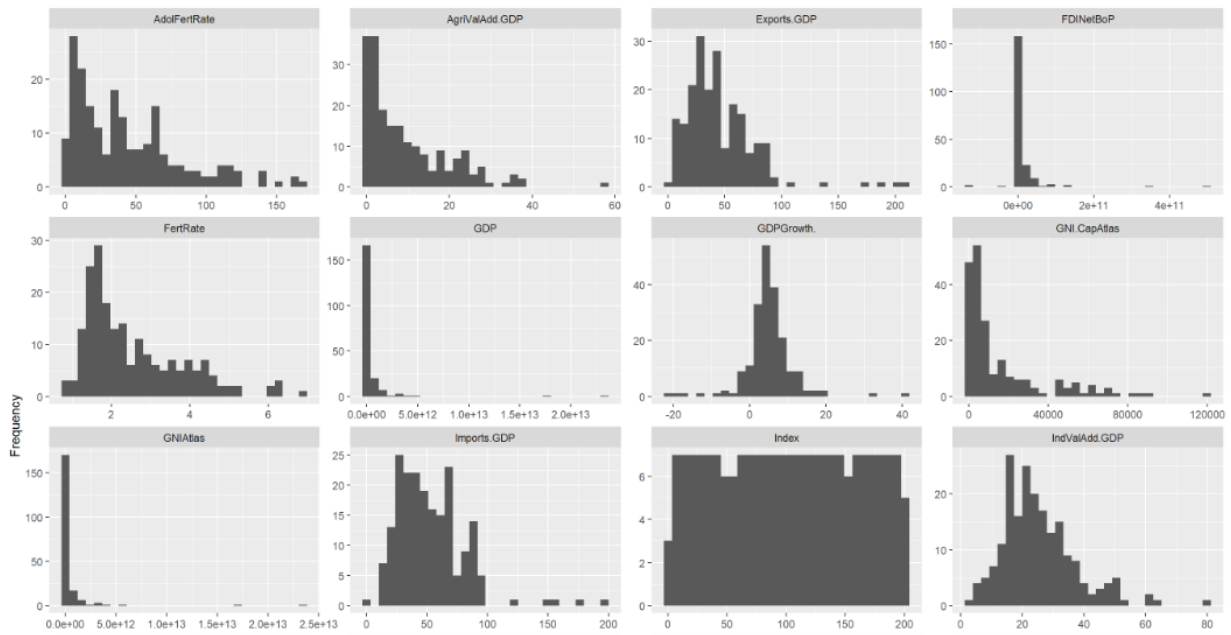
En el reporte generado podemos ver algunas gráficas útiles como un diagrama de entidad-relación.



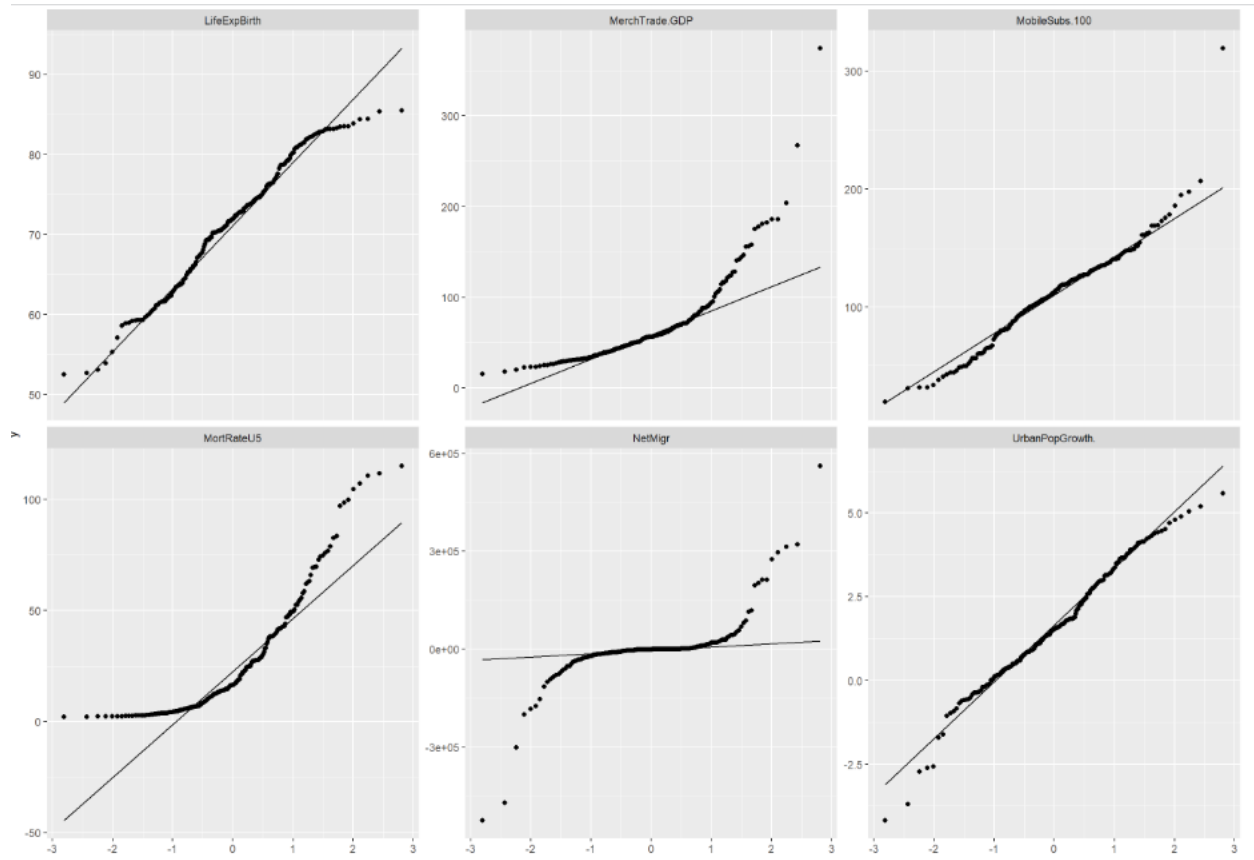
También podemos obtener en el reporte el histograma de las variables. Así podemos esbozar como se distribuyen las variables.

Univariate Distribution

Histogram



Otra de las gráficas que podemos obtener es un QQ-Plot, el cual nos muestra que si la distribución de las variables se asemeja a una normal entonces los puntos de la gráfica se encontrarán agrupados en la línea de 45°.



Pero hagamos la interpretación puntual de cada una de las variables:

```
attach(datosCrecimiento2021)
```

Recordemos las variables del modelo teórico que vamos a ocupar para nuestro modelo:

```
vars<-data.frame(GDPGrowth.,PopGrowth.,AdolFertRate,FertRate,InflConsPric.,LifeExpBirth,MobileSubs.100,
print(summary(vars))
```

```
##      GDPGrowth.      PopGrowth.      AdolFertRate      FertRate
##  Min.   :-20.700   Min.   :-4.2566   Min.    : 1.584   Min.    :0.772
##  1st Qu.: 2.562    1st Qu.: 0.2073   1st Qu.: 11.216   1st Qu.:1.595
##  Median : 4.792    Median : 0.9475   Median : 33.929   Median :2.091
##  Mean   : 4.958    Mean   : 0.9264   Mean    : 43.440   Mean    :2.545
##  3rd Qu.: 7.155    3rd Qu.: 1.8885   3rd Qu.: 64.054   3rd Qu.:3.303
##  Max.    : 41.745   Max.    : 3.7074   Max.    :170.462   Max.    :6.820
##  InflConsPric.    LifeExpBirth  MobileSubs.100    MortRateU5
##  Min.    : -4.178   Min.    :52.52    Min.    : 19.45   Min.    : 2.00
##  1st Qu.: 2.216    1st Qu.:65.82    1st Qu.: 88.09   1st Qu.: 6.40
##  Median : 4.685    Median :71.91    Median :112.44   Median : 16.60
##  Mean    : 10.581   Mean     :71.49    Mean    :110.07   Mean     : 26.42
##  3rd Qu.: 10.045   3rd Qu.:76.46    3rd Qu.:131.86   3rd Qu.: 38.60
##  Max.    :235.515   Max.     :85.49    Max.    :319.43   Max.     :115.20
##      NetMigr      UrbanPopGrowth.
##  Min.    :-525116.0   Min.    :-4.1703
##  1st Qu.: -11253.0   1st Qu.: 0.5033
```



```
## Median :   -742.0   Median : 1.5299
## Mean   :    137.1   Mean   : 1.5840
## 3rd Qu.:   2632.0   3rd Qu.: 2.7889
## Max.   :  561580.0   Max.   : 5.5835
```

Con `summary()` podemos ver las características generales de cada variable. Específicamente, apreciamos el mínimo, el primer cuartil, la mediana, el promedio, el tercer cuartil y el máximo de cada variable.

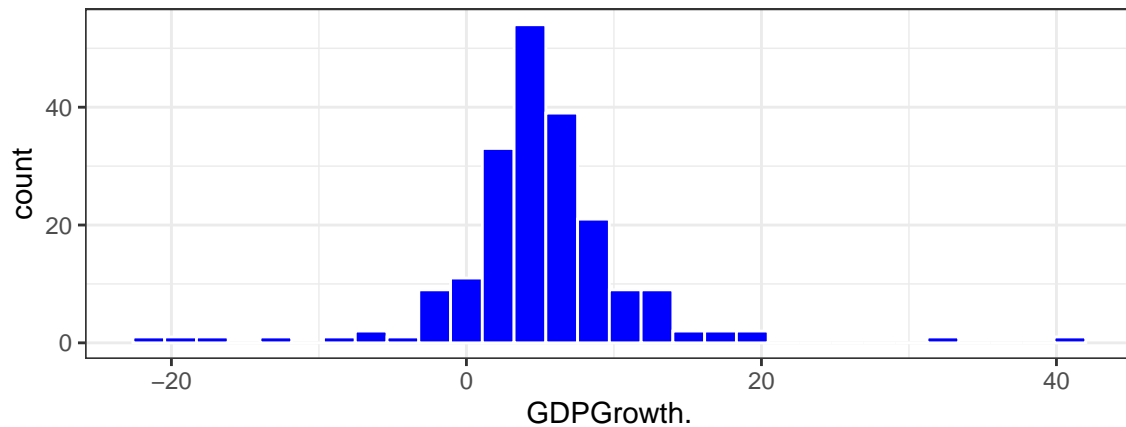
Lo primero que analizaremos es la distribución de las variables a través del histograma de cada una

```
# Histograma
ggplot(data=datosCrecimiento2021,aes(x=GDPGrowth.))+
  geom_histogram(col="white",fill="blue") +
  labs(title="Histograma",
        subtitle = "Variable: GDPGrowth.",
        caption= "Fuente: Elaboración propia")+
  theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Histograma

Variable: GDPGrowth.



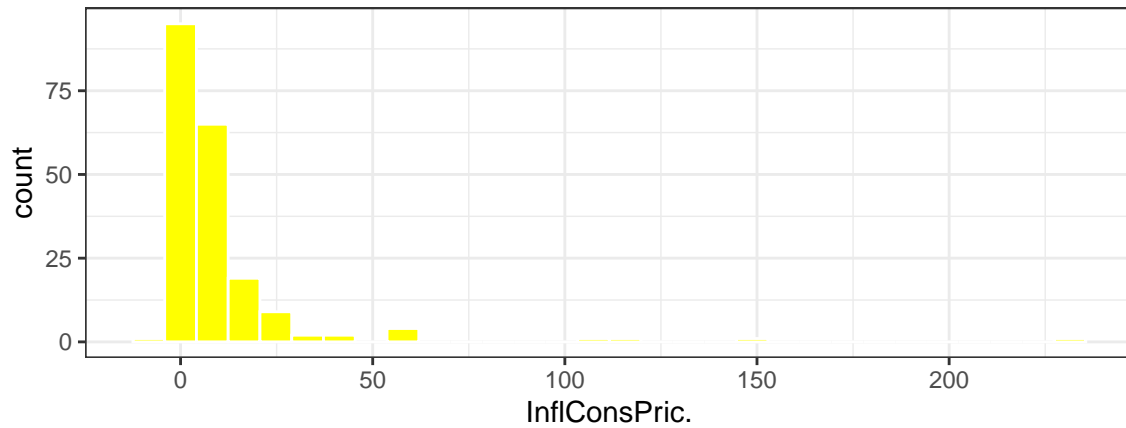
Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=InflConsPric.))+
  geom_histogram(col="white",fill="yellow") +
  labs(title="Histograma",
        subtitle = "Variable: InflConsPric.",
        caption= "Fuente: Elaboración propia")+
  theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Histograma

Variable: InflConsPric.



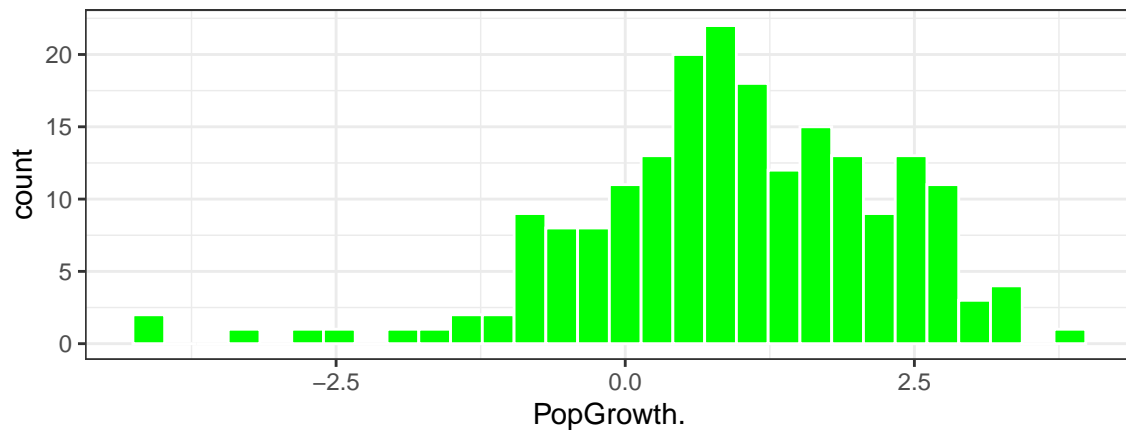
Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=PopGrowth.))+  
  geom_histogram(col="white",fill="green") +  
  labs(title="Histograma",  
        subtitle = "Variable: PopGrowth..",  
        caption= "Fuente: Elaboración propia")+  
  theme_bw()
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Histograma

Variable: PopGrowth..



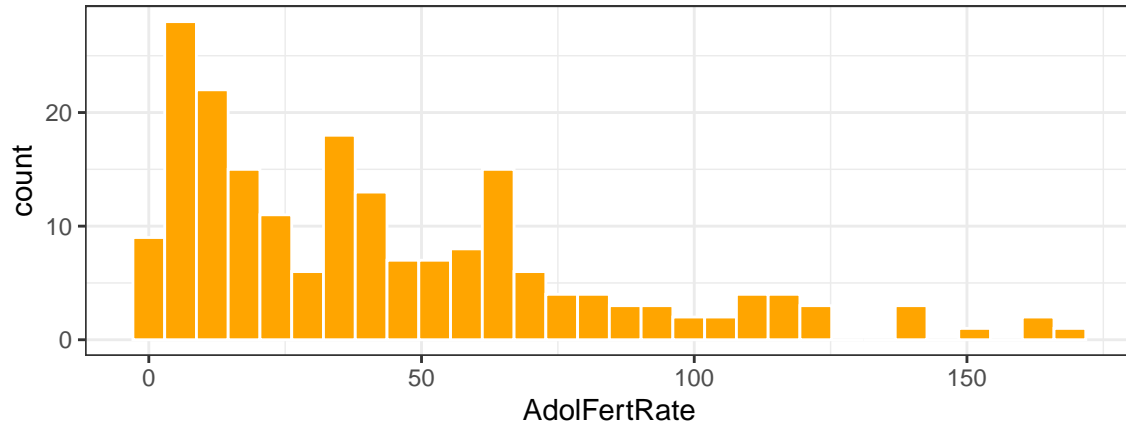
Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=AdolFertRate))+  
  geom_histogram(col="white",fill="orange") +  
  labs(title="Histograma",  
        subtitle = "Variable: AdolFertRate",  
        caption= "Fuente: Elaboración propia")+  
  theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Histograma

Variable: AdolFertRate



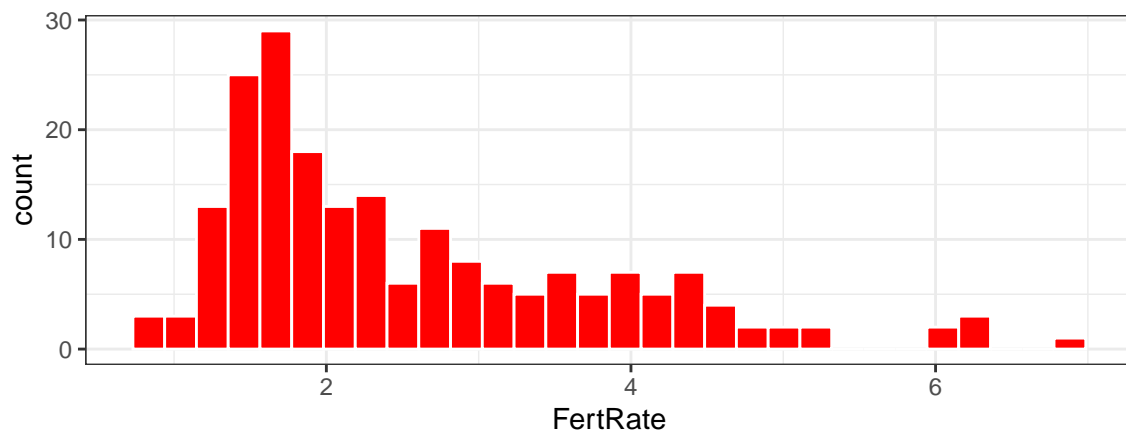
Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=FertRate))+  
  geom_histogram(col="white",fill="red") +  
  labs(title="Histograma",  
        subtitle = "Variable: FertRate",  
        caption= "Fuente: Elaboración propia")+  
  theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Histograma

Variable: FertRate



Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=LifeExpBirth))+  
  geom_histogram(col="white",fill="pink") +  
  labs(title="Histograma",
```

```

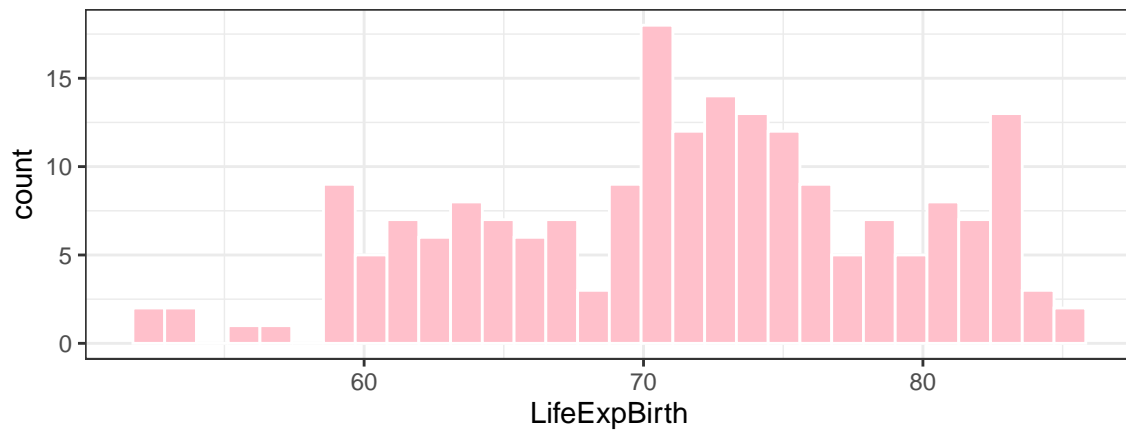
    subtitle = "Variable: LifeExpBirth",
    caption= "Fuente: Elaboración propia")+
  theme_bw()

```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Histograma

Variable: LifeExpBirth



Fuente: Elaboración propia

```

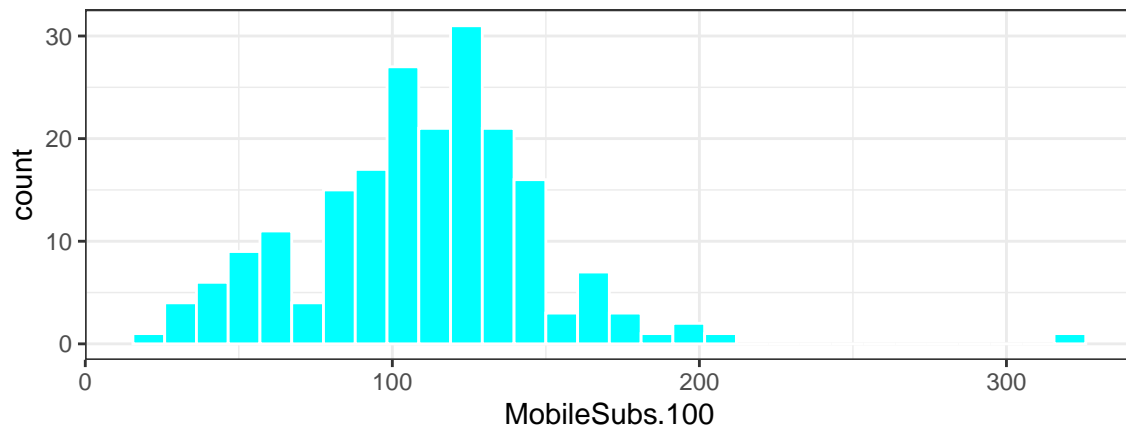
ggplot(data=datosCrecimiento2021,aes(x=MobileSubs.100))+
  geom_histogram(col="white",fill="cyan") +
  labs(title="Histograma",
       subtitle = "Variable: MobileSubs.100",
       caption= "Fuente: Elaboración propia")+
  theme_bw()

```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Histograma

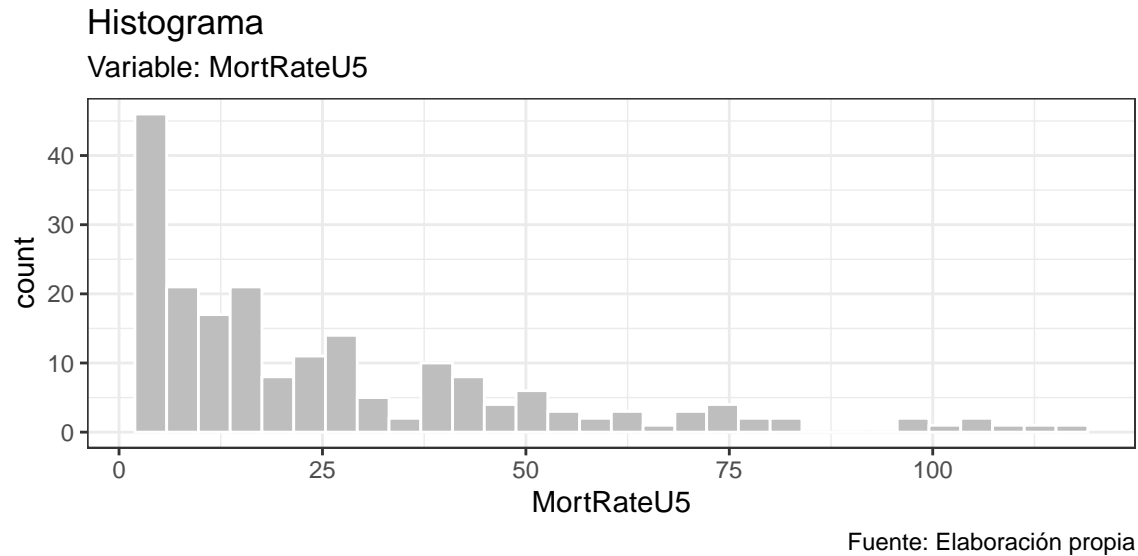
Variable: MobileSubs.100



Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=MortRateU5))+
  geom_histogram(col="white",fill="grey") +
  labs(title="Histograma",
        subtitle = "Variable: MortRateU5",
        caption= "Fuente: Elaboración propia")+
  theme_bw()
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

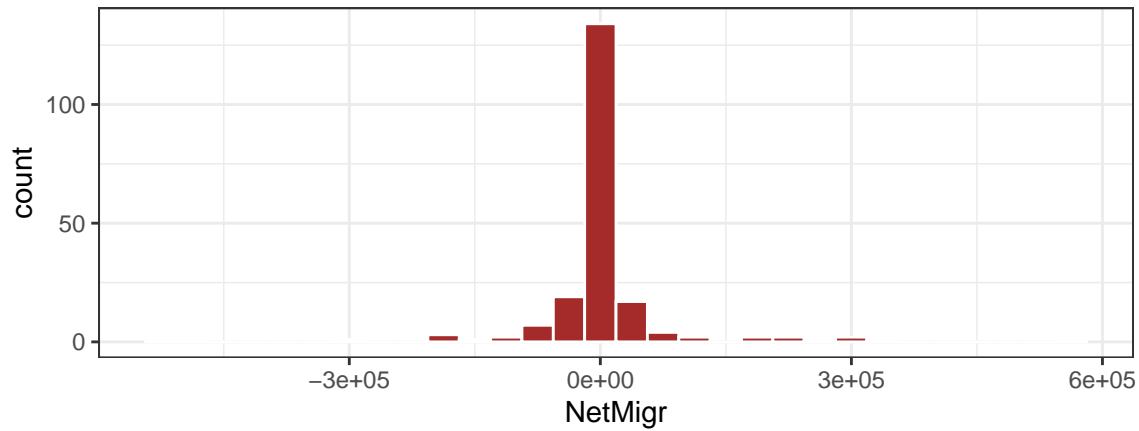


```
ggplot(data=datosCrecimiento2021,aes(x=NetMigr))+
  geom_histogram(col="white",fill="brown") +
  labs(title="Histograma",
        subtitle = "Variable: NetMigr",
        caption= "Fuente: Elaboración propia")+
  theme_bw()
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Histograma

Variable: NetMigr



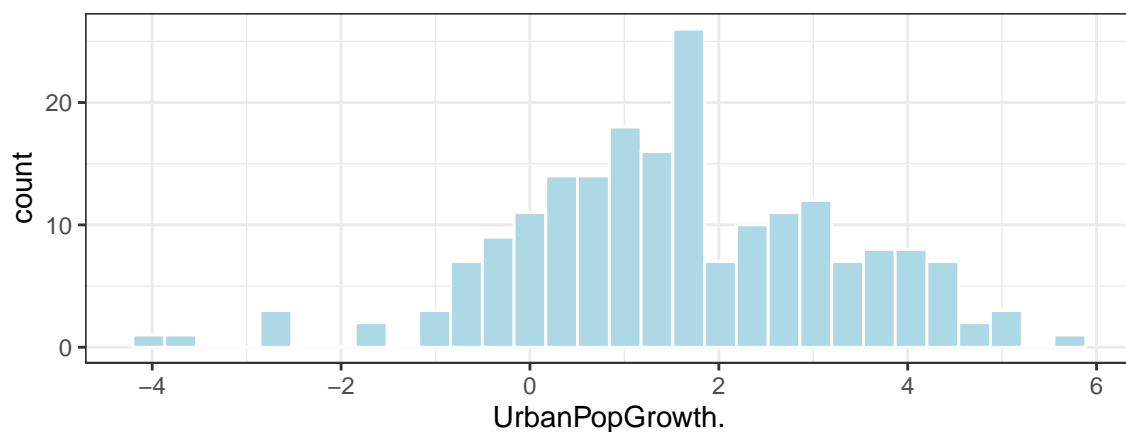
Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=UrbanPopGrowth.))+  
  geom_histogram(col="white",fill="lightblue") +  
  labs(title="Histograma",  
        subtitle = "Variable: UrbanPopGrowth.",  
        caption= "Fuente: Elaboración propia")+  
  theme_bw()
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Histograma

Variable: UrbanPopGrowth.



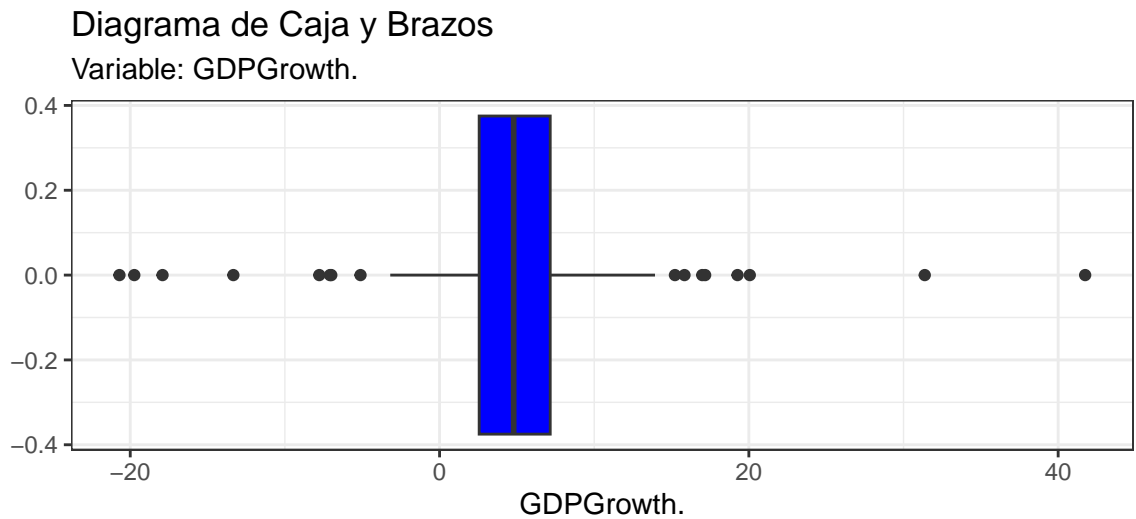
Fuente: Elaboración propia

El histograma que presenta una distribución más parecida a la normal es la variable explicada. Sin embargo, el crecimiento del PIB presenta más datos atípicos. Los gráficos de la inflación, la tasa de fertilidad adolescente, la tasa de fertilidad y, especialmente, la tasa de mortalidad infantil de menores a cinco años son aquellos que gozan de un sesgo a la derecha.

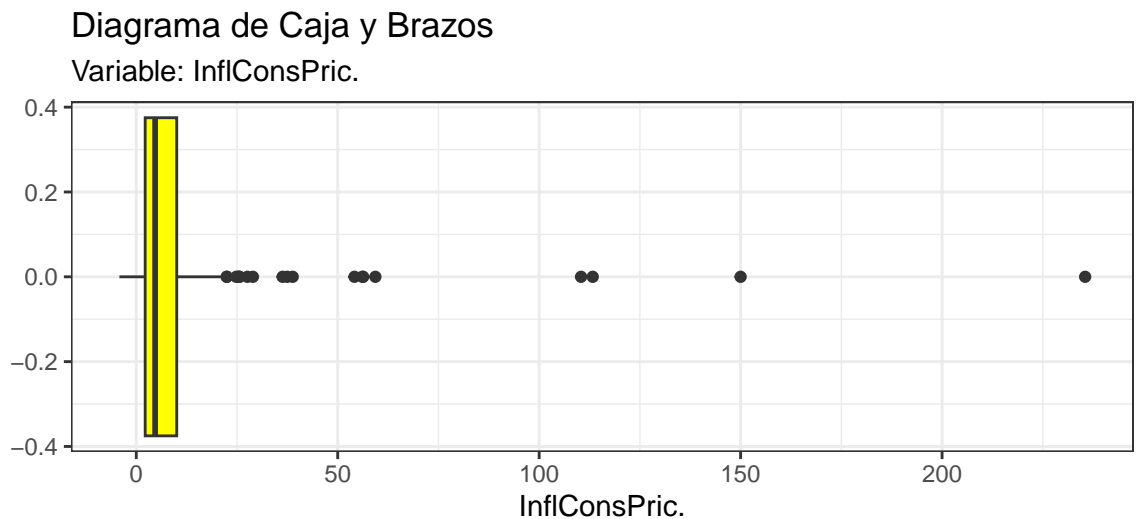
Por otro lado, el crecimiento poblacional, la esperanza de vida al nacer y el crecimiento de la población urbana presentan un sesgo a la izquierda. No obstante, estas tres variables tienen el sesgo menos marcado que la inflación. Esta variable y la tasa de mortalidad de niños menores a cinco años son las menos simétricas.

¿Por qué no vemos su dispersión con un diagrama de caja y brazos?

```
ggplot(data=datosCrecimiento2021,aes(x=GDPGrowth.))+  
  geom_boxplot(fill="blue") +  
  labs(title="Diagrama de Caja y Brazos",  
        subtitle = "Variable: GDPGrowth.",  
        caption= "Fuente: Elaboración propia")+  
  theme_bw()
```



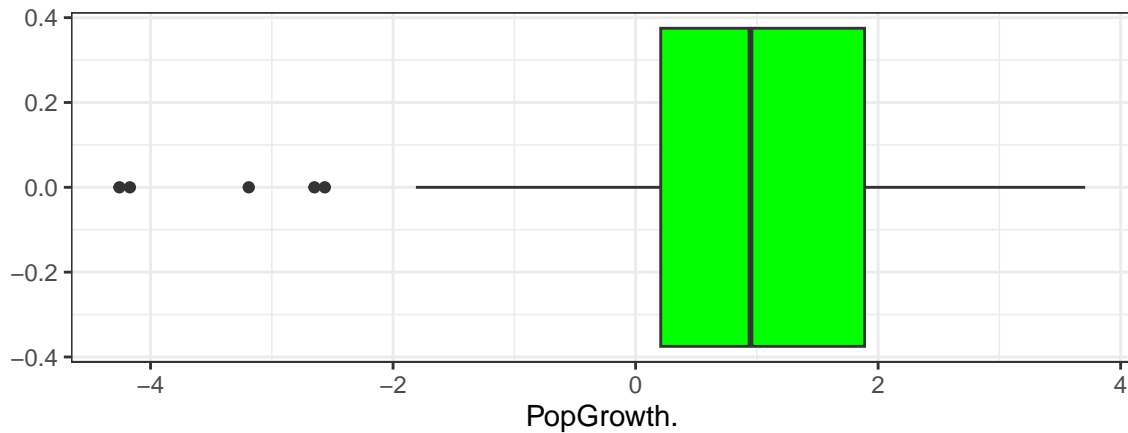
```
ggplot(data=datosCrecimiento2021,aes(x=InflConsPric.))+  
  geom_boxplot(fill="yellow") +  
  labs(title="Diagrama de Caja y Brazos",  
        subtitle = "Variable: InflConsPric.",  
        caption= "Fuente: Elaboración propia")+  
  theme_bw()
```



```
ggplot(data=datosCrecimiento2021,aes(x=PopGrowth.))+
  geom_boxplot(fill="green") +
  labs(title="Diagrama de Caja y Brazos",
        subtitle = "Variable: PopGrowth..",
        caption= "Fuente: Elaboración propia")+
  theme_bw()
```

Diagrama de Caja y Brazos

Variable: PopGrowth..

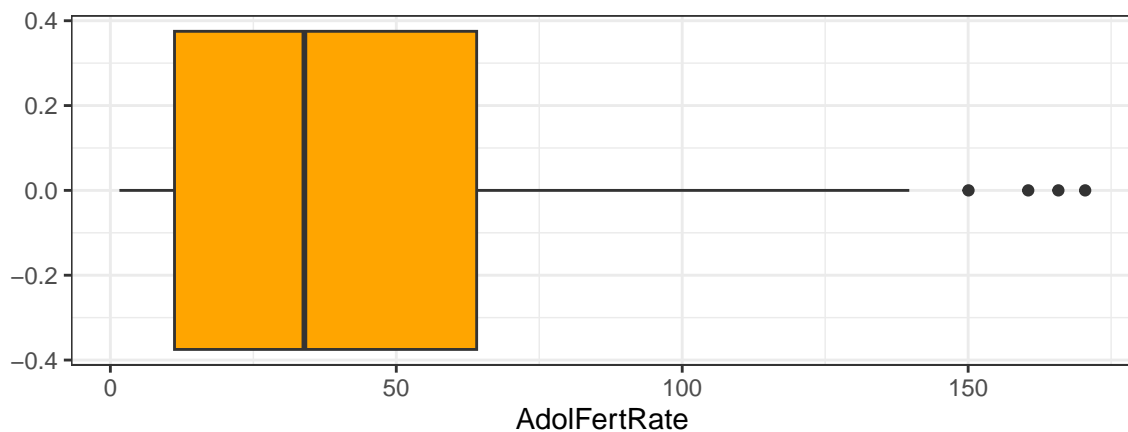


Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=AdolFertRate))+
  geom_boxplot(fill="orange") +
  labs(title="Diagrama de Caja y Brazos",
        subtitle = "Variable: AdolFertRate",
        caption= "Fuente: Elaboración propia")+
  theme_bw()
```

Diagrama de Caja y Brazos

Variable: AdolFertRate

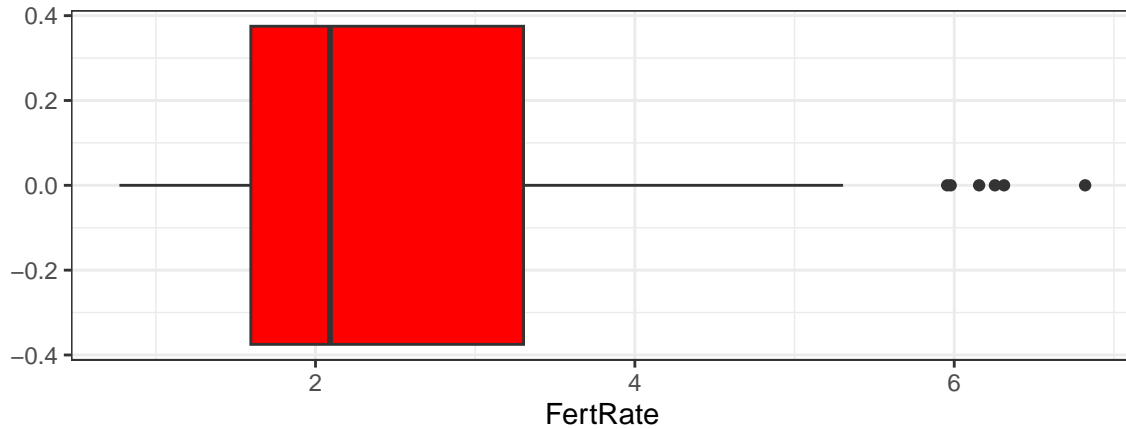


Fuente: Elaboración propia


```
ggplot(data=datosCrecimiento2021,aes(x=FertRate))+
  geom_boxplot(fill="red") +
  labs(title="Diagrama de Caja y Brazos",
        subtitle = "Variable: FertRate",
        caption= "Fuente: Elaboración propia")+
  theme_bw()
```

Diagrama de Caja y Brazos

Variable: FertRate

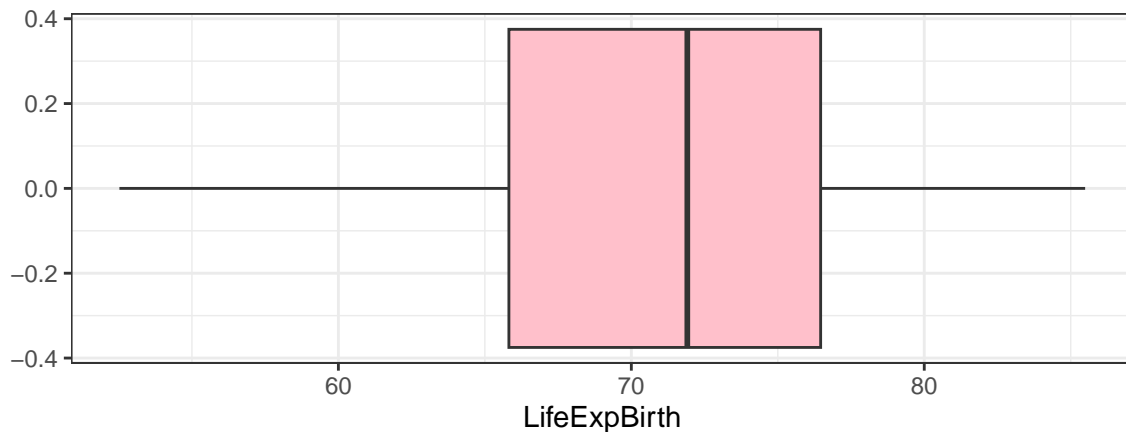


Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=LifeExpBirth))+
  geom_boxplot(fill="pink") +
  labs(title="Diagrama de Caja y Brazos",
        subtitle = "Variable: LifeExpBirth",
        caption= "Fuente: Elaboración propia")+
  theme_bw()
```

Diagrama de Caja y Brazos

Variable: LifeExpBirth

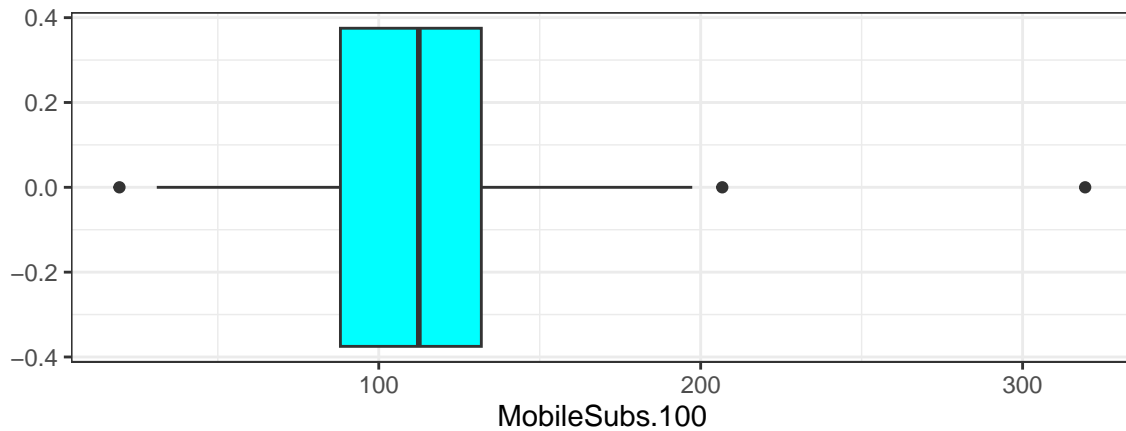


Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=MobileSubs.100))+
  geom_boxplot(fill="cyan") +
  labs(title="Diagrama de Caja y Brazos",
        subtitle = "Variable: MobileSubs.100",
        caption= "Fuente: Elaboración propia")+
  theme_bw()
```

Diagrama de Caja y Brazos

Variable: MobileSubs.100

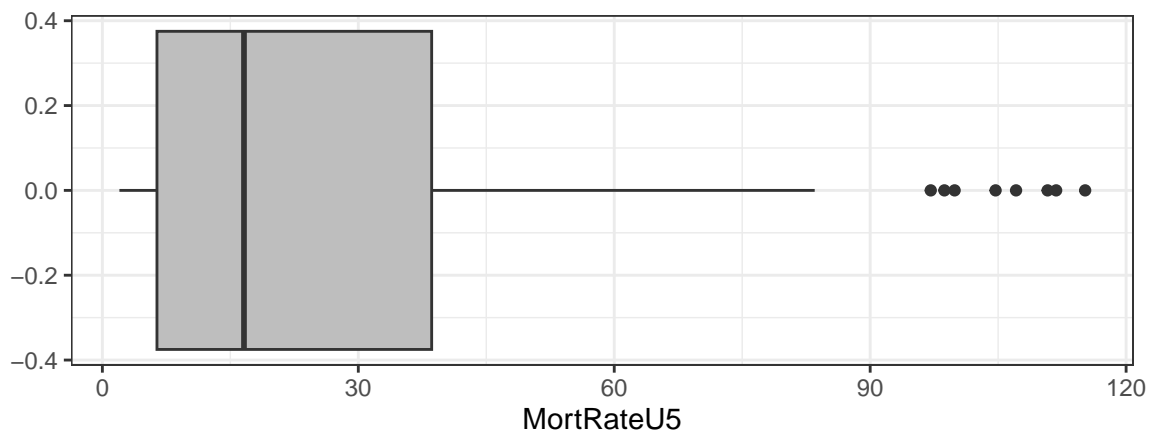


Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=MortRateU5))+
  geom_boxplot(fill="grey") +
  labs(title="Diagrama de Caja y Brazos",
        subtitle = "Variable: MortRateU5",
        caption= "Fuente: Elaboración propia")+
  theme_bw()
```

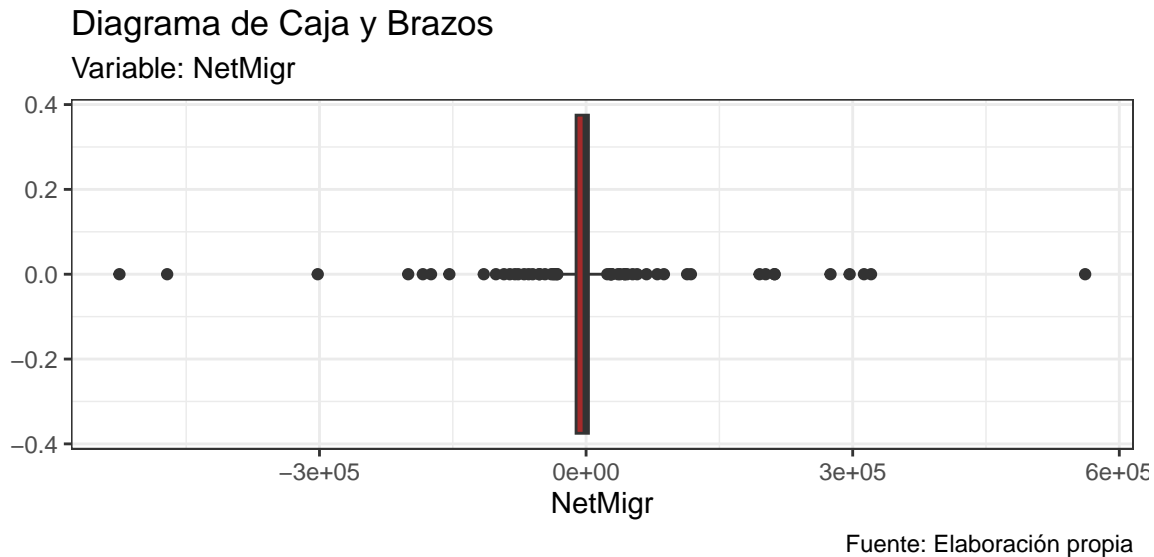
Diagrama de Caja y Brazos

Variable: MortRateU5

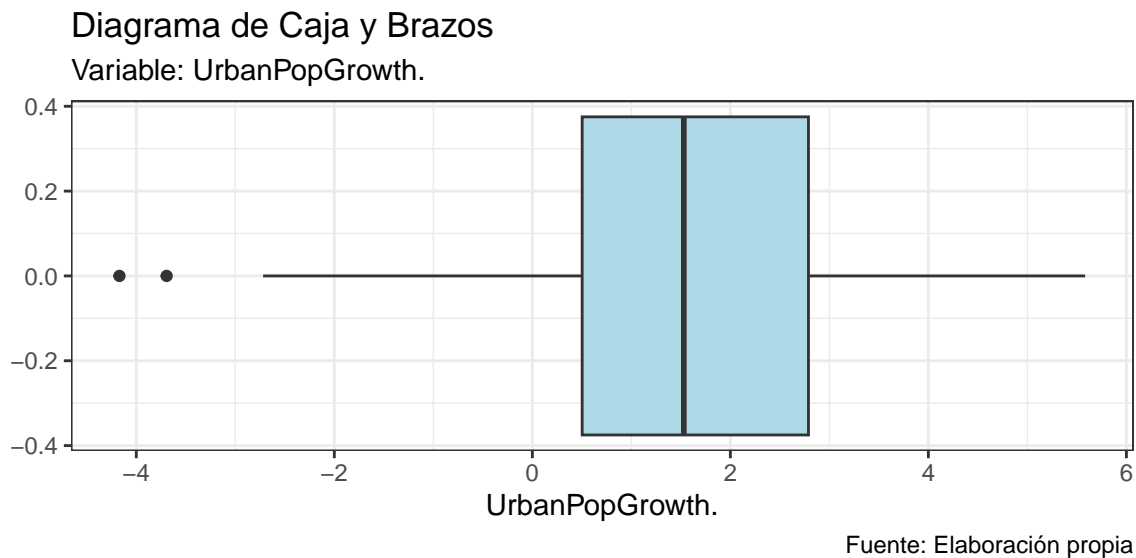


Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=NetMigr))+
  geom_boxplot(fill="brown") +
  labs(title="Diagrama de Caja y Brazos",
        subtitle = "Variable: NetMigr",
        caption= "Fuente: Elaboración propia")+
  theme_bw()
```



```
ggplot(data=datosCrecimiento2021,aes(x=UrbanPopGrowth.))+
  geom_boxplot(fill="lightblue") +
  labs(title="Diagrama de Caja y Brazos",
        subtitle = "Variable: UrbanPopGrowth.",
        caption= "Fuente: Elaboración propia")+
  theme_bw()
```



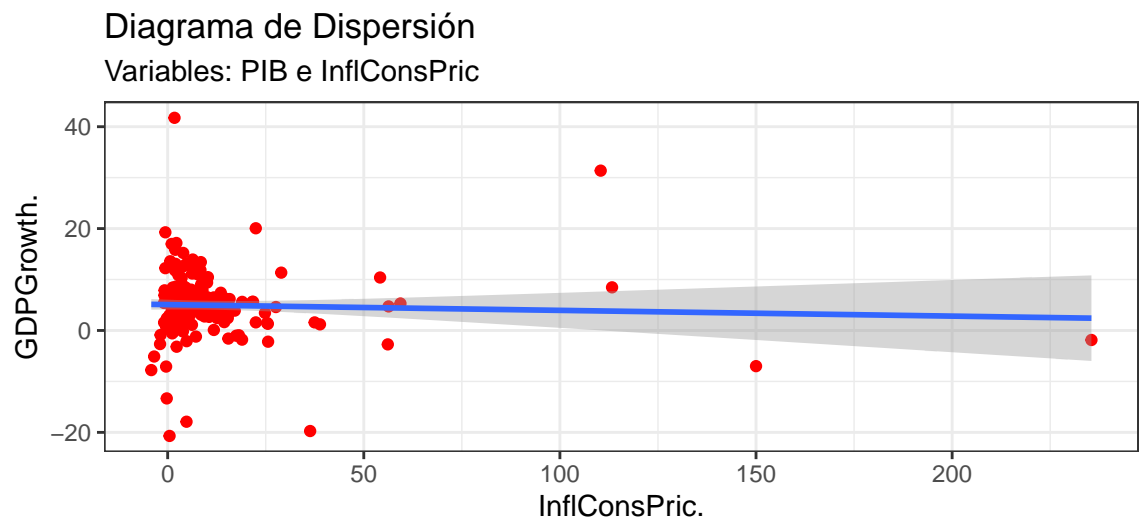
A simple vista, estos diagramas se comportan de manera muy similar a los histogramas. El crecimiento económico y la migración neta son variables que pareciera siguen una distribución normal; sin embargo,

ambas variables presentan muchos datos atípicos. La inflación, la tasa de fertilidad en adolescentes, la tasa de fertilidad y la tasa de mortalidad en niños menores a cinco años tienen un sesgo a la izquierda. Por otro lado, el crecimiento de la población, la esperanza de vida y el crecimiento de la población urbana presentan sesgos a la derecha. En general, los diagramas de cajas y brazos muestran que la distribución de la mayoría de las variables no son simétricas.

```
# Diagrama de dispersión
ggplot(data=datosCrecimiento2021,aes(x=InflConsPric.,y=GDPGrowth.))+
  geom_point(col="red",se=TRUE) +
  geom_smooth(method = "lm")+
  labs(title="Diagrama de Dispersión",
        subtitle = "Variables: PIB e InflConsPric",
        caption="Fuente: Elaboración propia")+
  theme_bw()
```

```
## Warning in geom_point(col = "red", se = TRUE): Ignoring unknown parameters:
## `se`
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

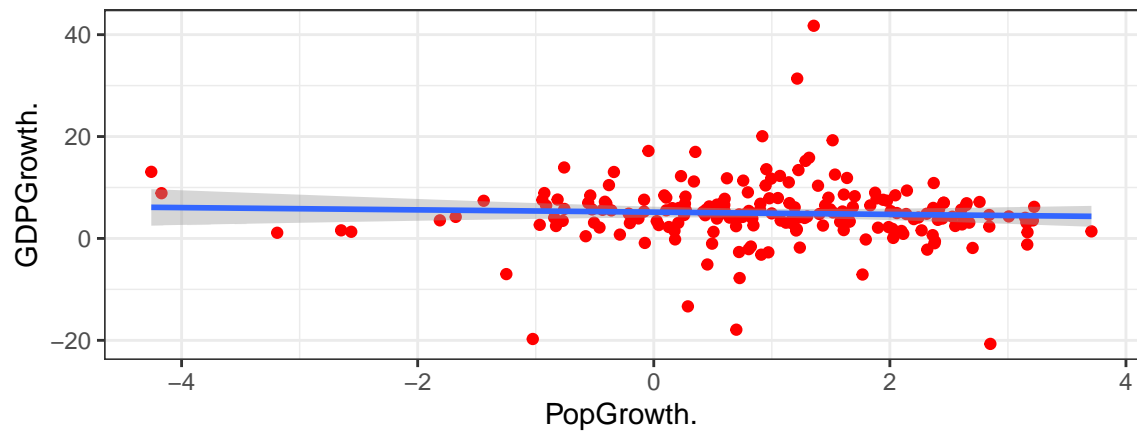


```
ggplot(data=datosCrecimiento2021,aes(x=PopGrowth.,y=GDPGrowth.))+
  geom_point(col="red") +
  geom_smooth(method = "lm",se=TRUE)+
  labs(title="Diagrama de Dispersión",
        subtitle = "Variables: PIB y x=PopGrowth.",
        caption="Fuente: Elaboración propia")+
  theme_bw()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Diagrama de Dispersión

Variables: PIB y x=PopGrowth.



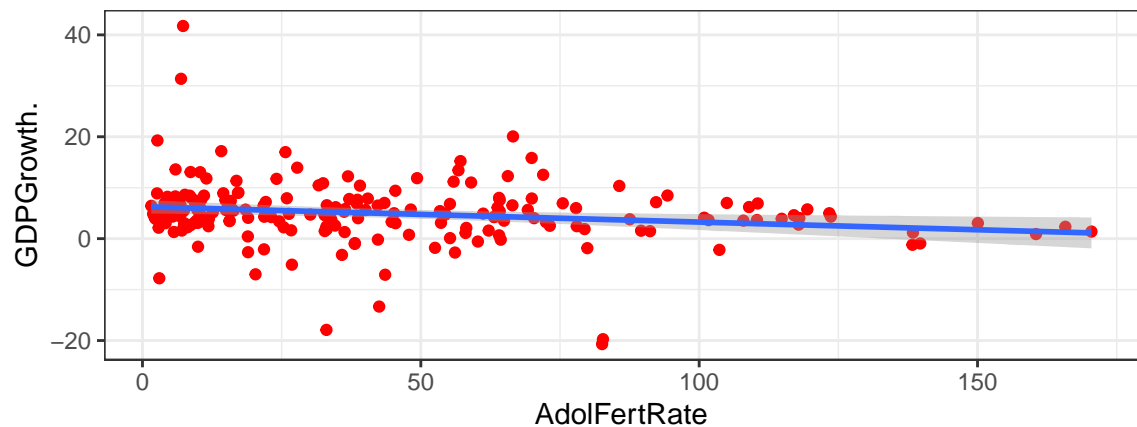
Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=AdolFertRate,y=GDPGrowth.))+  
  geom_point(col="red") +  
  geom_smooth(method = "lm",se=TRUE)+  
  labs(title="Diagrama de Dispersión",  
        subtitle = "Variables: PIB y AdolFertRate",  
        caption="Fuente: Elaboración propia")+  
  theme_bw()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Diagrama de Dispersión

Variables: PIB y AdolFertRate



Fuente: Elaboración propia

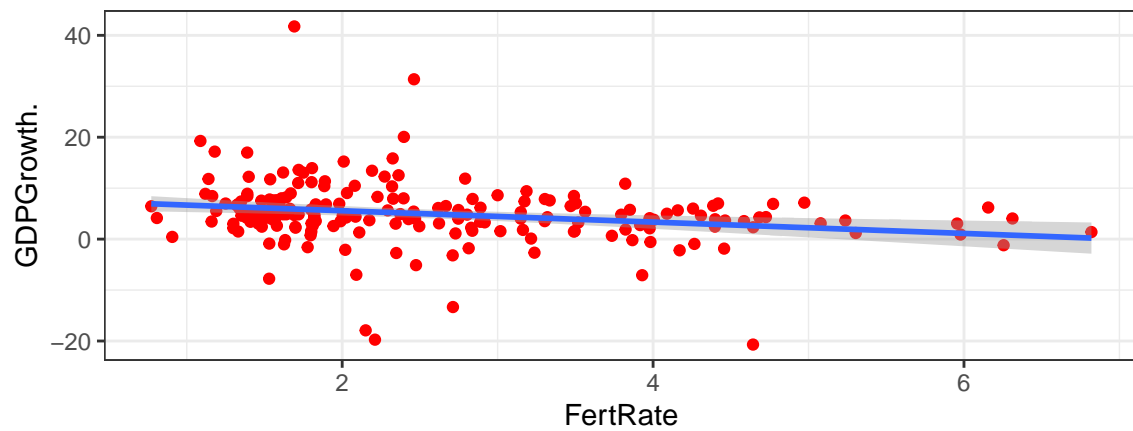
```
ggplot(data=datosCrecimiento2021,aes(x=FertRate,y=GDPGrowth.))+  
  geom_point(col="red") +  
  geom_smooth(method = "lm",se=TRUE)+  
  labs(title="Diagrama de Dispersión",  
        subtitle = "Variables: PIB y FertRate",
```

```
caption="Fuente: Elaboración propia")+
theme_bw()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Diagrama de Dispersión

Variables: PIB y FertRate



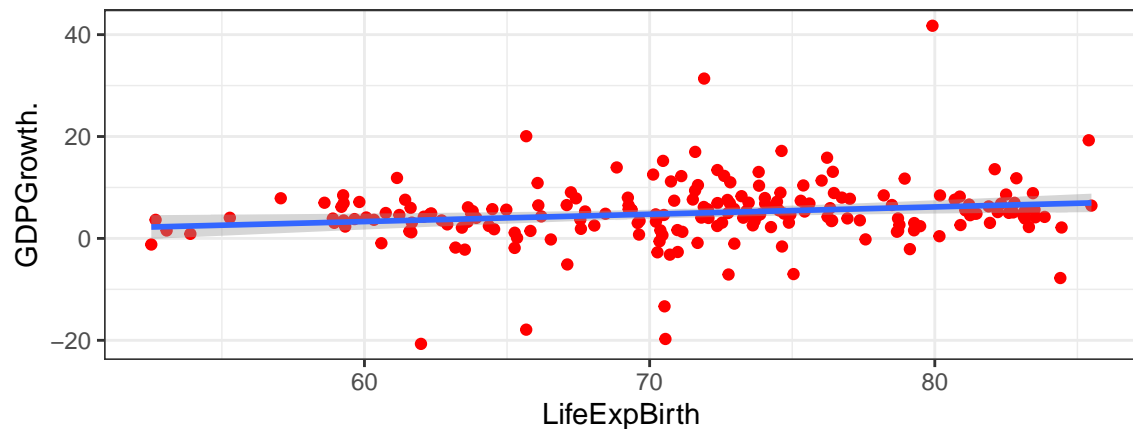
Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=LifeExpBirth,y=GDPGrowth.))+
  geom_point(col="red") +
  geom_smooth(method = "lm",se=TRUE)+
  labs(title="Diagrama de Dispersión",
        subtitle = "Variables: PIB y LifeExpBirth",
        caption="Fuente: Elaboración propia")+
  theme_bw()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Diagrama de Dispersión

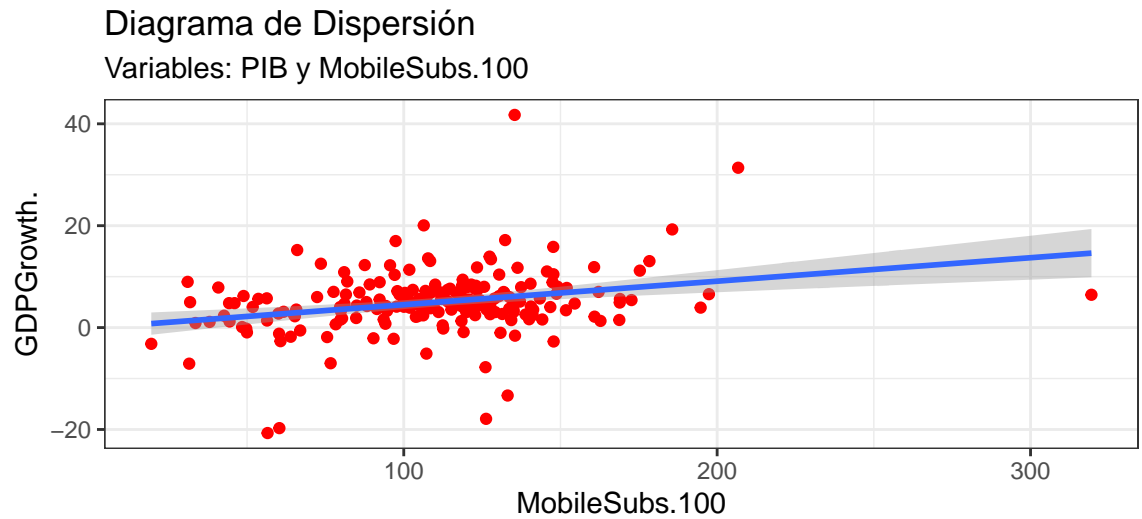
Variables: PIB y LifeExpBirth



Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=MobileSubs.100,y=GDPGrowth.))+
  geom_point(col="red") +
  geom_smooth(method = "lm",se=TRUE)+
  labs(title="Diagrama de Dispersión",
        subtitle = "Variables: PIB y MobileSubs.100",
        caption="Fuente: Elaboración propia")+
  theme_bw()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



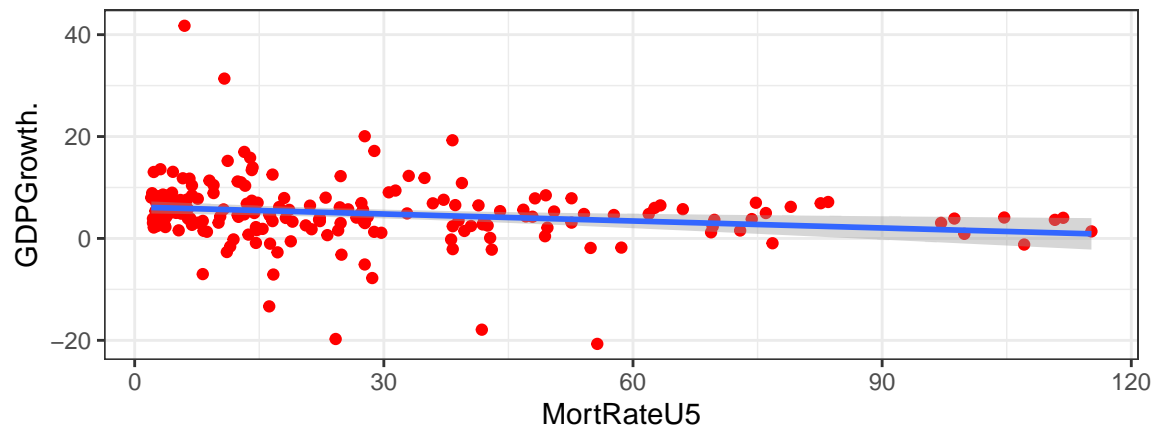
Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=MortRateU5,y=GDPGrowth.))+
  geom_point(col="red") +
  geom_smooth(method = "lm",se=TRUE)+
  labs(title="Diagrama de Dispersión",
        subtitle = "Variables: PIB y MortRateU5",
        caption="Fuente: Elaboración propia")+
  theme_bw()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Diagrama de Dispersión

Variables: PIB y MortRateU5



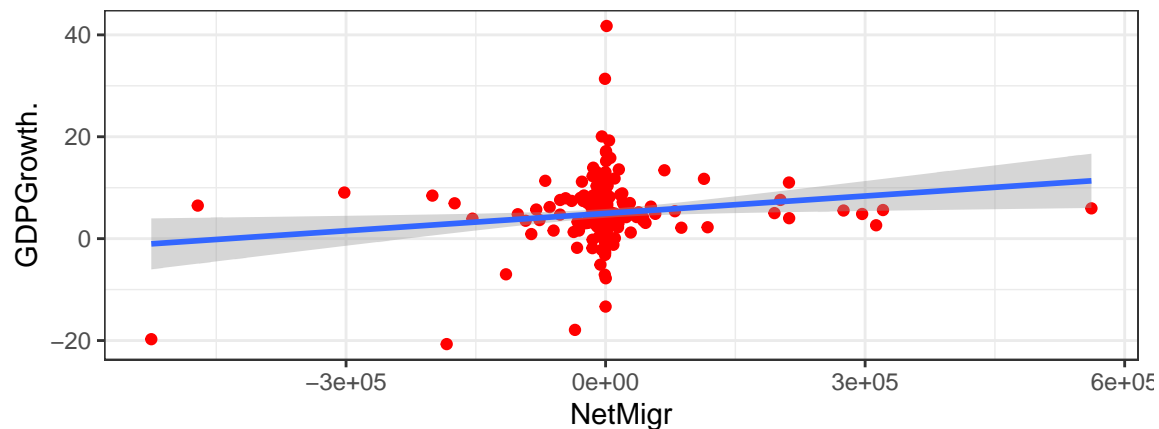
Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=NetMigr,y=GDPGrowth.))+
  geom_point(col="red") +
  geom_smooth(method = "lm",se=TRUE)+
  labs(title="Diagrama de Dispersión",
        subtitle = "Variables: PIB y NetMigr",
        caption="Fuente: Elaboración propia")+
  theme_bw()
```

`geom_smooth()` using formula = 'y ~ x'

Diagrama de Dispersión

Variables: PIB y NetMigr



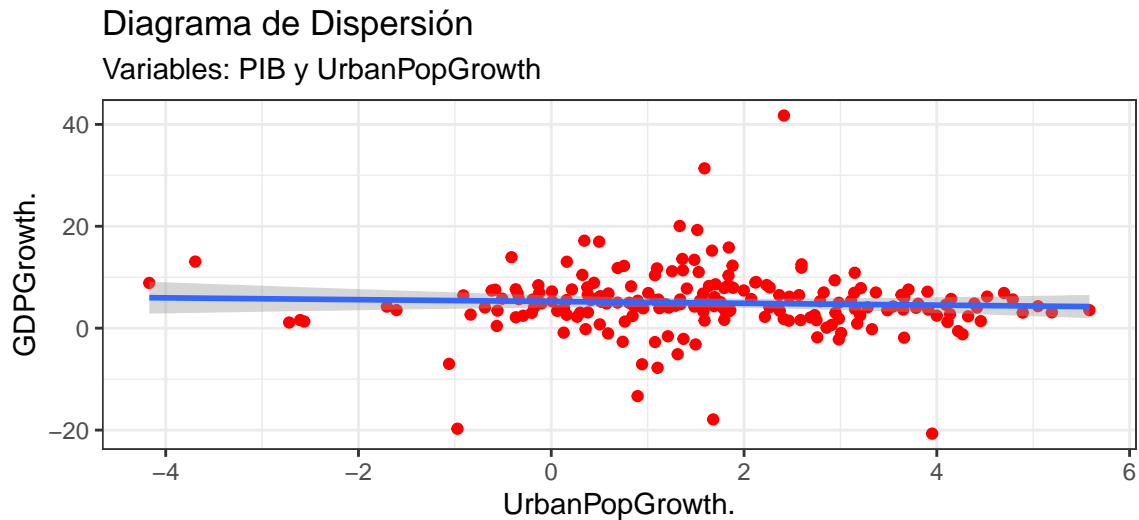
Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=UrbanPopGrowth.,y=GDPGrowth.))+
  geom_point(col="red") +
  geom_smooth(method = "lm",se=TRUE)+
  labs(title="Diagrama de Dispersión",
        subtitle = "Variables: PIB y UrbanPopGrowth",
```



```
caption="Fuente: Elaboración propia")+
theme_bw()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Fuente: Elaboración propia

A través de los diagramas de dispersión podemos ver la relación que existe entre cada una de las regresoras con la variable dependiente, que en este caso es el crecimiento del PIB. Las variables inflación, crecimiento de la población y crecimiento de la población urbana no muestran tener mucha relación con la variable dependiente. Esto porque la línea de tendencia de estos diagramas no tiene mucha pendiente. Cabe resaltar que en el diagrama de la inflación, aunque hay varios datos atípicos, la mayoría de las observaciones se encuentran entre 0 y 25. Por otro lado, los diagramas de fertilidad entre adolescentes, fertilidad y mortalidad en niños menores a cinco años señalan una relación negativa con el crecimiento económico. Finalmente, la esperanza de vida, las suscripciones de teléfono celular y la migración tienen una relación positiva con el crecimiento del PIB. Una observación interesante es que en el diagrama de migración neta, la mayoría de las observaciones se encuentran alrededor del cero. Una posible explicación de esto es que el flujo de las personas que migran es igual al flujo de personas que inmigran, por lo que el efecto es compensatorio y el número neto de migraciones es cero.

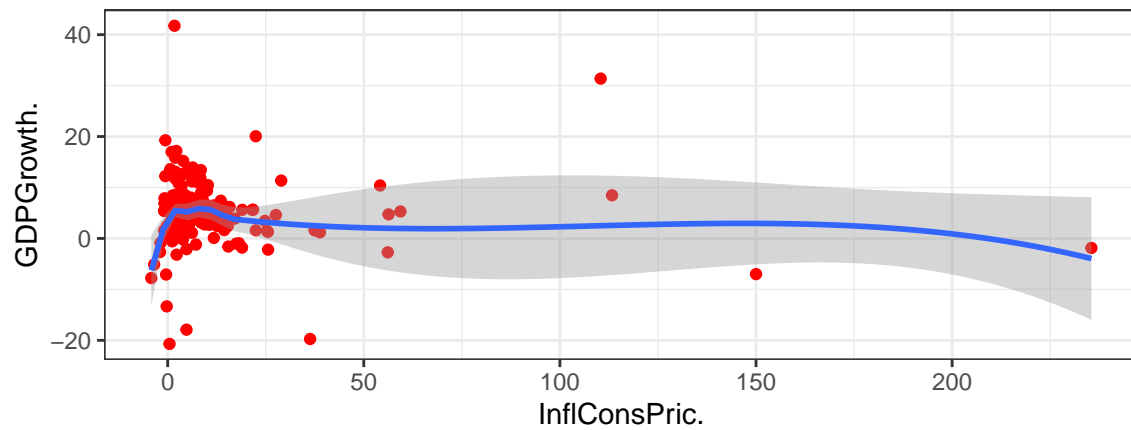
Observemos, solo por curiosidad, si quitamos el parámetro `lm` en `geom_smooth()` si es posible que las variables tengan un comportamiento diferente al lineal:

```
# Diagrama de dispersión
ggplot(data=datosCrecimiento2021,aes(x=InflConsPric.,y=GDPGrowth.))+
  geom_point(col="red") +
  geom_smooth()+
  labs(title="Diagrama de Dispersión",
        subtitle = "Variables: PIB e inflación",
        caption="Fuente: Elaboración propia")+
  theme_bw()
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

Diagrama de Dispersión

Variables: PIB e inflación



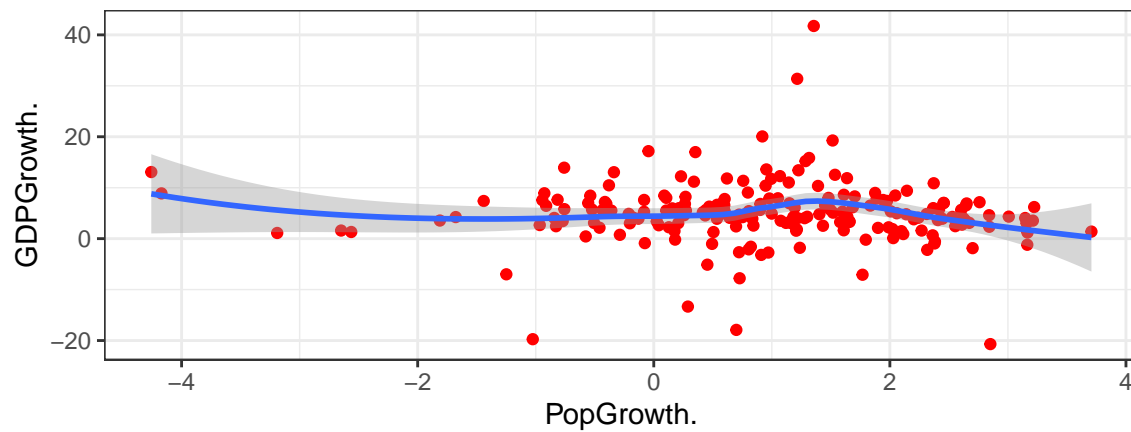
Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=PopGrowth.,y=GDPGrowth.))+  
  geom_point(col="red") +  
  geom_smooth()+  
  labs(title="Diagrama de Dispersión",  
        subtitle = "Variables: PIB y x=PopGrowth.",  
        caption="Fuente: Elaboración propia")+  
  theme_bw()
```

`geom_smooth()` using method = 'loess' and formula = 'y ~ x'

Diagrama de Dispersión

Variables: PIB y x=PopGrowth.



Fuente: Elaboración propia

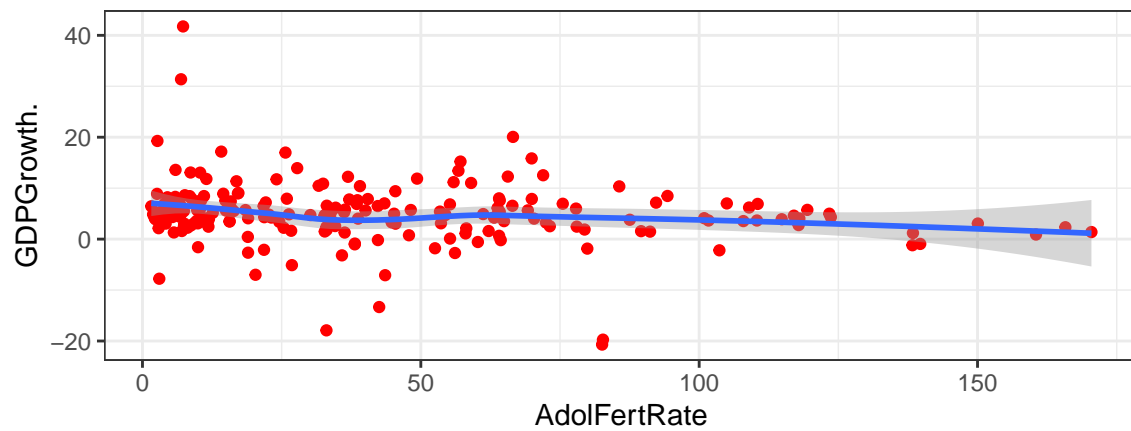
```
ggplot(data=datosCrecimiento2021,aes(x=AdolFertRate,y=GDPGrowth.))+  
  geom_point(col="red") +  
  geom_smooth()+  
  labs(title="Diagrama de Dispersión",  
        subtitle = "Variables: PIB y AdolFertRate",
```

```
caption="Fuente: Elaboración propia")+
theme_bw()
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

Diagrama de Dispersión

Variables: PIB y AdolFertRate



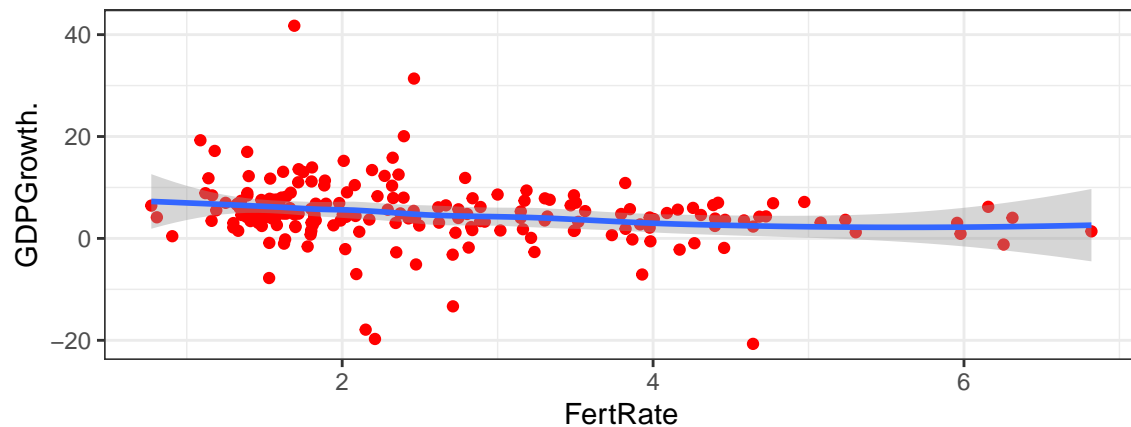
Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=FertRate,y=GDPGrowth.))+
  geom_point(col="red") +
  geom_smooth()+
  labs(title="Diagrama de Dispersión",
        subtitle = "Variables: PIB y FertRate",
        caption="Fuente: Elaboración propia")+
  theme_bw()
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

Diagrama de Dispersión

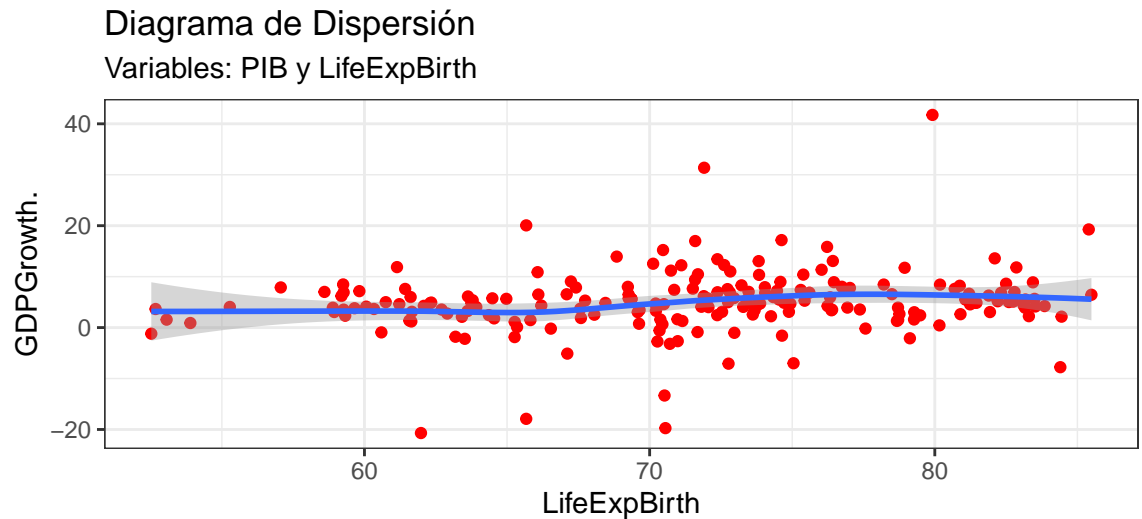
Variables: PIB y FertRate



Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=LifeExpBirth,y=GDPGrowth.))+
  geom_point(col="red") +
  geom_smooth()+
  labs(title="Diagrama de Dispersión",
        subtitle = "Variables: PIB y LifeExpBirth",
        caption="Fuente: Elaboración propia")+
  theme_bw()
```

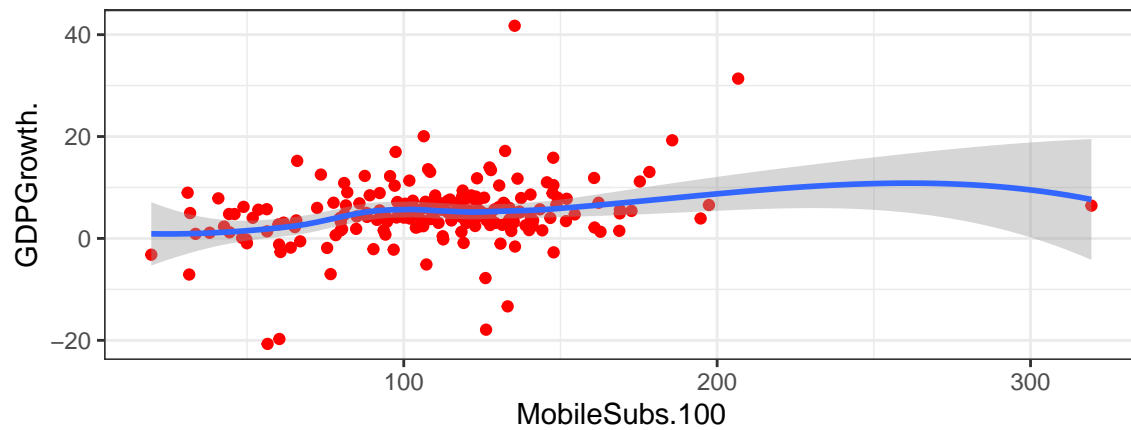
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



```
ggplot(data=datosCrecimiento2021,aes(x=MobileSubs.100,y=GDPGrowth.))+
  geom_point(col="red") +
  geom_smooth()+
  labs(title="Diagrama de Dispersión",
        subtitle = "Variables: PIB y MobileSubs.100",
        caption="Fuente: Elaboración propia")+
  theme_bw()
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

Diagrama de Dispersión
Variables: PIB y MobileSubs.100

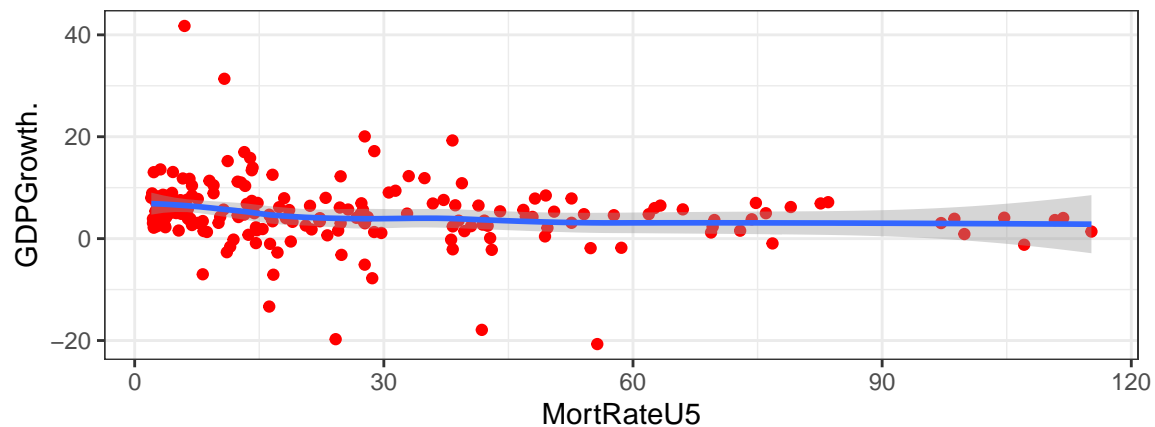


Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=MortRateU5,y=GDPGrowth.))+
  geom_point(col="red") +
  geom_smooth()+
  labs(title="Diagrama de Dispersión",
        subtitle = "Variables: PIB y MortRateU5",
        caption="Fuente: Elaboración propia")+
  theme_bw()
```

`geom_smooth()` using method = 'loess' and formula = 'y ~ x'

Diagrama de Dispersión
Variables: PIB y MortRateU5



Fuente: Elaboración propia

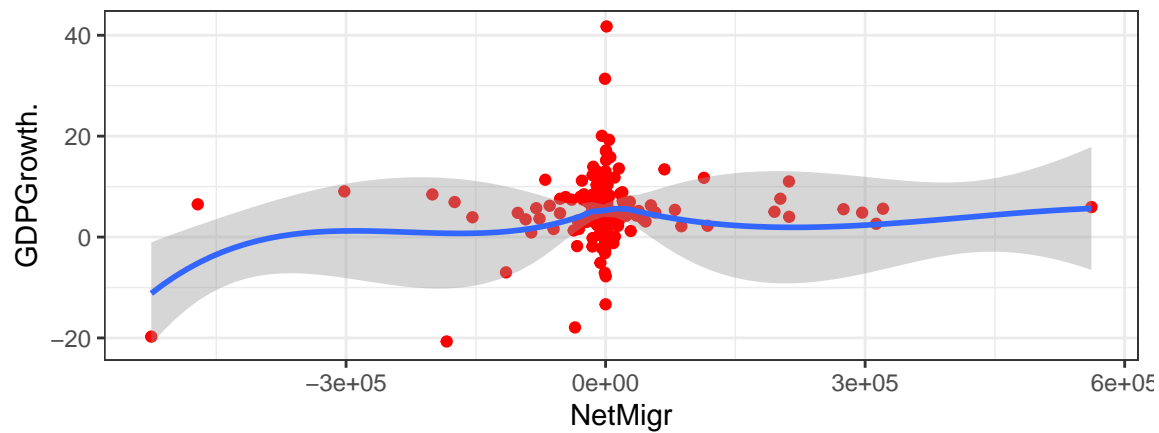
```
ggplot(data=datosCrecimiento2021,aes(x=NetMigr,y=GDPGrowth.))+
  geom_point(col="red") +
  geom_smooth()+
  labs(title="Diagrama de Dispersión",
        subtitle = "Variables: PIB y NetMigr",
```

```
caption="Fuente: Elaboración propia")+
theme_bw()
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

Diagrama de Dispersión

Variables: PIB y NetMigr



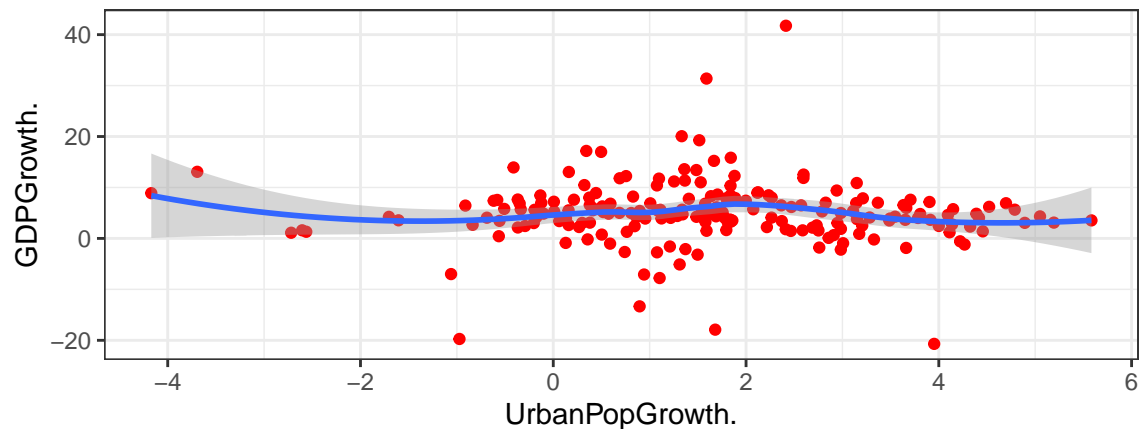
Fuente: Elaboración propia

```
ggplot(data=datosCrecimiento2021,aes(x=UrbanPopGrowth.,y=GDPGrowth.))+
  geom_point(col="red") +
  geom_smooth()+
  labs(title="Diagrama de Dispersión",
        subtitle = "Variables: PIB y UrbanPopGrowth.",
        caption="Fuente: Elaboración propia")+
  theme_bw()
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

Diagrama de Dispersión

Variables: PIB y UrbanPopGrowth.



Fuente: Elaboración propia

Modelo teórico y Modelo estadístico

Nuestro modelo teórico es el siguiente:

$\text{CRECIMIENTO ECONOMICO} = f(\text{CRECIMIENTO POBLACION}, \text{INFLACION}, \text{ESPERANZA VIDA NACER}, \text{SUSCRIPCIONES TELEFONOS CELULARES}, \text{MORTALIDAD NINOS MENORES 5 AÑOS}, \text{MIGRACION NETA}, \text{CRECIMIENTO POBLACION URBANA})$

Del lado derecho de la ecuación se encuentra nuestra variable dependiente y del lado izquierdo tenemos a las variables independientes. Queremos encontrar la relación entre ambos tipos de variables.

El modelo estadístico es el siguiente:

$$Y_i = \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i3} + \alpha_4 X_{i4} + \alpha_5 X_{i5} + \alpha_6 X_{i6} + \alpha_7 X_{i7} + \alpha_8 X_{i8} + \alpha_9 X_{i9} + u_i$$

En este modelo X_{i1} se refiere al crecimiento de la población, X_{i2} a la tasa de fertilidad en adolescentes, X_{i3} a la tasa de fertilidad, X_{i4} a la inflación, X_{i5} a la esperanza de vida, X_{i6} a las suscripciones de teléfonos celulares, X_{i7} a la tasa de mortalidad en niños menores a 5 años, X_{i8} a la migración neta y X_{i9} al crecimiento de la población urbana.

Estimación del modelo de regresión mediante mínimos cuadrados ordinarios

```
# Regresión Lineal Múltiple
n<-length(PopGrowth.)
k<-9
reg1<-lm(GDPGrowth.~PopGrowth.+AdolFertRate+FertRate+InflConsPric.+LifeExpBirth+MobileSubs.100+MortRateU5+
summary(reg1)

##
## Call:
## lm(formula = GDPGrowth. ~ PopGrowth. + AdolFertRate + FertRate +
##      InflConsPric. + LifeExpBirth + MobileSubs.100 + MortRateU5 +
##      NetMigr + UrbanPopGrowth., data = datosCrecimiento2021)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.109  -2.813   -0.223    2.330   33.647
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.593e+00  9.385e+00   0.703   0.4832
## PopGrowth.     1.453e-02  1.145e+00   0.013   0.9899
## AdolFertRate   -5.206e-03  2.308e-02  -0.226   0.8218
## FertRate       -1.538e+00  8.167e-01  -1.884   0.0611 .
## InflConsPric.  -8.534e-04  1.835e-02  -0.047   0.9630
## LifeExpBirth   -4.189e-02  1.119e-01  -0.374   0.7085
## MobileSubs.100  3.612e-02  1.318e-02   2.742   0.0067 **
## MortRateU5     -7.595e-03  3.371e-02  -0.225   0.8220
## NetMigr         6.772e-06  4.782e-06   1.416   0.1584
## UrbanPopGrowth. 1.086e+00  9.335e-01   1.164   0.2460
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 6.008 on 191 degrees of freedom
## Multiple R-squared:  0.1333, Adjusted R-squared:  0.09241
## F-statistic: 3.263 on 9 and 191 DF,  p-value: 0.001028
```

```
stargazer(reg1,type="text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               GDPGrowth.
## -----
## PopGrowth.                  0.015
##                               (1.145)
##
## AdolFertRate                -0.005
##                               (0.023)
##
## FertRate                   -1.538*
##                               (0.817)
##
## InflConsPric.              -0.001
##                               (0.018)
##
## LifeExpBirth               -0.042
##                               (0.112)
##
## MobileSubs.100             0.036***
##                               (0.013)
##
## MortRateU5                 -0.008
##                               (0.034)
##
## NetMigr                     0.00001
##                               (0.00000)
##
## UrbanPopGrowth.            1.086
##                               (0.934)
##
## Constant                   6.593
##                               (9.385)
##
## -----
## Observations                201
## R2                          0.133
## Adjusted R2                 0.092
## Residual Std. Error        6.008 (df = 191)
## F Statistic                 3.263*** (df = 9; 191)
## =====
## Note:                       *p<0.1; **p<0.05; ***p<0.01
```

A través del método de mínimos cuadrados ordinarios construimos un modelo de regresión lineal múltiple. Esto explica de qué manera y a qué grado se relacionan nuestras variables independientes con el crecimiento

económico. Los resultados fueron los siguientes: por un lado, las suscripciones de teléfono celular resultaron ser la única variable significativa y la tasa de fertilidad es marginalmente significativa; por otro lado, la R^2 fue igual a 0.1333 y la R^2 ajustada fue igual a 0.09241. En pocas palabras, los resultados de esta regresión no eran los que esperábamos, pues la mayoría de estas variables no son significativas y los valores de la R^2 y R^2 ajustada fueron muy pequeños.

A primera vista, este modelo no es muy bueno para estimar cambios en el crecimiento económico y, por lo tanto, debe ser corregido. No obstante, antes de realizar cualquier corrección, hay que calcular la significancia individual y conjunta de las variables independientes, verificar el resto de supuestos, etc.

```
ygorro<-reg1$fitted.values
ugorro<-reg1$residuals
sigma2<-sum(ugorro^2)/(n-k-1)
```

Definimos ygorro: valores estimados, ugorro: errores estimados, sigma2: varianza de los residuales.

Ahora vamos a probar la significancia individual de cada variable.

```
# Pruebas de significancia (individual)
linearHypothesis(reg1,"PopGrowth.=0")
```

```
## Linear hypothesis test
##
## Hypothesis:
## PopGrowth. = 0
##
## Model 1: restricted model
## Model 2: GDPGrowth. ~ PopGrowth. + AdolFertRate + FertRate + InflConsPric. +
##      LifeExpBirth + MobileSubs.100 + MortRateU5 + NetMigr + UrbanPopGrowth.
##
##      Res.Df  RSS Df Sum of Sq      F Pr(>F)
## 1      192 6895
## 2      191 6895  1 0.0058142 2e-04 0.9899
```

```
linearHypothesis(reg1,"AdolFertRate=0")
```

```
## Linear hypothesis test
##
## Hypothesis:
## AdolFertRate = 0
##
## Model 1: restricted model
## Model 2: GDPGrowth. ~ PopGrowth. + AdolFertRate + FertRate + InflConsPric. +
##      LifeExpBirth + MobileSubs.100 + MortRateU5 + NetMigr + UrbanPopGrowth.
##
##      Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1      192 6896.9
## 2      191 6895.0  1      1.837 0.0509 0.8218
```

```
linearHypothesis(reg1,"FertRate=0")
```

```
## Linear hypothesis test
```

```
##
## Hypothesis:
## FertRate = 0
##
## Model 1: restricted model
## Model 2: GDPGrowth. ~ PopGrowth. + AdolFertRate + FertRate + InflConsPric. +
##      LifeExpBirth + MobileSubs.100 + MortRateU5 + NetMigr + UrbanPopGrowth.
##
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1     192 7023.1
## 2     191 6895.0  1    128.07 3.5478 0.06114 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
linearHypothesis(reg1,"InflConsPric.=0")
```

```
## Linear hypothesis test
##
## Hypothesis:
## InflConsPric. = 0
##
## Model 1: restricted model
## Model 2: GDPGrowth. ~ PopGrowth. + AdolFertRate + FertRate + InflConsPric. +
##      LifeExpBirth + MobileSubs.100 + MortRateU5 + NetMigr + UrbanPopGrowth.
##
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1     192 6895.1
## 2     191 6895.0  1   0.078076 0.0022 0.963
```

```
linearHypothesis(reg1,"LifeExpBirth=0")
```

```
## Linear hypothesis test
##
## Hypothesis:
## LifeExpBirth = 0
##
## Model 1: restricted model
## Model 2: GDPGrowth. ~ PopGrowth. + AdolFertRate + FertRate + InflConsPric. +
##      LifeExpBirth + MobileSubs.100 + MortRateU5 + NetMigr + UrbanPopGrowth.
##
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1     192 6900.1
## 2     191 6895.0  1    5.0615 0.1402 0.7085
```

```
linearHypothesis(reg1,"MobileSubs.100=0")
```

```
## Linear hypothesis test
##
## Hypothesis:
## MobileSubs.100 = 0
##
## Model 1: restricted model
```

```
## Model 2: GDPGrowth. ~ PopGrowth. + AdolFertRate + FertRate + InflConsPric. +
##      LifeExpBirth + MobileSubs.100 + MortRateU5 + NetMigr + UrbanPopGrowth.
##
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     192 7166.4
## 2     191 6895.0   1    271.32 7.5159 0.006697 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
linearHypothesis(reg1,"MortRateU5=0")
```

```
## Linear hypothesis test
##
## Hypothesis:
## MortRateU5 = 0
##
## Model 1: restricted model
## Model 2: GDPGrowth. ~ PopGrowth. + AdolFertRate + FertRate + InflConsPric. +
##      LifeExpBirth + MobileSubs.100 + MortRateU5 + NetMigr + UrbanPopGrowth.
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     192 6896.9
## 2     191 6895.0   1    1.8327 0.0508 0.822
```

```
linearHypothesis(reg1,"UrbanPopGrowth.=0")
```

```
## Linear hypothesis test
##
## Hypothesis:
## UrbanPopGrowth. = 0
##
## Model 1: restricted model
## Model 2: GDPGrowth. ~ PopGrowth. + AdolFertRate + FertRate + InflConsPric. +
##      LifeExpBirth + MobileSubs.100 + MortRateU5 + NetMigr + UrbanPopGrowth.
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     192 6943.9
## 2     191 6895.0   1    48.892 1.3543 0.246
```

Ahora vamos a probar la significancia individual de cada variable. En esta prueba confirmamos lo que anteriormente habíamos observado. Primero, el crecimiento de la población, la tasa de fertilidad en adolescentes, la inflación, la esperanza de vida, la tasa de mortalidad en niños menores a cinco años y el crecimiento de la población no son variables significativas. Esto se debe a que el valor p de cada una de estas variables es mayor a 0.01. Segundo, la tasa de fertilidad es ligeramente significativa, pues su valor p se encuentra entre los parámetros 0.05 y 0.1. Por último, la única variable significativa son las suscripciones de teléfono celular. Su valor p fue igual a 0.006697.

```
# Prueba de significancia (conjunta)
```

```
linearHypothesis(reg1,c("PopGrowth.=0", "InflConsPric.=0", "FertRate=0", "LifeExpBirth=0", "AdolFertRate=0"))
```

```
## Linear hypothesis test
```

```
##
## Hypothesis:
## PopGrowth. = 0
## InflConsPric. = 0
## FertRate = 0
## LifeExpBirth = 0
## AdolFertRate = 0
## MobileSubs.100 = 0
## MortRateU5 = 0
## NetMigr = 0
## UrbanPopGrowth. = 0
##
## Model 1: restricted model
## Model 2: GDPGrowth. ~ PopGrowth. + AdolFertRate + FertRate + InflConsPric. +
##      LifeExpBirth + MobileSubs.100 + MortRateU5 + NetMigr + UrbanPopGrowth.
##
##      Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1      200 7955.1
## 2      191 6895.0   9      1060 3.2626 0.001028 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Posteriormente, realizaremos una prueba de significancia conjunta. En esta prueba evaluaremos si en el grupo de variables independientes existe alguna significativa en el modelo de regresión. El resultado de esta prueba arrojó un valor p igual a 0.001028. Esto indica que existe suficiente evidencia para rechazar la hipótesis nula y, por lo tanto, existe al menos una variable significativa. Por las pruebas individuales que realizamos anteriormente, podemos inferir que las variables significativas son la tasa de fertilidad en adolescentes y las suscripciones de teléfono celular.

Para continuar con el análisis, aplicaremos la función `anova()`. Esta se utiliza para realizar análisis de la varianza en modelos lineales.

```
anova(reg1)
```

```
## Analysis of Variance Table
##
## Response: GDPGrowth.
##
##      Df Sum Sq Mean Sq F value    Pr(>F)
## PopGrowth.      1    17.2    17.235    0.4774 0.490432
## AdolFertRate      1   315.0   315.047    8.7271 0.003529 **
## FertRate          1   316.0   316.026    8.7543 0.003479 **
## InflConsPric.      1     0.7     0.661    0.0183 0.892490
## LifeExpBirth       1     4.9     4.926    0.1364 0.712251
## MobileSubs.100     1   287.3   287.347    7.9598 0.005288 **
## MortRateU5         1     2.7     2.747    0.0761 0.782972
## NetMigr            1    67.1    67.138    1.8598 0.174254
## UrbanPopGrowth.    1    48.9    48.892    1.3543 0.245971
## Residuals        191  6895.0   36.100
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Podemos observar una tabla que presenta datos para cada variable sobre los grados de libertad, la suma de cuadrados, la media de cuadrados, el valor F y el valor P. En cuanto a la suma de cuadrados de cada

variable, sabemos que entre mayor sea el valor de esta columna mayor será la variabilidad explicada por cada variable independiente. Cabe destacar que en esta columna también se presentan datos sobre los residuales del modelo. De esta manera, podemos inferir que la variabilidad que no se explica por ninguna variable independiente es mayor a la variabilidad explicada por cualquier variable del modelo.

La siguiente columna (media de cuadrados) brinda información de la variabilidad promedio que se le puede atribuir a cada variable independiente. Es interesante observar que ahora la variabilidad promedio atribuida a la tasa de fertilidad en adolescentes, tasa de fertilidad y suscripciones de teléfonos celulares es mayor a la media de cuadrados de los residuales. Por último, debido al valor F y al valor P, podemos observar nuevamente la significancia de cada variable del modelo. Las interpretaciones de esto se pueden encontrar en las secciones anteriores con las pruebas de significancia conjunta e individual.

A continuación, haremos dos intervalos de confianza para cada variable del modelo. El primero consiste en un intervalo de confianza al 95 % y el segundo en un intervalo de confianza al 90%.

```
# Intervalos de confianza
```

```
confint(reg1)
```

```
##                2.5 %      97.5 %
## (Intercept)    -1.191847e+01 2.510447e+01
## PopGrowth.     -2.243312e+00 2.272366e+00
## AdolFertRate   -5.073082e-02 4.031797e-02
## FertRate       -3.149332e+00 7.261357e-02
## InflConsPric.  -3.705036e-02 3.534349e-02
## LifeExpBirth   -2.625663e-01 1.787826e-01
## MobileSubs.100  1.013338e-02 6.211333e-02
## MortRateU5     -7.408313e-02 5.889319e-02
## NetMigr        -2.660424e-06 1.620433e-05
## UrbanPopGrowth -7.549263e-01 2.927697e+00
```

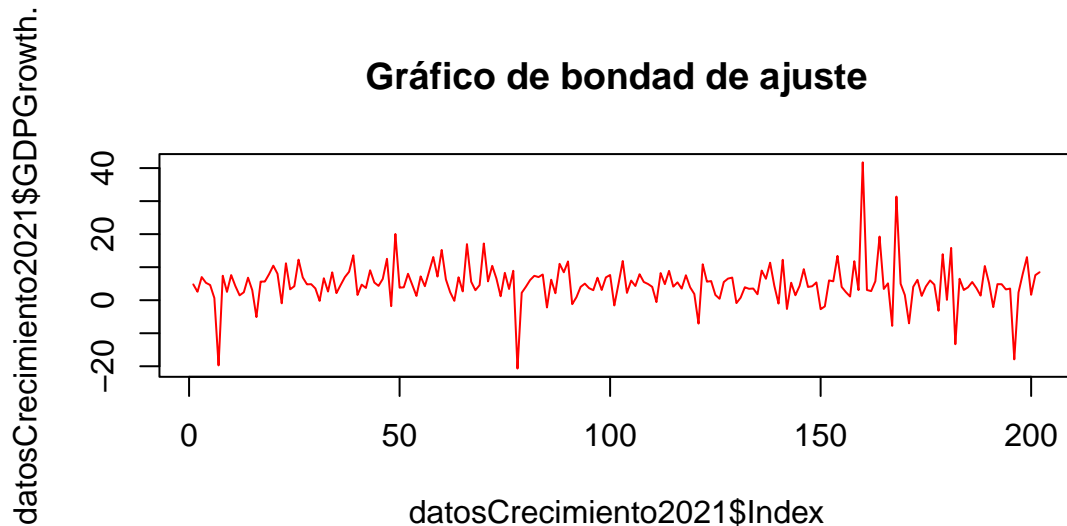
```
confint(reg1,level=0.90)
```

```
##                5 %      95 %
## (Intercept)    -8.919117e+00 2.210512e+01
## PopGrowth.     -1.877482e+00 1.906536e+00
## AdolFertRate   -4.335465e-02 3.294181e-02
## FertRate       -2.888311e+00 -1.884069e-01
## InflConsPric.  -3.118549e-02 2.947862e-02
## LifeExpBirth   -2.268112e-01 1.430274e-01
## MobileSubs.100  1.434444e-02 5.790226e-02
## MortRateU5     -6.331027e-02 4.812034e-02
## NetMigr        -1.132128e-06 1.467603e-05
## UrbanPopGrowth -4.565847e-01 2.629356e+00
```

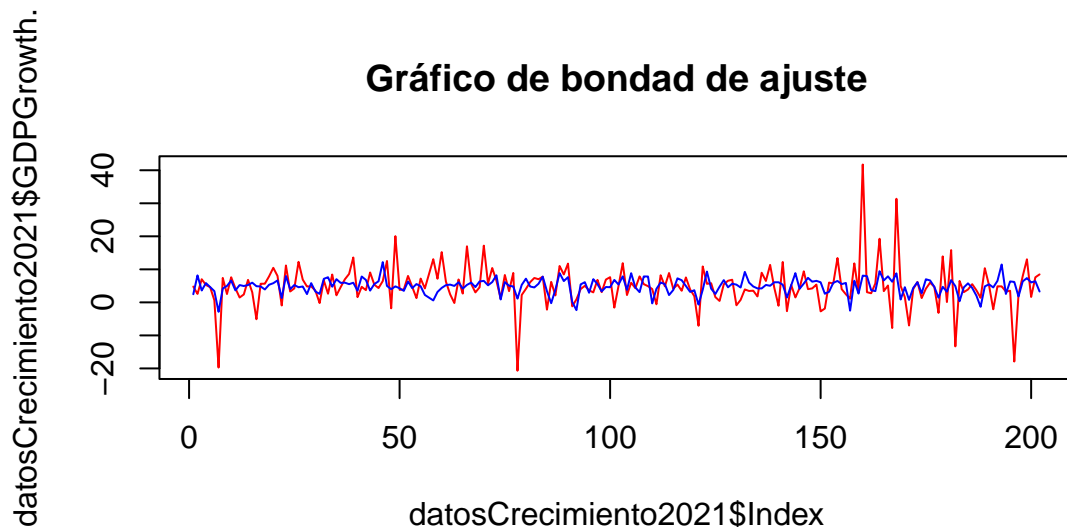
El problema de estos intervalos es que para la mayoría de las variables, el intervalo incluye al cero. Sólo el intervalo de confianza de las suscripciones de teléfono celular no incluye al cero. Esto es problemático, porque esto podría ser indicativo de que la variable independiente no tiene una relación significativa con el crecimiento del PIB.

Ahora vamos a hacer el gráfico de bondad de ajuste, a través de los cuales compararemos los valores observados con los valores estimados de crecimiento económico. A partir del gráfico podemos señalar que no existe un problema de colinealidad, pues los datos observados difieren de los datos estimados.

```
# Gráfico de bondad de ajuste
plot(datosCrecimiento2021$Index,datosCrecimiento2021$GDPPGrowth.,type="l",
     col="red",
     main="Gráfico de bondad de ajuste")
```

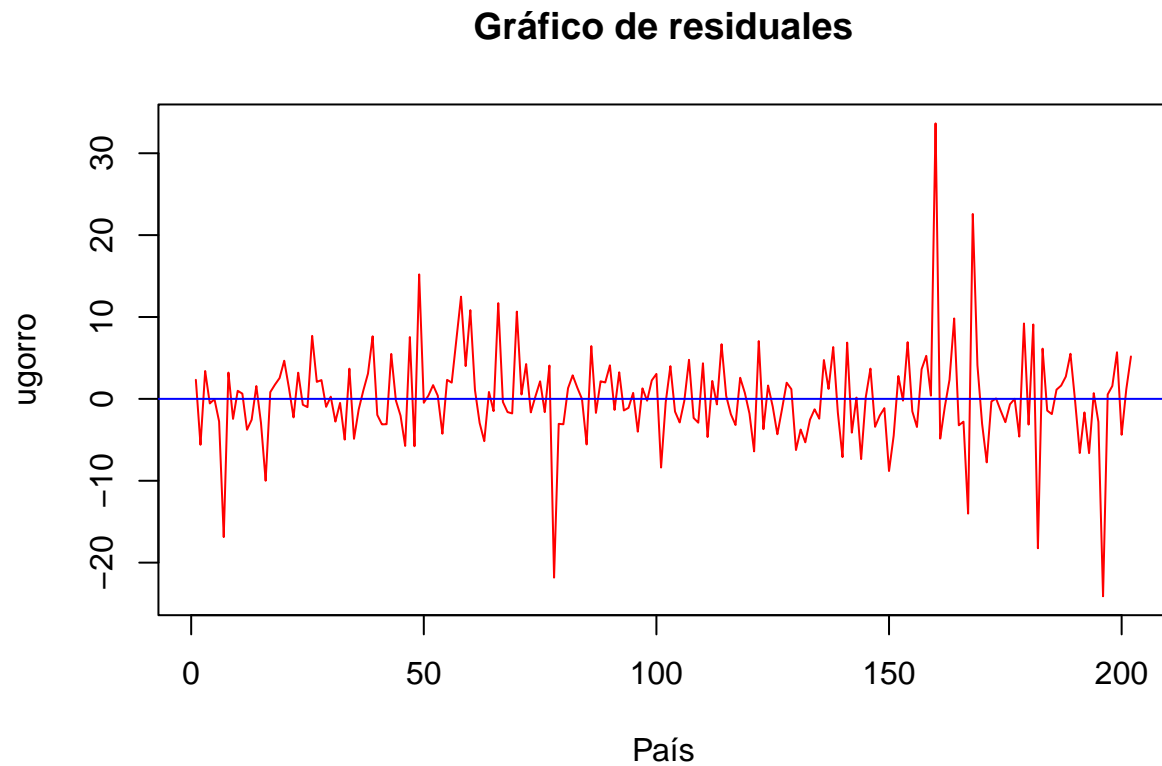


```
plot(datosCrecimiento2021$Index,datosCrecimiento2021$GDPPGrowth.,type="l",
     col="red",
     main="Gráfico de bondad de ajuste")
lines(datosCrecimiento2021$Index,ygorro,type="l",col="blue")
```



Finalmente realizaremos el gráfico de residuales. En este observamos que no hay patrones en los residuales, lo que significa que no hay autocorrelación. De igual manera, podemos decir que no hay heterocedasticidad, pues los residuales se distribuyen de una manera homogénea.

```
# Gráfico de residuales
plot(datosCrecimiento2021$Index,ugorro,type="l",
     col="red",
     main="Gráfico de residuales",
     xlab = "País")
abline(h=0, col="blue")
```

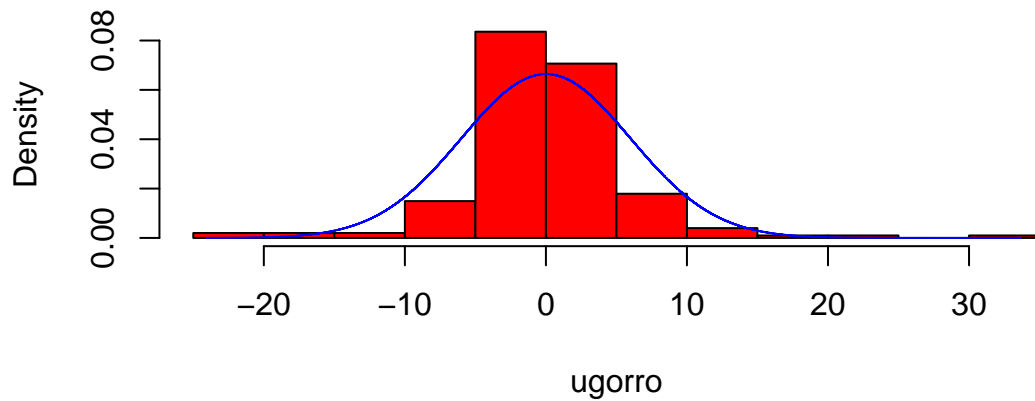


Verificación de los supuestos

Verificación de Normalidad

```
hist(ugorro,breaks=10,col="red",freq = F,
     main="Histograma con curva Normal")
dz<-seq(min(ugorro), max(ugorro), 0.001)
lines(dnorm(dz,0,sqrt(sigma2))~dz, type="l",
     col="blue", lwd=1)
```

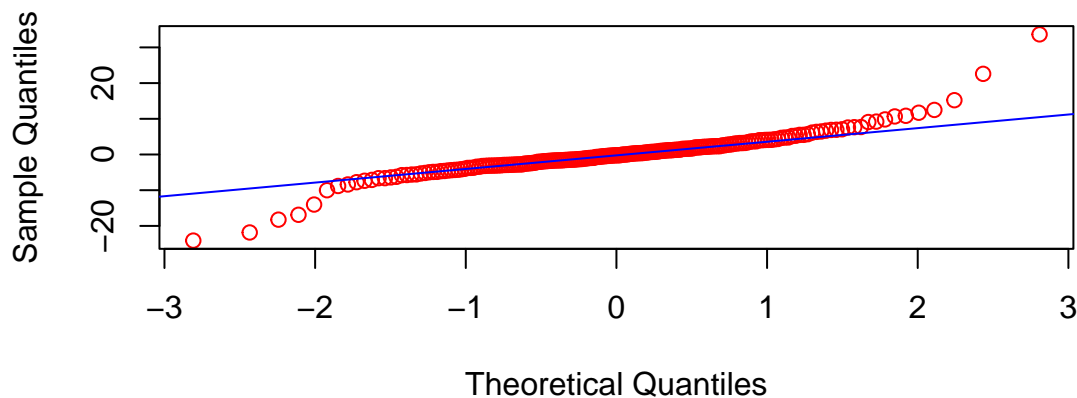
Histograma con curva Normal



Si bien el histograma no nos sugiere que la distribución de nuestros residuales es normal, es importante mencionar que no presenta ningún sesgo. El gráfico debería de parecerse lo más posible a la curva de la distribución normal, pero aquí podemos observar que en el intervalo de $[-5, 5]$ hay frecuencia de más. Por otro lado, los intervalos de $[-15, -5]$ y $[5, 15]$ tienen una frecuencia menor a la esperada.

```
qqnorm(ugorro,col="red",main="Gráfico QQ Normal")
qqline(ugorro,col="blue")
```

Gráfico QQ Normal



Este gráfico nos muestra que la mayoría de nuestros residuales se encuentran sobre la línea azul de 45 grados; sin embargo, podemos observar claramente la presencia de datos atípicos en los extremos. Los intervalos de $[-3, -2]$ y $[1.8, 3]$ son aquellos en los que se muestran estos outliers. Todos los puntos deberían de pertenecer a un mismo rango.

Prueba Jarque-Bera

Esta prueba claramente nos indica que la distribución de nuestro término estocástico no es normal. El valor p es considerablemente menor que mi α , el nivel de significancia. De este modo, se rechaza la hipótesis nula, la cual nos indica que la distribución es normal.


```
jarque.bera.test(ugorro)
```

```
##
## Jarque Bera Test
##
## data: ugorro
## X-squared = 526.61, df = 2, p-value < 2.2e-16
```

Para interpretar la prueba de Jarque Bera, recordemos como funciona el p-value:

El criterio para rechazar H_0 es comparar el valor-p con mi nivel de significancia. El valor-p funciona como mi crítico a partir el cual rechazo.

- Si valor $p >$ nivel de significancia, no rechazo H_0
- Valor $p <$ nivel de significancia, rechazo H_0

Verificación de Multicolinealidad

Primero, revisemos si nuestra matriz es de rango completo:

```
design_matrix <- model.matrix(reg1)
rank_design_matrix <- qr(design_matrix)$rank
print(rank_design_matrix)
```

```
## [1] 10
```

El rango tiene que ser de $K+1$, es decir, si tiene rango completo (10).

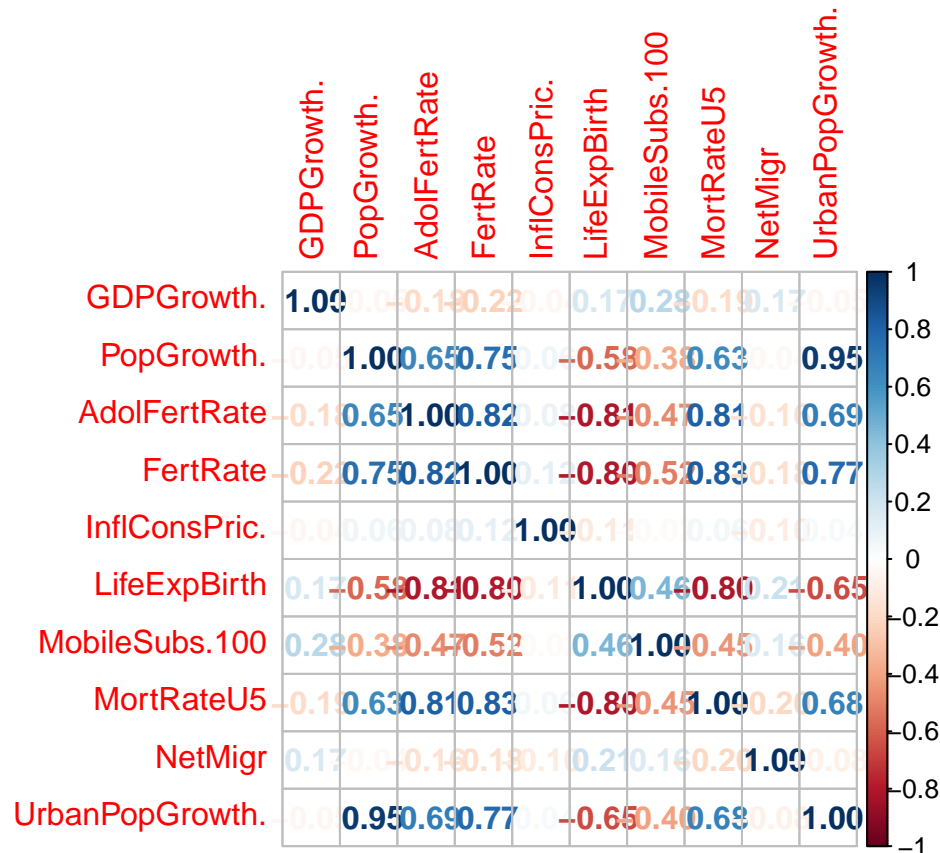
Recordemos, como lo establecimos anteriormente, que la multicolinealidad surge cuando hay una alta correlación entre mis regresoras. Comprobemos eso:

```
# Matriz de correlaciones
vardep<-data.frame(GDPGrowth.,PopGrowth., AdolFertRate, FertRate, InflConsPric., LifeExpBirth, MobileSubs.100, MortRateU5, NetMigr, UrbanPopGrowth.)
(M<-round(cor(vardep),4))
```

```
##          GDPGrowth. PopGrowth. AdolFertRate FertRate InflConsPric.
## GDPGrowth.      1.0000   -0.0465    -0.1813  -0.2203    -0.0425
## PopGrowth.     -0.0465    1.0000     0.6515   0.7520     0.0607
## AdolFertRate   -0.1813    0.6515     1.0000   0.8210     0.0828
## FertRate       -0.2203    0.7520     0.8210   1.0000     0.1183
## InflConsPric.  -0.0425    0.0607     0.0828   0.1183     1.0000
## LifeExpBirth    0.1732   -0.5790    -0.8068  -0.7985    -0.1124
## MobileSubs.100  0.2779   -0.3785    -0.4723  -0.5153    -0.0236
## MortRateU5     -0.1879    0.6300     0.8120   0.8277     0.0577
## NetMigr         0.1675   -0.0389    -0.1558  -0.1799    -0.1002
## UrbanPopGrowth. -0.0477    0.9542     0.6940   0.7712     0.0359
##          LifeExpBirth MobileSubs.100 MortRateU5 NetMigr UrbanPopGrowth.
## GDPGrowth.      0.1732                0.2779   -0.1879  0.1675    -0.0477
## PopGrowth.     -0.5790                0.6300  -0.0389    0.9542
## AdolFertRate   -0.8068                0.8120  -0.1558    0.6940
```

```
## FertRate          -0.7985      -0.5153      0.8277 -0.1799      0.7712
## InflConsPric.     -0.1124      -0.0236      0.0577 -0.1002      0.0359
## LifeExpBirth       1.0000       0.4557     -0.8002  0.2137     -0.6508
## MobileSubs.100     0.4557       1.0000     -0.4471  0.1585     -0.4043
## MortRateU5         -0.8002     -0.4471      1.0000 -0.2042      0.6769
## NetMigr            0.2137       0.1585     -0.2042  1.0000     -0.0764
## UrbanPopGrowth.   -0.6508     -0.4043      0.6769 -0.0764      1.0000
```

```
corrplot(cor(vardep), method= "number")
```



Esta tabla nos muestra las correlaciones lineales que existen entre las regresoras de nuestro modelo. La variable explicada tiene una correlación relativamente pequeña con todas las variables. Del mismo modo, el crecimiento del PIB muestra una correlación negativa con el crecimiento poblacional, la tasa de fertilidad adolescente, tasa de fertilidad, la inflación, la tasa de mortalidad infantil de niños menores a cinco años y el crecimiento de la población urbana. Esto nos indica que si alguna de estas seis regresoras aumenta en una unidad o, en el caso de las tasas, en un 0.01, la variable dependiente disminuirá en un 1%. Por otro lado, el crecimiento del PIB goza de una correlación directa con la esperanza de vida al nacer y la tasa de suscripciones de teléfono.

Ahora bien, podemos observar que el crecimiento poblacional y el crecimiento de la población urbana tienen una correlación positiva de más de 0.7. Las variables relacionadas con los nacimientos o la fertilidad están altamente correlacionadas con la tasa de mortalidad infantil de menores a cinco años. La migración neta no se encuentra muy correlacionada con ninguna variable; sin embargo, tiene una correlación negativa con la tasa de mortalidad infantil de menores a cinco años. Las variables de las tasas de fertilidad y la esperanza de vida al nacer claramente están altamente correlacionadas entre ellas.

```
# R2 y Significancia
stargazer(reg1,type="text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               GDPGrowth.
## -----
## PopGrowth.                  0.015
##                               (1.145)
##
## AdolFertRate                -0.005
##                               (0.023)
##
## FertRate                   -1.538*
##                               (0.817)
##
## InflConsPric.              -0.001
##                               (0.018)
##
## LifeExpBirth               -0.042
##                               (0.112)
##
## MobileSubs.100              0.036***
##                               (0.013)
##
## MortRateU5                 -0.008
##                               (0.034)
##
## NetMigr                     0.00001
##                               (0.00000)
##
## UrbanPopGrowth.             1.086
##                               (0.934)
##
## Constant                    6.593
##                               (9.385)
## -----
## Observations                 201
## R2                           0.133
## Adjusted R2                  0.092
## Residual Std. Error         6.008 (df = 191)
## F Statistic                  3.263*** (df = 9; 191)
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01
```

El R^2 de nuestro modelo es bajo y también el R^2 ajustado. Sin embargo, las variables que muestran significancia son la tasa de suscripciones de teléfono y la tasa de fertilidad. Las otras variables no son significativas. Ambas tasas de fertilidad están relacionadas indirectamente con el crecimiento económico, pero la tasa de fertilidad es la significativa.

Regresiones Auxiliares de Klein La Regla de Klein se basa en estimar k regresiones auxiliares, una para cada regresor en función de los regresores del modelo. Cuando la R^2 de las regresiones auxiliares es mayor a la R^2 del modelo original decimos que hay multicolinealidad en el modelo.

FIV y TOL: Flactor Inflador de la Varianza y Tolerancia

El FIV nos indica que tan “inflada” está la varianza de un estimador por la presencia de multicolinealidad. Cuando el FIV es mayor a 10 comenzamos a preocuparnos por la multicolinealidad.

El TOL es el recíproco del FIV, por lo que queremos que este sea un número mayor a cero, pero menor a 1.

```
# Factor inflador de la varianza (FIV) y Tolerancia (TOL)
```

```
vif(reg1)
```

##	PopGrowth.	AdolFertRate	FertRate	InflConsPric.	LifeExpBirth
##	12.664912	4.237572	5.761203	1.049391	4.018555
##	MobileSubs.100	MortRateU5	NetMigr	UrbanPopGrowth.	
##	1.390548	4.232377	1.091766	13.684090	

```
(tol <- 1/vif(reg1))
```

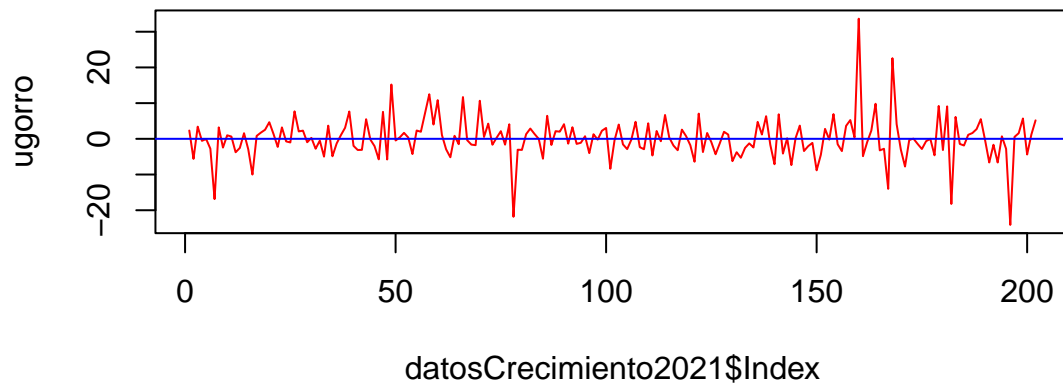
##	PopGrowth.	AdolFertRate	FertRate	InflConsPric.	LifeExpBirth
##	0.07895831	0.23598421	0.17357485	0.95293365	0.24884567
##	MobileSubs.100	MortRateU5	NetMigr	UrbanPopGrowth.	
##	0.71914072	0.23627386	0.91594696	0.07307757	

En este caso, es preferible cuando el factor inflador de la varianza es menor a diez. Dicho esto, nuestro modelo muestra que las variables de crecimiento poblacional y el crecimiento de la población urbana claramente “inflan” su propia varianza. De este modo, podemos observar la existencia de la multicolinealidad en nuestro modelo. Ahora bien, la tolerancia muestra lo mismo. Pero es importante mencionar que las variables más cercanas a uno son la inflación y la migración neta.

Verificación de Heteroscedasticidad

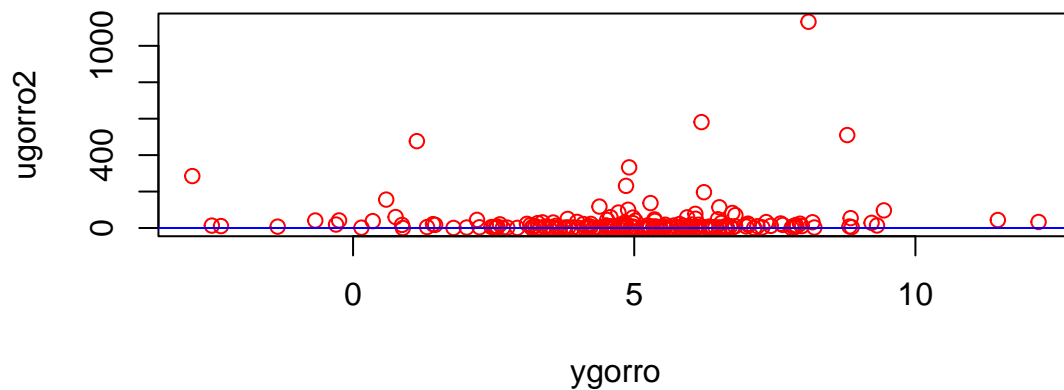
La heteroscedasticidad ocurre cuando la varianza de los residuales no es constante. Lo que buscamos es que los residuales varíen sobre un mismo rango. Este gráfico nos indica que nuestros residuales se están comportando como lo esperado. Es posible apreciar algunas partes en las que puede aumentar o disminuir la varianza; sin embargo, estos cambios son pocos.

```
plot(datosCrecimiento2021$Index,ugorro,type="l",col="red")
abline(h=0, col="blue")
```



```
ugorro2<-ugorro^2
plot(ygorro,ugorro2,type="p",col="red",main="Método gráfico")
abline(h=0, col="blue")
```

Método gráfico



Este gráfico de dispersión de los residuales nos muestra como nuestro término estocástico varía en una misma banda de valores. Se pueden apreciar observaciones atípicas, las cuales gozan de una variabilidad diferente a la mayoría. Esto podría ser ocasionado por la misma naturaleza de nuestros datos.

Prueba de Breusch, Pagan y Godfrey

La prueba de Breusch, Pagan y Godfrey necesita el supuesto de normalidad. Sin embargo, para fines de un ejercicio de práctica intentemos interpretar sus resultados.

```
bptest(reg1)
```

```
##
```

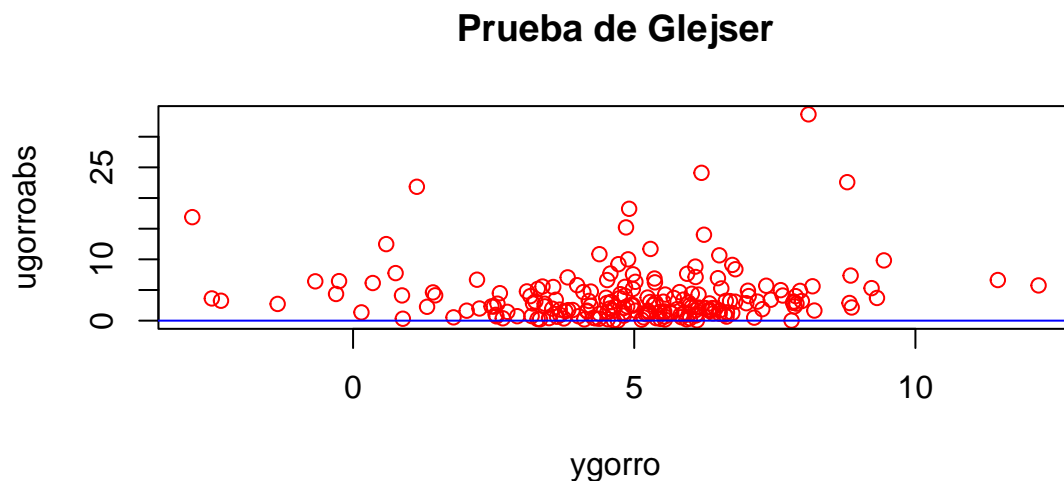
```
## studentized Breusch-Pagan test
##
## data: reg1
## BP = 8.5218, df = 9, p-value = 0.4825
```

Nuestro valor p es mayor que el nivel de significancia. Dicho esto, no rechazamos la hipótesis nula, la cual se refiere al hecho de que hay homocedasticidad. Nuestro valor p es considerablemente grande. Esto nos indica que, a pesar de la presencia de datos atípicos en los gráficos pasados, no hay heterocedasticidad en nuestro término aleatorio.

Prueba de Glejser y Prueba de Park

```
# Prueba de Glejser

ugorroabs<-abs(ugorro)
plot(ygorro,ugorroabs,type="p",col="red",
     main="Prueba de Glejser")
abline(h=0, col="blue")
```



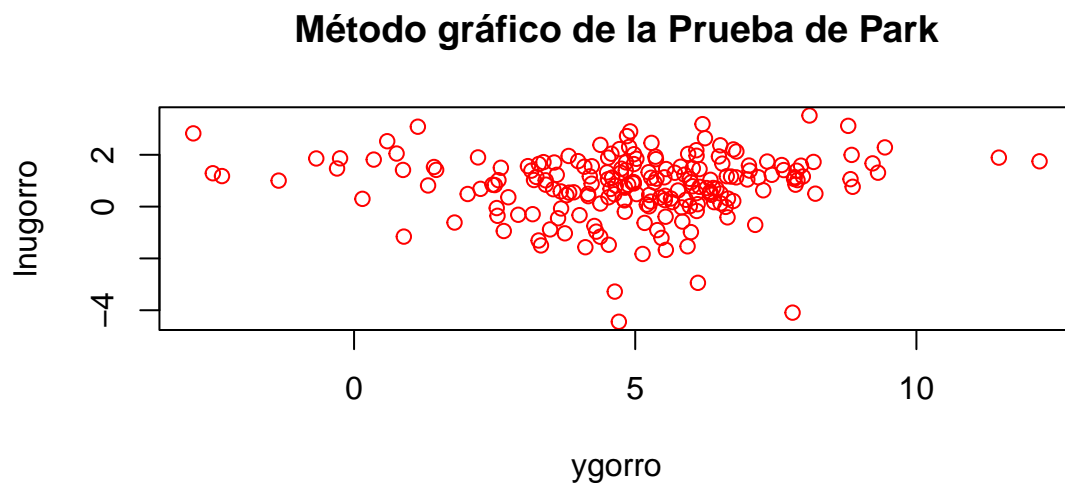
```
aux3h<-lm(ugorroabs~PopGrowth.+AdolFertRate+FertRate+InflConsPric.+LifeExpBirth+MobileSubs.100+MortRateU5)
H03h<-c("PopGrowth.=0", "InflConsPric.=0","FertRate=0","LifeExpBirth=0","AdolFertRate=0","MobileSubs.100=0")
linearHypothesis(aux3h,H03h)
```

```
## Linear hypothesis test
##
## Hypothesis:
## PopGrowth. = 0
## InflConsPric. = 0
## FertRate = 0
## LifeExpBirth = 0
## AdolFertRate = 0
## MobileSubs.100 = 0
## MortRateU5 = 0
## NetMigr = 0
```

```
## UrbanPopGrowth. = 0
##
## Model 1: restricted model
## Model 2: ugorroabs ~ PopGrowth. + AdolFertRate + FertRate + InflConsPric. +
##      LifeExpBirth + MobileSubs.100 + MortRateU5 + NetMigr + UrbanPopGrowth.
##
##      Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1      200 3935.0
## 2      191 3696.5   9    238.49 1.3692 0.2045
```

```
# Prueba de Park
```

```
lnugorro<-log(ugorroabs)
plot(ygorro,lnugorro,type="p",col="red",
     main="Método gráfico de la Prueba de Park")
```



```
aux4h<-lm(lnugorro~PopGrowth.+AdolFertRate+FertRate+InflConsPric.+LifeExpBirth+MobileSubs.100+MortRateU5)
H04h<-c("PopGrowth.=0", "InflConsPric.=0", "FertRate=0", "LifeExpBirth=0", "AdolFertRate=0", "MobileSubs.100=0")
linearHypothesis(aux4h,H04h)
```

```
## Linear hypothesis test
##
## Hypothesis:
## PopGrowth. = 0
## InflConsPric. = 0
## FertRate = 0
## LifeExpBirth = 0
## AdolFertRate = 0
## MobileSubs.100 = 0
## MortRateU5 = 0
## NetMigr = 0
## UrbanPopGrowth. = 0
##
## Model 1: restricted model
```

```
## Model 2: lnugorro ~ PopGrowth. + AdolFertRate + FertRate + InflConsPric. +
##      LifeExpBirth + MobileSubs.100 + MortRateU5 + NetMigr + UrbanPopGrowth.
##
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1     200 292.16
## 2     191 278.47   9    13.694 1.0436 0.4069
```

Ambas pruebas nos muestran los mismos resultados que la prueba de Breusch, Pagan y Godfrey, el p-value es mayor a 0.1, por lo que rechazo que haya Heteroscedasticidad en el modelo. Pero, ¿por qué no intentamos con la prueba más general

Prueba de White

```
white(reg1,interactions = TRUE)
```

```
## # A tibble: 1 x 5
##   statistic p.value parameter method      alternative
##   <dbl>    <dbl>    <dbl> <chr>      <chr>
## 1      63.8    0.170        54 White's Test greater
```

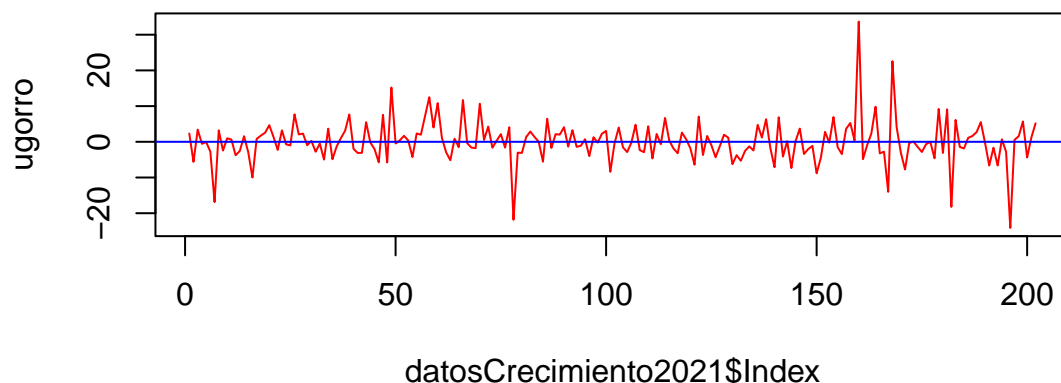
La prueba de White utiliza una regresión auxiliar más general, pues incluye como regresores no solo las variables originales lineales, sino también de manera cuadrática y de orden dos. En este caso el p-value confirma que el modelo no tiene un problema de Heteroscedasticidad.

Ahora bien, necesitamos comprobar que no existe autocorrelación entre nuestros residuales para asegurarnos que no se viola el quinto supuesto del modelo de regresión lineal múltiple.

No Autocorrelación

Método Gráfico

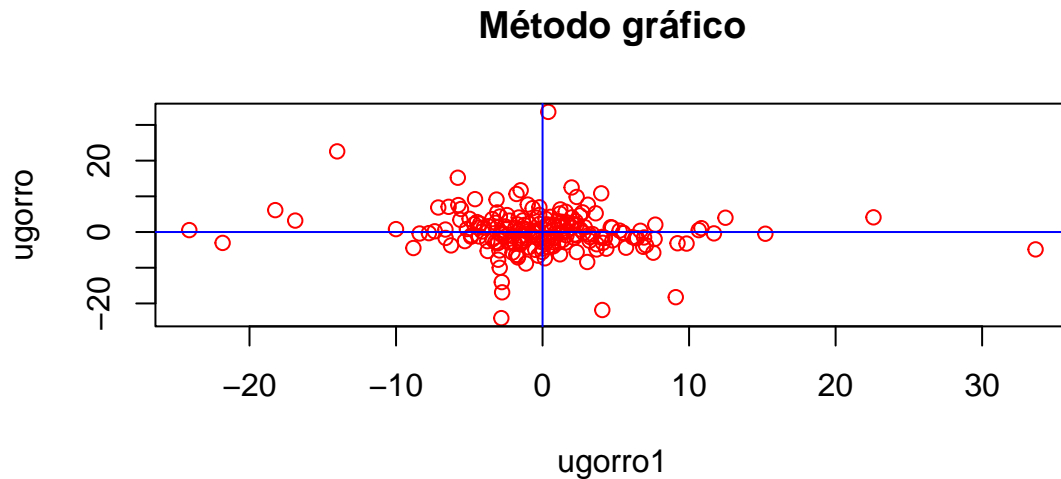
```
# Método gráfico
plot(datosCrecimiento2021$Index,ugorro,type="l",col="red")
abline(h=0, col="blue")
```




```

ugorro1<-vector()
for(i in 2:n){ugorro1[i]<-ugorro[i-1]}
plot(ugorro1,ugorro,type="p",col="red",main="Método gráfico")
abline(h=0,col="blue")
abline(v=0,col="blue")

```



Estos dos gráficos nos muestran que el término estocástico del modelo no sigue ningún tipo de patrón. El diagrama de dispersión revela el hecho de que los residuos se comportan como un término esférico.

Prueba de Durbin - Watson

```
# Prueba de Durbin-Watson
```

```
dwtest(reg1)
```

```

##
## Durbin-Watson test
##
## data: reg1
## DW = 2.2627, p-value = 0.9671
## alternative hypothesis: true autocorrelation is greater than 0

```

El valor p nos arroja un número considerablemente grande. Dicho esto, no se rechaza la hipótesis nula, la cual se refiere a que el modelo no presenta autocorrelación de primer orden.

Prueba de Breusch y Godfrey

```
# Prueba de Breusch y Godfrey
```

```
bgtest(reg1)
```

```

##
## Breusch-Godfrey test for serial correlation of order up to 1
##
## data: reg1
## LM test = 3.8292, df = 1, p-value = 0.05037

```

Si bien el valor p es muy parecido a nuestro nivel de significancia, si es mayor. De este modo, no se rechaza la hipótesis nula. Nuestro modelo no presenta autocorrelación de orden superior.

Verificación de Exogeneidad

Probemos si cumple el supuesto de Endogeneidad con el Reset Test.

```
# *****  
#      Supuesto: Exogeneidad  
# *****  
  
resettest(reg1, power=2, type="fitted")  
  
##  
## RESET test  
##  
## data:  reg1  
## RESET = 1.1647, df1 = 1, df2 = 190, p-value = 0.2819
```

En este caso, el valor p es mayor que el nivel de significancia del modelo. Por esta razón, ¡nuestro modelo parece estar bien especificado!

Corrección del modelo

Corrección 1. Modelo ascendente

Elaboramos diferentes modelos para analizar la significancia, de cada una de las diferentes variables. Esto se llevó a cabo debido a que la correlación entre algunas variables era considerablemente alta, como en el caso del crecimiento poblacional y el crecimiento de la población urbana. Las regresoras de tasa de fertilidad y tasa de fertilidad adolescente también presentan una correlación elevada entre ellas.

Por otro lado, es importante aclarar que decidimos eliminar la observación que correspondía al país de Cuba, por lo que, ahora solo son 201 observaciones. La razón de esta decisión está relacionada con el hecho de que Cuba se comportaba como un dato atípico. Decidimos eliminar primero a variable del crecimiento poblacional y la de inflación ya que su significancia era muy pequeña y estaba altamente correlacionada con el crecimiento de la población urbana, la cual sí era significativa. Luego volvimos a correr otra serie de modelos ascendentes para saber cual era el más adecuado.

```
RegAux1 <- lm (GDPGrowth.~MobileSubs.100, data=datosCrecimiento2021)  
RegAux2 <- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.,data=datosCrecimiento2021)  
RegAux3 <- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.+FertRate,data=datosCrecimiento2021)  
RegAux4 <- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.+FertRate+LifeExpBirth, data=datosCrecimiento2021)  
RegAux5<- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.+FertRate+LifeExpBirth+MortRateU5, data=datosCrecimiento2021)  
RegAux6<- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.+FertRate+LifeExpBirth+MortRateU5+NetMigr, data=datosCrecimiento2021)  
RegAux7<- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.+FertRate+LifeExpBirth+MortRateU5+NetMigr+InflCor, data=datosCrecimiento2021)  
RegAux8<- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.+FertRate+LifeExpBirth+MortRateU5+NetMigr+InflCor, data=datosCrecimiento2021)  
RegAux9<- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.+FertRate+LifeExpBirth+MortRateU5+NetMigr+InflCor, data=datosCrecimiento2021)  
  
library (stargazer)  
lista <- list (RegAux1,RegAux2,RegAux3,RegAux4,RegAux5,RegAux6,RegAux7,RegAux8,RegAux9)  
  
stargazer(lista,type = "text")
```

```

##
## =====
##
## -----
##              (1)              (2)              (3)              (4)
## -----
## MobileSubs.100      0.046***      0.051***      0.038***      0.038***
##                   (0.011)      (0.012)      (0.013)      (0.013)
##
## UrbanPopGrowth.      0.290      1.144***      1.144***
##                   (0.279)      (0.393)      (0.393)
##
## FertRate      -1.713***
##                   (0.565)
##
## LifeExpBirth
##
## MortRateU5
##
## NetMigr
##
## InflConsPric.
##
## AdolFertRate
##
## PopGrowth.
##
## Constant      -0.115      -1.144      3.367      3.367
##                   (1.315)      (1.646)      (2.195)      (2.195)
## -----
## Observations      201      201      201      201
## R2      0.077      0.082      0.123      0.123
## Adjusted R2      0.073      0.073      0.110      0.110
## Residual Std. Error      6.074 (df = 199)      6.072 (df = 198)      5.951 (df = 197)      5.966 (df = 197)
## F Statistic      16.652*** (df = 1; 199)      8.868*** (df = 2; 198)      9.220*** (df = 3; 197)      6.880*** (df = 4; 196)
## =====
## Note:

```

El crecimiento de la población urbana no fue eliminado debido a que podemos analizar cómo su integración en el segundo modelo aumenta el R^2 . Esto quiere decir que las variaciones de esta regresora sí ayudan a explicar las variaciones del crecimiento económico. Del mismo modo, la tasa de fertilidad aumenta el R^2 ajustado en el tercer modelo.

Ahora bien, realizando todos estos modelos nos dimos cuenta que las regresiones que presentaban el R^2 mayor eran la sexta y séptima. Sin embargo, el R^2 ajustado de la última regresión disminuye con la presencia de la tasa de fertilidad adolescente. Dicho esto, optamos por el penúltimo modelo. Nuestro nuevo modelo goza de un R^2 y R^2 ajustado mayor.

```

RegAux1B <- lm (GDPGrowth.~MobileSubs.100, data=datosCrecimiento2021)
RegAux2B <- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.,data=datosCrecimiento2021)
RegAux3B <- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.+FertRate,data=datosCrecimiento2021)
RegAux4B <- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.+FertRate+LifeExpBirth, data=datosCrecimiento2021)
RegAux5B<- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.+FertRate+LifeExpBirth+MortRateU5, data=datosCrecimiento2021)
RegAux6B<- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.+FertRate+LifeExpBirth+MortRateU5+NetMigr, data=datosCrecimiento2021)
RegAux7B<- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.+FertRate+LifeExpBirth+MortRateU5+NetMigr+AdolFertRate, data=datosCrecimiento2021)

listaB <- list(RegAux1B,RegAux2B,RegAux3B,RegAux4B,RegAux5B,RegAux6B,RegAux7B)

stargazer(listaB,type = "text")

```

```

##
## =====
##                                     Dependent
## -----
##                                     GDPGr
##                                     (4)
##      (1)          (2)          (3)
## -----
## MobileSubs.100      0.046***      0.051***      0.038***      0.038***
##                      (0.011)      (0.012)      (0.013)      (0.013)
##
## UrbanPopGrowth.      0.290      1.144***      1.144***
##                      (0.279)      (0.393)      (0.393)
##
## FertRate      -1.713***
##                (0.565)
##
## LifeExpBirth
##
## MortRateU5
##
## NetMigr
##
## AdolFertRate
##
## Constant      -0.115      -1.144      3.367      3.367
##                (1.315)      (1.646)      (2.195)      (2.195)
## -----
## Observations      201      201      201      201
## R2      0.077      0.082      0.123      0.123
## Adjusted R2      0.073      0.073      0.110      0.110
## Residual Std. Error      6.074 (df = 199)      6.072 (df = 198)      5.951 (df = 197)      5.966 (df = 197)
## F Statistic      16.652*** (df = 1; 199)      8.868*** (df = 2; 198)      9.220*** (df = 3; 197)      6.880*** (df = 3; 197)
## =====
## Note:

```

```
# Modelo Simplificado
stargazer(RegAux6B,type = "text")
```

```
##
## =====
##                      Dependent variable:
##                      -----
##                      GDPGrowth.
## -----
## MobileSubs.100          0.036***
##                        (0.013)
##
## UrbanPopGrowth.        1.088***
##                        (0.399)
##
## FertRate               -1.582**
##                        (0.760)
##
## LifeExpBirth           -0.033
##                        (0.102)
##
## MortRateU5             -0.010
##                        (0.032)
##
## NetMigr                 0.00001
##                        (0.00000)
##
## Constant                5.905
##                        (8.444)
##
## -----
## Observations            201
## R2                      0.133
## Adjusted R2             0.106
## Residual Std. Error     5.962 (df = 194)
## F Statistic             4.960*** (df = 6; 194)
## =====
## Note:                   *p<0.1; **p<0.05; ***p<0.01
```

Ahora, para comprobar como se desempeña este nuevo modelo, realizamos las pruebas para la verificación de supuestos.

```
modelo<-RegAux6B
jarque.bera.test(modelo$residuals)
```

```
##
## Jarque Bera Test
##
## data:  modelo$residuals
## X-squared = 528.11, df = 2, p-value < 2.2e-16
```

```
vif(modelo)
```

```
## MobileSubs.100 UrbanPopGrowth. FertRate LifeExpBirth MortRateU5
##      1.378299      2.538110      5.064774      3.376760      3.867005
##      NetMigr
##      1.071126
```

```
bptest(modelo)
```

```
##
## studentized Breusch-Pagan test
##
## data: modelo
## BP = 6.8619, df = 6, p-value = 0.3338
```

```
white(modelo, interactions = TRUE)
```

```
## # A tibble: 1 x 5
##   statistic p.value parameter method alternative
##   <dbl>    <dbl>    <dbl> <chr>      <chr>
## 1      26.1    0.511      27 White's Test greater
```

```
dwtest(modelo)
```

```
##
## Durbin-Watson test
##
## data: modelo
## DW = 2.2596, p-value = 0.966
## alternative hypothesis: true autocorrelation is greater than 0
```

```
bgtest(modelo)
```

```
##
## Breusch-Godfrey test for serial correlation of order up to 1
##
## data: modelo
## LM test = 3.6359, df = 1, p-value = 0.05654
```

```
resettest(modelo, power=2, type="fitted")
```

```
##
## RESET test
##
## data: modelo
## RESET = 1.0893, df1 = 1, df2 = 193, p-value = 0.2979
```

- **Prueba Jarque Bera:** El término aleatorio claramente sigue sin ser normal. El valor p es considerablemente menor que el nivel de significancia. La hipótesis nula es rechazada, por lo que, los residuales no se comportan de manera normal.

- **FIV:** La eliminación de la variables del crecimiento poblacional y la tasa de fertilidad adolescente ayudaron a que el factor que infla a la varianza disminuya. Ninguna de las regresoras presenta un factor mayor a 10.
- **Prueba Breush Pagan y Godfrey:** El valor p sigue siendo mayor que el nivel de significancia. No se rechaza la hipótesis nula. Por ello, el término estocástico no presenta heterocedasticidad.
- **Prueba Durbin-Watson:** El valor p es mayor que el nivel de significancia. Dicho esto, el término aleatorio no presenta autocorrelación de primer orden.
- **Prueba de especificación:** El valor continúa siendo mayor que el nivel de significancia. A pesar de haber eliminado variables y observaciones, el modelo está bien especificado.
- **Prueba de White:** En este caso el p- value confirma que el modelo no tiene un problema de Heteroscedasticidad cuando tomamos un modelo más general (términos cuadráticos y de orden dos).

Regresiones Auxiliares de Klein.

```
aux1m<-lm(MobileSubs.100~UrbanPopGrowth.+FertRate+LifeExpBirth+MortRateU5+NetMigr,data=datosCrecimiento)
aux2m<-lm(UrbanPopGrowth.~MobileSubs.100+FertRate+LifeExpBirth+MortRateU5+NetMigr,data=datosCrecimiento)
aux3m<-lm(FertRate~UrbanPopGrowth.+MobileSubs.100+LifeExpBirth+MortRateU5+NetMigr,data=datosCrecimiento)
aux4m<-lm(LifeExpBirth~UrbanPopGrowth.+FertRate+MobileSubs.100+MortRateU5+NetMigr,data=datosCrecimiento)
aux5m<-lm(MortRateU5~UrbanPopGrowth.+FertRate+LifeExpBirth+MobileSubs.100+NetMigr,data=datosCrecimiento)
aux6m<-lm(NetMigr~UrbanPopGrowth.+FertRate+LifeExpBirth+MortRateU5+MobileSubs.100,data=datosCrecimiento)

stargazer(aux1m,aux2m,aux3m,aux4m,aux5m,aux6m,type="text")
```

```
##
## =====
##                                     Dependent variable:
##                                     -----
##                                     MobileSubs.100 UrbanPopGrowth. FertRate LifeExpBirth MortRateU5
##                                     (1) (2) (3) (4) (5)
## -----
## UrbanPopGrowth. -0.379 0.237*** -0.238 1.056 9
##                  (2.195) (0.034) (0.280) (0.890) (6
##
## MobileSubs.100 -0.0004 -0.004*** 0.009 -0.001
##                  (0.002) (0.001) (0.009) (0.029) (
##
## FertRate -12.232*** 0.860*** -2.288*** 9.792*** -3
##            (4.088) (0.122) (0.509) (1.551) (11
##
## LifeExpBirth 0.532 -0.015 -0.041*** -1.263*** 1
##              (0.559) (0.018) (0.009) (0.209) (1
##
## MortRateU5 -0.008 0.007 0.017*** -0.125*** -
##             (0.176) (0.006) (0.003) (0.021) (4
##
## NetMigr 0.00003 0.00000 -0.00000 0.00000 -0.00001
##          (0.00003) (0.00000) (0.00000) (0.00000) (0.00001)
##
## Constant 103.949** 0.367 5.040*** 80.022*** 90.275*** -14
##           (45.851) (1.515) (0.709) (1.570) (17.773) (12
##
```

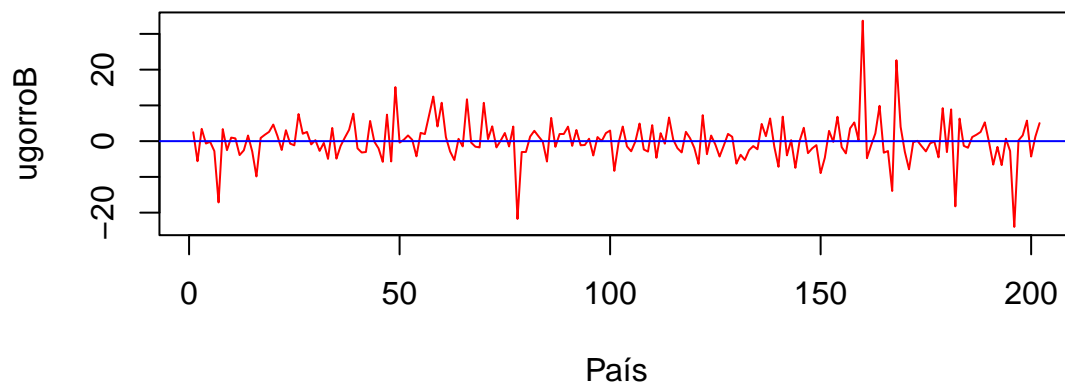
```
## -----
## Observations          201          201          201          201          201
## R2                    0.274          0.606          0.803          0.704          0.741
## Adjusted R2           0.256          0.596          0.797          0.696          0.735
## Residual Std. Error (df = 195) 32.799          1.070          0.562          4.195          13.354
## F Statistic (df = 5; 195) 14.754***        59.986***        158.526***    92.694***    111.813***
## =====
## Note:                                                         *p<0.1; **p<0.05;
```

A pesar de que el factor inflador de la varianza no muestra multicolinealidad, la regla de Klein nos permite darnos cuenta de que todos los R^2 de cada una de las regresiones auxiliares es mayor que el R^2 de nuestro modelo, el cual es de 0.133. Con base en lo anterior, es necesario corregir este problema.

```
k<-6
ygorroB<-RegAux6B$fitted.values
ugorroB<-RegAux6B$residuals
sigma2B<-sum(ugorroB^2)/(n-k-1)

plot(datosCrecimiento2021$Index,ugorroB,type="l",
     col="red",
     main="Gráfico de residuales",
     xlab = "País")
abline(h=0, col="blue")
```

Gráfico de residuales



No obstante, nuestros residuales continúan comportándose de buena manera. No presentan autocorrelación ni heterocedasticidad. Este gráfico nos muestra que los residuales varían sobre un mismo rango y que no existe ningún patrón.

Corrección 2. Modelos ascendentes para corregir multicolinealidad

En este caso, decidimos realizar cuatro diferentes regresiones para intentar eliminar el problema de la multicolinealidad. Desechamos en tres de estas regresiones algunas variables que considerábamos que podía estar ocasionando el problema de la correlación entre las regresoras. En la primera regresión, eliminamos la tasa de mortalidad infantil de menores a cinco años. La segunda regresión se realiza sin tomar en cuenta la

esperanza de vida al nacer. Por otro lado, la tercera regresión no cuenta con la migración neta. De este modo, podemos observar que el R^2 mayor es el del tercer modelo; sin embargo su R^2 ajustado es menor que el de la primera o segunda regresora. Esto nos indica que los otros dos modelos son mejores opciones.

```
RegAux1C<- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.+FertRate+LifeExpBirth+NetMigr, data=datosCrecim
RegAux2C<- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.+FertRate+NetMigr+MortRateU5, data=datosCrecim
RegAux3C<- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.+FertRate+LifeExpBirth+MortRateU5, data=datosCrec
RegAux4C<- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.+FertRate+LifeExpBirth+MortRateU5+NetMigr, data=

listaC <- list(RegAux1C,RegAux2C,RegAux3C,RegAux4C)
stargazer(listaC,type = "text")
```

```
##
## =====
##                                     Dependent variable:
## -----
##                                     GDPGrowth.
##          (1)                (2)                (3)                (4)
## -----
## MobileSubs.100              0.036***              0.036***              0.038***              0.036***
##                               (0.013)              (0.013)              (0.013)              (0.013)
##
## UrbanPopGrowth.             1.077***              1.095***              1.155***              1.088***
##                               (0.397)              (0.397)              (0.397)              (0.397)
##
## FertRate                    -1.675**              -1.505**              -1.608**              -1.588**
##                               (0.691)              (0.722)              (0.762)              (0.762)
##
## LifeExpBirth                 -0.021              -0.021              -0.021              -0.021
##                               (0.093)              (0.102)              (0.102)              (0.102)
##
## NetMigr                      0.00001              0.00001              0.00001              0.00001
##                               (0.00000)              (0.00000)              (0.00000)              (0.00000)
##
## MortRateU5                  -0.005              -0.013              -0.013              -0.013
##                               (0.029)              (0.032)              (0.032)              (0.032)
##
## Constant                     5.039              3.231              4.906              5.906
##                               (7.917)              (2.227)              (8.438)              (8.438)
##
## -----
## Observations                  201              201              201              201
## R2                           0.133              0.133              0.124              0.133
## Adjusted R2                   0.110              0.110              0.101              0.110
## Residual Std. Error    5.949 (df = 195)    5.949 (df = 195)    5.979 (df = 195)    5.962 (df = 195)
## F Statistic              5.962*** (df = 5; 195)  5.958*** (df = 5; 195)  5.512*** (df = 5; 195)  5.960*** (df = 5; 195)
## =====
## Note:                                     *p<0.1; **p<0.05
```

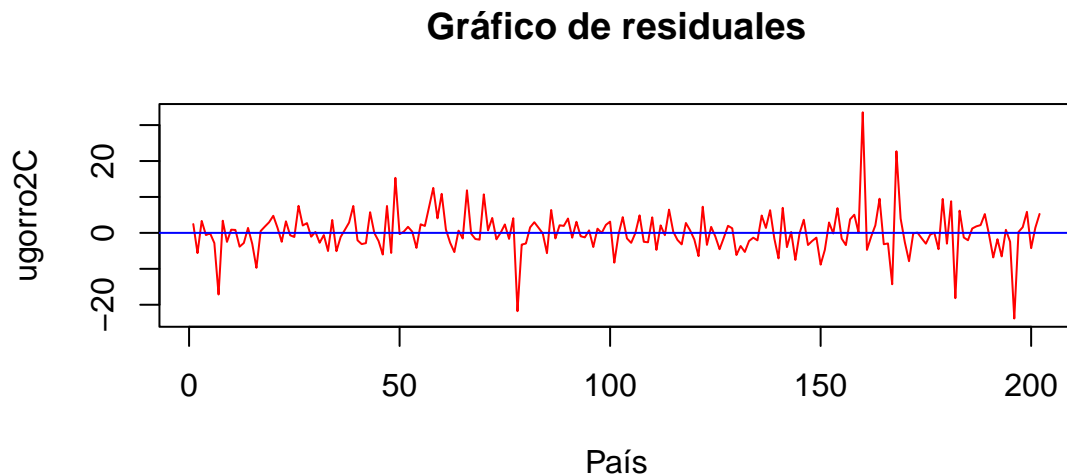
Ahora bien, la primera regresión no mostró buenos resultados en el supuesto de homocedasticidad. El modelo tenía un valor p menor que el nivel de significancia en la prueba de Breusch, Godfrey y Pagan. Dicho esto, optamos por elegir el segundo modelo. Esta regresión ahora no incluye a la esperanza de vida al nacer. Asimismo, esta variable estaba altamente correlacionada con la tasa de fertilidad, la cual es significativa en este nuevo modelo.

```

k<-5
ygorro2C<-RegAux2C$fitted.values
ugorro2C<-RegAux2C$residuals
sigma2C<-sum(ugorro2C^2)/(n-k-1)

plot(datosCrecimiento2021$Index,ugorro2C,type="l",
     col="red",
     main="Gráfico de residuales",
     xlab = "País")
abline(h=0, col="blue")

```



Los residuales de nuestra nueva regresión se siguen comportando de buena manera. no presentan ningún patrón y su variabilidad está sujeta a un mismo rango.

Recordemos el modelo que hasta el momento se ha desempeñado mejor y apliquemos las pruebas para comprobar los supuestos.

```

RegAlmost <- RegAux2C
stargazer(RegAlmost,type = "text")

```

```

##
## =====
##                               Dependent variable:
##                               -----
##                               GDPGrowth.
##                               -----
## MobileSubs.100                0.036***
##                               (0.013)
##
## UrbanPopGrowth.              1.095***
##                               (0.397)
##
## FertRate                     -1.505**
##                               (0.722)
##

```

```
## NetMigr                0.00001
##                        (0.00000)
##
## MortRateU5             -0.005
##                        (0.029)
##
## Constant               3.231
##                        (2.227)
##
## -----
## Observations           201
## R2                     0.133
## Adjusted R2            0.110
## Residual Std. Error    5.949 (df = 195)
## F Statistic            5.958*** (df = 5; 195)
## =====
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

```
modelo<-RegAlmost
jarque.bera.test(modelo$residuals)
```

```
##
## Jarque Bera Test
##
## data:  modelo$residuals
## X-squared = 516.33, df = 2, p-value < 2.2e-16
```

```
vif(modelo)
```

```
## MobileSubs.100 UrbanPopGrowth.      FertRate      NetMigr      MortRateU5
##      1.371907      2.528761      4.589361      1.063175      3.258038
```

```
bptest(modelo)
```

```
##
## studentized Breusch-Pagan test
##
## data:  modelo
## BP = 6.9024, df = 5, p-value = 0.228
```

```
white(modelo,interactions = TRUE)
```

```
## # A tibble: 1 x 5
##   statistic p.value parameter method      alternative
##   <dbl>    <dbl>    <dbl> <chr>      <chr>
## 1      20.1    0.454      20 White's Test greater
```

```
dwtest(modelo)
```

```
##
```

```
## Durbin-Watson test
##
## data: modelo
## DW = 2.2596, p-value = 0.9666
## alternative hypothesis: true autocorrelation is greater than 0
```

```
bgtest(modelo)
```

```
##
## Breusch-Godfrey test for serial correlation of order up to 1
##
## data: modelo
## LM test = 3.6266, df = 1, p-value = 0.05686
```

```
resettest(modelo, power=2, type="fitted")
```

```
##
## RESET test
##
## data: modelo
## RESET = 1.0796, df1 = 1, df2 = 194, p-value = 0.3001
```

- **Prueba Jarque Bera:** El valor p continúa siendo menor que el nivel de significancia. Por ello, nuestro término estocástico no se distribuye de manera normal.
- **FIV:** El factor inflador de la varianza continúa siendo menor que diez para las cinco regresoras que tenemos. Sin embargo, tenemos que aplicar la regla de Klein para asegurarnos de que no exista multicolinealidad imperfecta, pero elevada.
- **Prueba Breush Pagan y Godfrey:** El valor p es muy parecido a nuestro nivel de significancia, pero sigue siendo mayor. Nuestro modelo no presenta autocorrelación de orden superior.
- **Prueba Durbin-Watson:** El valor p es mayor que nuestro nivel de significancia, el es 0.05. Dicho esto, nuestro modelo no presenta autocorrelación de primer orden.
- **Prueba de especificación:** El valor p es mayor que nuestro nivel de significancia. A pesar de haber eliminado la variable de esperanza de vida al nacer, el modelo se encuentra bien especificado.
- **Prueba de White:** En este caso el p- value confirma que el modelo no tiene un problema de Heteroscedasticidad cuando tomamos un modelo más general (términos cuadráticos y de orden dos).

Prueba de Klein Modelo RegAlmost

```
aux1m<-lm(MobileSubs.100~UrbanPopGrowth.+FertRate+MortRateU5+NetMigr,data=datosCrecimiento2021)
aux2m<-lm(UrbanPopGrowth.~MobileSubs.100+FertRate+MortRateU5+NetMigr,data=datosCrecimiento2021)
aux3m<-lm(FertRate~UrbanPopGrowth.+MobileSubs.100+MortRateU5+NetMigr,data=datosCrecimiento2021)
aux4m<-lm(MortRateU5~UrbanPopGrowth.+FertRate+MobileSubs.100+NetMigr,data=datosCrecimiento2021)
aux5m<-lm(NetMigr~UrbanPopGrowth.+FertRate+MortRateU5+MobileSubs.100,data=datosCrecimiento2021)

stargazer(aux1m,aux2m,aux3m,aux4m,aux5m,type="text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               MobileSubs.100 UrbanPopGrowth. FertRate MortRateU5 NetMigr
##                               (1) (2) (3) (4) (5)
## -----
## UrbanPopGrowth. -0.508 0.272*** 1.610* 9,567.850
## (2.190) (0.034) (0.962) (6,035.595)
##
## MobileSubs.100 -0.001 -0.004*** -0.015 209.662
## (0.002) (0.001) (0.032) (197.535)
##
## FertRate -13.513*** 0.899*** 15.052*** -8,252.685
## (3.860) (0.113) (1.394) (11,017.940)
##
## MortRateU5 -0.075 0.009* 0.025*** -685.337
## (0.161) (0.005) (0.002) (444.970)
##
## NetMigr 0.00003 0.00000 -0.00000 -0.00002
## (0.00003) (0.00000) (0.00000) (0.00001)
##
## Constant 147.235*** -0.875** 1.938*** -12.820** 1,016.317
## (6.327) (0.395) (0.171) (5.354) (34,039.400)
##
## -----
## Observations 201 201 201 201 201
## R2 0.271 0.605 0.782 0.693 0.059
## Adjusted R2 0.256 0.596 0.778 0.687 0.040
## Residual Std. Error (df = 196) 32.791 1.069 0.589 14.511 90,943.890
## F Statistic (df = 4; 196) 18.223*** 74.909*** 175.879*** 110.644*** 3.096**
## =====
## Note: *p<0.1; **p<0.05; ***p<0.01
```

A pesar de haber eliminado la esperanza de vida al nacer, nuestro modelo continúa presentando multicolinealidad. Sin embargo, la migración neta tiene un R^2 de 0.059, el cual es menor que el R^2 de nuestro modelo. Las otras cuatro regresoras tienen un R^2 mayor que el del modelo original.

Corrección 3. Regresiones por Continentes

Dado que en nuestro modelo aún persiste el problema de multicolinealidad, decidimos correr nuestra última regresión, pero ahora para cada continente. Anteriormente, eliminamos algunas variables que disminuían nuestro R^2 ajustado y la significancia de otras variables y con ello, el modelo mejoró. Esto lo hicimos considerando todas los países. No obstante, creemos que al realizar varios modelos (uno por continente) nuestra R^2 puede mejorar. Cabe mencionar que al separar por regiones, hay modelos que no cumplen con el número mínimo de observaciones, tal es el caso de Oceanía que tiene sólo 16. De hecho el único continente que cumple con el número de observaciones recomendado es África. El resto de regiones tiene un número de observaciones cercano al recomendable.

```
datosCrecimientoEuropa <- subset(datosCrecimiento2021, Region == "Europe")

RegAlmostEu <- lm (datosCrecimientoEuropa$GDPPGrowth.~
                  datosCrecimientoEuropa$MobileSubs.100+
```

```

        datosCrecimientoEuropa$UrbanPopGrowth.+
        datosCrecimientoEuropa$FertRate+
        datosCrecimientoEuropa$NetMigr+
        datosCrecimientoEuropa$MortRateU5,
        data=datosCrecimiento2021)

datosCrecimientoAsia <- subset(datosCrecimiento2021, Region == "Asia")

RegAlmostAs<- lm (datosCrecimientoAsia$GDPGrowth.~
        datosCrecimientoAsia$MobileSubs.100+
        datosCrecimientoAsia$UrbanPopGrowth.+
        datosCrecimientoAsia$FertRate+
        datosCrecimientoAsia$NetMigr+
        datosCrecimientoAsia$MortRateU5,
        data=datosCrecimiento2021)

datosCrecimientoAmericas <- subset(datosCrecimiento2021, Region == "Americas")

RegAlmostAme<- lm (datosCrecimientoAmericas$GDPGrowth.~
        datosCrecimientoAmericas$MobileSubs.100+
        datosCrecimientoAmericas$UrbanPopGrowth.+
        datosCrecimientoAmericas$FertRate+
        datosCrecimientoAmericas$NetMigr+
        datosCrecimientoAmericas$MortRateU5,
        data=datosCrecimiento2021)

datosCrecimientoAfrica <- subset(datosCrecimiento2021, Region == "Africa")

RegAlmostAfri <- lm (datosCrecimientoAfrica$GDPGrowth.~
        datosCrecimientoAfrica$MobileSubs.100+
        datosCrecimientoAfrica$UrbanPopGrowth.+
        datosCrecimientoAfrica$FertRate+
        datosCrecimientoAfrica$NetMigr+
        datosCrecimientoAfrica$MortRateU5,
        data=datosCrecimiento2021)

datosCrecimientoOceania <- subset(datosCrecimiento2021, Region == "Oceania")

RegAlmostOcea <- lm (datosCrecimientoOceania$GDPGrowth.~
        datosCrecimientoOceania$MobileSubs.100+
        datosCrecimientoOceania$UrbanPopGrowth.+
        datosCrecimientoOceania$FertRate+
        datosCrecimientoOceania$NetMigr+
        datosCrecimientoOceania$MortRateU5,
        data=datosCrecimiento2021)

stargazer(RegAlmostEu,RegAlmostAs,RegAlmostAme,RegAlmostAfri,RegAlmostOcea,type="text")

```

```

##
## =====
##                                     Dependent variable:
## -----
##          GDPGrowth.          GDPGrowth.          GDPGrowth.          GDPGrowth.
##

```

##	(1)	(2)	(3)	(4)
## -----				
## MobileSubs.100	0.002			
##	(0.035)			
##				
## UrbanPopGrowth.	-0.829			
##	(0.576)			
##				
## FertRate	3.202			
##	(2.421)			
##				
## NetMigr	-0.00001			
##	(0.00001)			
##				
## MortRateU5	-0.242**			
##	(0.102)			
##				
## MobileSubs.100		0.020		
##		(0.035)		
##				
## UrbanPopGrowth.		1.760*		
##		(0.914)		
##				
## FertRate		-2.810		
##		(1.934)		
##				
## NetMigr		-0.00001		
##		(0.00001)		
##				
## MortRateU5		-0.143		
##		(0.100)		
##				
## MobileSubs.100			0.032	
##			(0.030)	
##				
## UrbanPopGrowth.			4.971***	
##			(1.409)	
##				
## FertRate			-5.555**	
##			(2.701)	
##				
## NetMigr			0.00001	
##			(0.00001)	
##				
## MortRateU5			-0.010	
##			(0.086)	
##				
## MobileSubs.100				0.060***
##				(0.018)
##				
## UrbanPopGrowth.				0.756
##				(0.947)
##				
## FertRate				-0.864

```

##                                                                    (1.245)
##                                                                    (0.00003)
## NetMigr                                                            -0.00002
##                                                                    (0.00003)
## MortRateU5                                                         -0.001
##                                                                    (0.035)
## MobileSubs.100
##
## UrbanPopGrowth.
##
## FertRate
##
## NetMigr
##
## MortRateU5
##
## Constant                    2.711                    8.447                    8.474                    -0.181
##                             (5.786)                   (7.487)                   (5.523)                   (3.714)
## -----
## Observations                 42                     49                     41                     53
## R2                          0.208                   0.156                   0.398                   0.293
## Adjusted R2                 0.098                   0.057                   0.312                   0.218
## Residual Std. Error  3.477 (df = 36)    7.938 (df = 43)    5.830 (df = 35)    4.180 (df = 47)
## F Statistic           1.890 (df = 5; 36) 1.585 (df = 5; 43) 4.636*** (df = 5; 35) 3.895*** (df = 5; 47)
## =====
## Note:                                                                *p<0.1;

```

Los resultados de este experimento fueron los siguientes: en el caso de Europa, el R^2 ajustado disminuyó a 0.098 y sólo la tasa de mortalidad fue significativa; en el modelo de Asia, el R^2 también disminuyó a 0.057 y sólo la tasa de crecimiento de la población fue ligeramente significativa; en el caso de América obtuvimos el mejor modelo, pues el R^2 ajustado fue de 0.312 y obtuvimos dos variables significativas; en África el modelo también mejoró, porque el R^2 ajustado fue igual a 0.218 y una variable resultó significativa; finalmente, el modelo de Oceanía fracasó, porque no obtuvimos ninguna R^2 ajustada. Creemos que esto último se debe a que no tenemos el número necesario de observaciones.

Posteriormente, estudiamos los residuales de algunas regresiones. Particularmente, hicimos un gráfico de residuales para Oceanía, África y Europa. Estos mostraron autocorrelación entre los residuales y heterocedasticidad, pues la dispersión de los residuales no parece ser homogénea.

```

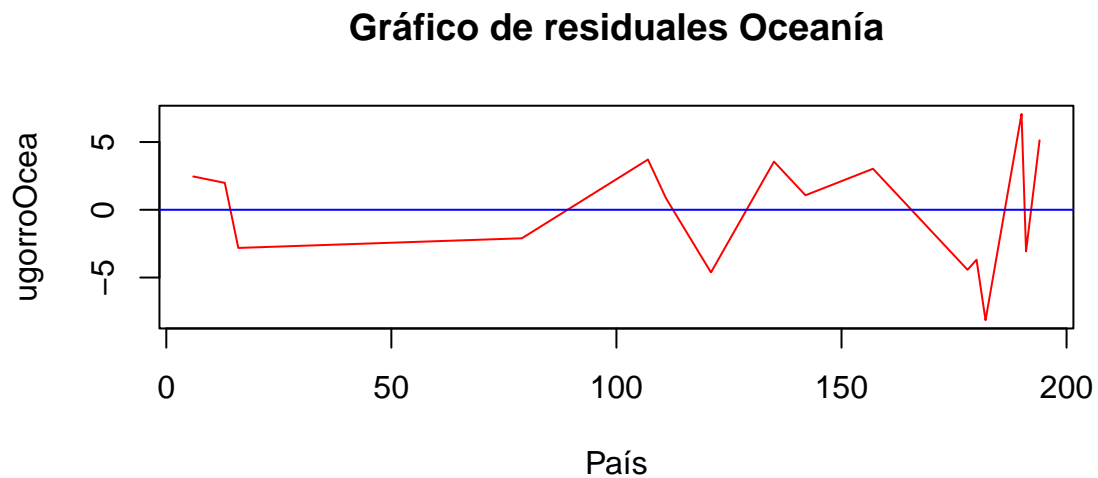
k<-5
n2 = sum(datosCrecimientoOceania$Index)
ygorroOcea<-RegAlmostOcea$fitted.values
ugorroOcea<-RegAlmostOcea$residuals
sigmaOcea<-sum(ugorroOcea^2)/(n2-k-1)

plot(datosCrecimientoOceania$Index,ugorroOcea,type="l",

```



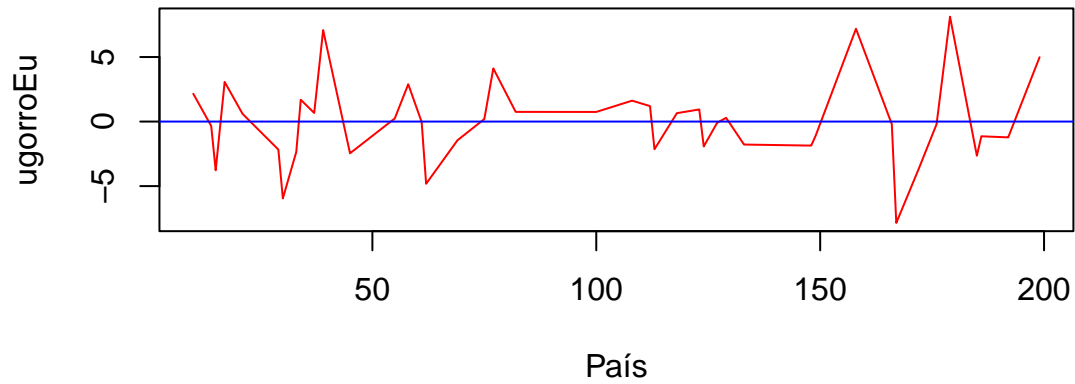
```
col="red",
main="Gráfico de residuales Oceanía",
xlab = "País")
abline(h=0, col="blue")
```



```
k<-5
nEurope = sum(datosCrecimientoEuropa$Index)
ygorroEu<-RegAlmostEu$fitted.values
ugorroEu<-RegAlmostEu$residuals
sigmaEu<-sum(ugorroEu^2)/(nEurope-k-1)

plot(datosCrecimientoEuropa$Index,ugorroEu,type="l",
col="red",
main="Gráfico de residuales Europe",
xlab = "País")
abline(h=0, col="blue")
```

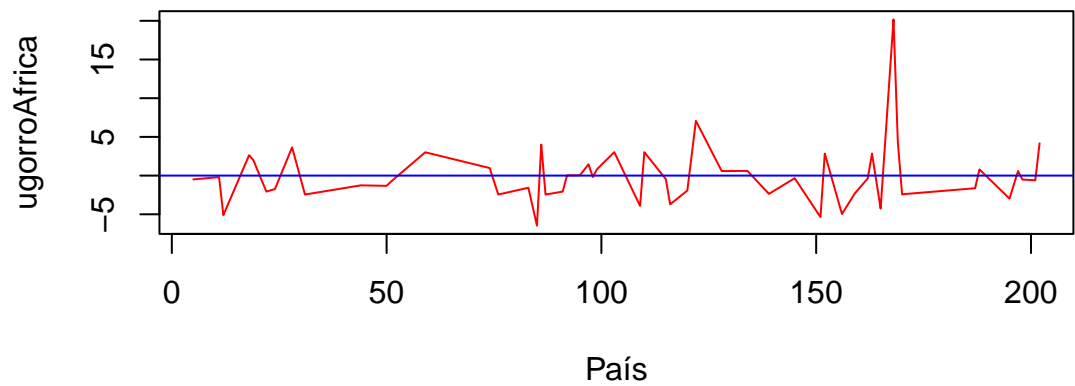
Gráfico de residuales Europe



```
k<-5
nAfrica = sum(datosCrecimientoAfrica$Index)
ygorroAfrica<-RegAlmostAfri$fitted.values
ugorroAfrica<-RegAlmostAfri$residuals
sigmaAfrica<-sum(ugorroAfrica^2)/(nAfrica-k-1)

plot(datosCrecimientoAfrica$Index,ugorroAfrica,type="l",
     col="red",
     main="Gráfico de residuales Africa",
     xlab = "País")
abline(h=0, col="blue")
```

Gráfico de residuales Africa



Corrección 4. NetMigr Dicotómica

Por lo anterior, concluimos que realizar un modelo para cada región no es conveniente y en su lugar, nos dispusimos a hacer una regresión dicotómica, en la que tomamos a la migración neta como variable categórica. Antes de continuar es importante mencionar que la migración neta se define como número de inmigrantes menos número de emigrantes. Así, si la migración neta de un país es negativa, la observación se codifica como 1 y si es positiva como 0.

```
(dico1<-(NetMigr<0)*1)
```

```
## [1] 1 1 1 1 0 1 1 1 1 0 1 0 1 0 0 1 0 1 1 1 0 0 1 1 0 1 0 0 0 0 1 0 0 1 0 1 0
## [38] 0 0 1 0 1 1 1 0 0 1 1 1 1 1 1 1 1 0 0 1 0 0 0 0 1 1 1 0 0 0 0 0 1 0 0 0 0
## [75] 1 1 1 0 0 1 1 1 0 1 1 1 0 1 0 0 1 0 0 1 1 1 1 1 1 1 0 0 0 0 1 1 0 0 1 1 0
## [112] 0 0 1 0 1 0 1 1 1 1 0 1 1 1 0 1 1 1 1 1 0 0 1 0 1 1 1 0 0 1 1 0 1 1 0 0 0
## [149] 1 1 1 1 0 1 1 1 0 1 0 0 1 1 0 1 0 0 1 1 1 1 1 1 1 0 1 1 1 1 0 0 0 1 1 0 1
## [186] 1 1 1 0 1 0 0 1 1 1 1 1 1 1 0 1 1
```

```
dico1
```

```
## [1] 1 1 1 1 0 1 1 1 1 0 1 0 1 0 0 1 0 1 1 1 0 0 1 1 0 1 0 0 0 0 1 0 0 1 0 1 0
## [38] 0 0 1 0 1 1 1 0 0 1 1 1 1 1 1 1 1 0 0 1 0 0 0 0 1 1 1 0 0 0 0 0 1 0 0 0 0
## [75] 1 1 1 0 0 1 1 1 0 1 1 1 0 1 0 0 1 0 0 1 1 1 1 1 1 1 0 0 0 0 1 1 0 0 1 1 0
## [112] 0 0 1 0 1 0 1 1 1 1 0 1 1 1 0 1 1 1 1 1 0 0 1 0 1 1 1 0 0 1 1 0 1 1 0 0 0
## [149] 1 1 1 1 0 1 1 1 0 1 0 0 1 1 0 1 0 0 1 1 1 1 1 1 1 0 1 1 1 1 0 0 0 1 1 0 1
## [186] 1 1 1 0 1 0 0 1 1 1 1 1 1 1 0 1 1
```

```
regDic <- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.+FertRate+NetMigr+MortRateU5+dico1, data=datosCr
stargazer(regDic,type="text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               GDPGrowth.
## -----
## MobileSubs.100                0.036***
##                               (0.013)
##
## UrbanPopGrowth.              1.086***
##                               (0.399)
##
## FertRate                     -1.458**
##                               (0.738)
##
## NetMigr                      0.00001
##                               (0.00001)
##
## MortRateU5                   -0.006
##                               (0.029)
##
## dico1                        -0.324
```

```
## (0.999)
##
## Constant 3.348
## (2.261)
##
## -----
## Observations 201
## R2 0.133
## Adjusted R2 0.106
## Residual Std. Error 5.963 (df = 194)
## F Statistic 4.960*** (df = 6; 194)
## =====
## Note: *p<0.1; **p<0.05; ***p<0.01
```

Corrección 5. Múltiples dicotómicas

Finalmente, desarrollamos una última regresión dicotómica en la que controlamos por región. En especial, elegimos a África como control y observamos el efecto individual de cada continente sobre el crecimiento económico.

```
res <- model.matrix(~Region, data = datosCrecimiento2021)
```

```
regDic2 <- lm (GDPGrowth.~MobileSubs.100+UrbanPopGrowth.+FertRate+NetMigr+MortRateU5+res, data=datosCrecimiento2021)
stargazer(regDic2,type="text")
```

```
##
## =====
## Dependent variable:
## -----
## GDPGrowth.
## -----
## MobileSubs.100 0.028**
## (0.014)
##
## UrbanPopGrowth. 0.994**
## (0.404)
##
## FertRate -1.361*
## (0.783)
##
## NetMigr 0.00001
## (0.00000)
##
## MortRateU5 -0.028
## (0.030)
##
## res(Intercept)
##
##
## resRegionAmericas -0.414
## (1.729)
```

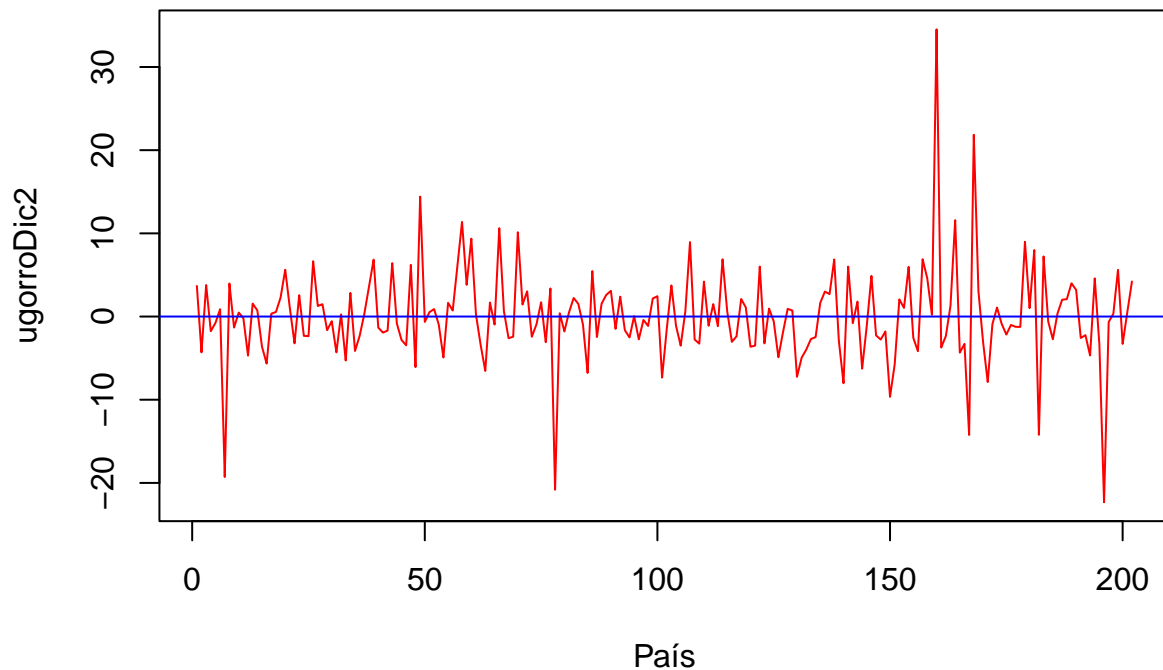
```
##
## resRegionAsia          -2.316
##                        (1.549)
##
## resRegionEurope        -1.090
##                        (1.884)
##
## resRegionOceania       -5.392***
##                        (1.982)
##
## Constant               5.803*
##                        (3.357)
##
## -----
## Observations           201
## R2                     0.181
## Adjusted R2            0.143
## Residual Std. Error    5.839 (df = 191)
## F Statistic            4.706*** (df = 9; 191)
## =====
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

Los resultados de la regresión fueron prometedores, porque el R^2 ajustado fue igual a 0.143 y en particular, fue mayor al obtenido en cualquiera de las regresiones anteriores. No obstante, cuando observamos el efecto de Oceanía sobre la variable dependiente vemos que el coeficiente es negativo y significativo. Esto nos parece particularmente extraño tanto porque el número de observaciones de Oceanía es sumamente pequeño como para realizar conclusiones.

```
k<-6
ygorroDic2<-regDic2$fitted.values
ugorroDic2<-regDic2$residuals
sigma2C<-sum(ugorro2C^2)/(n-k-1)

plot(datosCrecimiento2021$Index,ugorroDic2,type="l",
     col="red",
     main="Gráfico de residuales",
     xlab = "País")
abline(h=0, col="blue")
```

Gráfico de residuales



Hagamos por último las pruebas de supuestos:

```
modelo<-regDic2
jarque.bera.test(modelo$residuals)
```

```
##
## Jarque Bera Test
##
## data: modelo$residuals
## X-squared = 675.54, df = 2, p-value < 2.2e-16
```

```
#vif(modelo)
bptest(modelo)
```

```
##
## studentized Breusch-Pagan test
##
## data: modelo
## BP = 10.348, df = 9, p-value = 0.3231
```

```
white(modelo,interactions = TRUE)
```

```
## # A tibble: 1 x 5
##   statistic p.value parameter method      alternative
##   <dbl>    <dbl>    <dbl> <chr>      <chr>
## 1     38.6    0.944      54 White's Test greater
```

```
dwtest(modelo)
```

```
##  
## Durbin-Watson test  
##  
## data: modelo  
## DW = 2.2255, p-value = 0.762  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
bgtest(modelo)
```

```
##  
## Breusch-Godfrey test for serial correlation of order up to 1  
##  
## data: modelo  
## LM test = 2.7749, df = 1, p-value = 0.09575
```

```
resettest(modelo, power=2, type="fitted")
```

```
##  
## RESET test  
##  
## data: modelo  
## RESET = 0.48986, df1 = 1, df2 = 189, p-value = 0.4849
```

- **Prueba Jarque Bera:** El valor p continúa es menor que el nivel de significancia. Por ello, nuestro término estocástico no se distribuye de manera normal.
- **Prueba Breush Pagan y Godfrey:** En esta prueba el valor p es igual a 0.3231 y por lo tanto, es mayor que el nivel de significancia. Entonces, no hay heterocedasticidad en el modelo.
- **Prueba Durbin-Watson:** El valor p es igual a 0.762 y , por ende, es mayor que nuestro nivel de significancia 0.05. Entonces, nuestro modelo no presenta autocorrelación de primer orden.
- **Prueba de Breusch y Godfrey:** El valor p es igual a 0.09575 y en consecuencia, mayor a nuestro nivel de significancia. Cabe mencionar que en este modelo, la diferencia entre el valor p y el nivel de significancia es la más grande que hemos encontrado. Y en consecuencia, nuestro modelo no presenta autocorrelación de orden superior.
- **Prueba de especificación:** El valor p es mayor que el nivel de significancia, ya que es igual a 0.4849. Y de esta manera, el modelo está bien especificado.
- **Prueba de White:** En este caso el p- value confirma que el modelo no tiene un problema de Heteroscedasticidad.

Conclusiones

Ahora bien, nuestro modelo aún presenta multicolinealidad; sin embargo, no cuenta con la presencia de la heterocedasticidad ni de la autocorrelación. La variable dicotómica que está controlada por la región, o bien el continente, a la que pertenecían los países si aumentó el R^2 ajustado. Como ya se mencionó, estos resultados no nos convencieron, ya que el continente de Oceanía presenta solo dieciséis observaciones y este es el coeficiente significativo de la nueva regresión que incluya a la dicotómica por continente.

Asimismo, este modelo presentaba cuatro nuevas regresoras del continente asiático, americano, europeo y de Oceanía. No obstante, la única variable significativa era la de este último continente. Por otro lado, suponemos que no logramos corregir el problema de la multicolinealidad debido a la tendencia común de nuestros datos. Nuestras variables presentaban alta colinealidad entre ellas, como fue en el caso de la tasa de fertilidad, esperanza de vida al nacer y la tasa de fertilidad adolescente. Por ello, decidimos eliminar a estas dos últimas variables.

Además, suponemos que el eliminar variables, como la tasa de fertilidad adolescente y la esperanza de vida al nacer, y la observación del país de Cuba no afectó a la especificación del modelo. La prueba de la especificación de la regresión no tuvo ningún problema.

Nos parece importante mencionar que el proceso para llegar a este modelo incluyó no solo eliminación, sino la elaboración de distintas pruebas que nos permitieron tomar estas decisiones. Para la eliminación de Cuba observamos el histograma y el diagrama de dispersión, los cuales indican la presencia de datos atípicos. Para eliminar variables consideramos que tuvieran correlaciones altas y poca significancia. Elaboramos múltiples modelos ascendentes para comparar la significancia y la aportación de cada una de las regresoras en el modelo y en el R^2 . Claramente nuestro mayor problema está enfocado en la multicolinealidad. A pesar de agrupar a los 201 países en sus continentes, no todos los coeficientes mostraron ser significativos.

Nuestro último modelo, el cual incluye a la variable dicotómica de las diferentes regiones, demuestra que la migración neta y la tasa de mortalidad de niños menores a cinco años no son significativas para el crecimiento económico. Es decir, las variaciones del crecimiento del PIB no son explicadas por la diferencia entre el número de inmigrantes y emigrantes ni por la tasa de mortalidad de niños menores a cinco años. Por otro lado, las variaciones del término dependiente sí son explicadas por la tasa de fertilidad, la tasa de suscripciones a servicios de telefonía y el crecimiento de la población urbana.

No obstante, nuestro modelo muestra que el hecho de que un país pertenezca al continente americano, asiático y europeo no tiene ninguna significancia. Las variaciones del crecimiento económico en los países de África y Oceanía si son explicadas por el hecho de que estos países pertenecen a estas dos regiones.

Ahora bien, otra de las razones por las que podríamos estar obteniendo estos datos es por la pandemia del COVID-19. La causa muestral de la multicolinealidad nos indica que algunas variables pueden presentar alta correlación en ciertos intervalos o periodos de tiempo. Recordemos que diversas economías se estaban recuperando del confinamiento o incluso algunas aún seguían tomando ciertas medidas sanitarias que afectaron a algunas industrias.

La realidad es que los resultados de nuestro modelo nos sorprendieron. No esperábamos que la migración neta no fuera influyente en el crecimiento económico. Sin embargo, el contexto que envuelve al año 2021 puede influir en estas conclusiones a las que llegó la regresión. Pero el hecho de que una tasa de mortalidad no haya afectado al crecimiento del PIB en un periodo en el que exactamente apareció una enfermedad que mató a millones de personas captó nuestra atención. Esto podría ser debido a que la tasa de mortalidad es de infantes y no de personas de la tercera edad, el cual fue un grupo altamente afectado por la pandemia.

En miras al futuro, creemos que una continuación de este proyecto vale la pena en el contexto de un trabajo de investigación más robusto, como una tesis de licenciatura, en el que en lugar de ocupar la tasa de crecimiento como dependiente, sea sustituida por el residuo de Solow. Al hacer esto, esperaríamos un mejor desempeño del R^2 ajustado; no obstante, eso queda fuera del alcance de este análisis de variables alternativas de crecimiento.