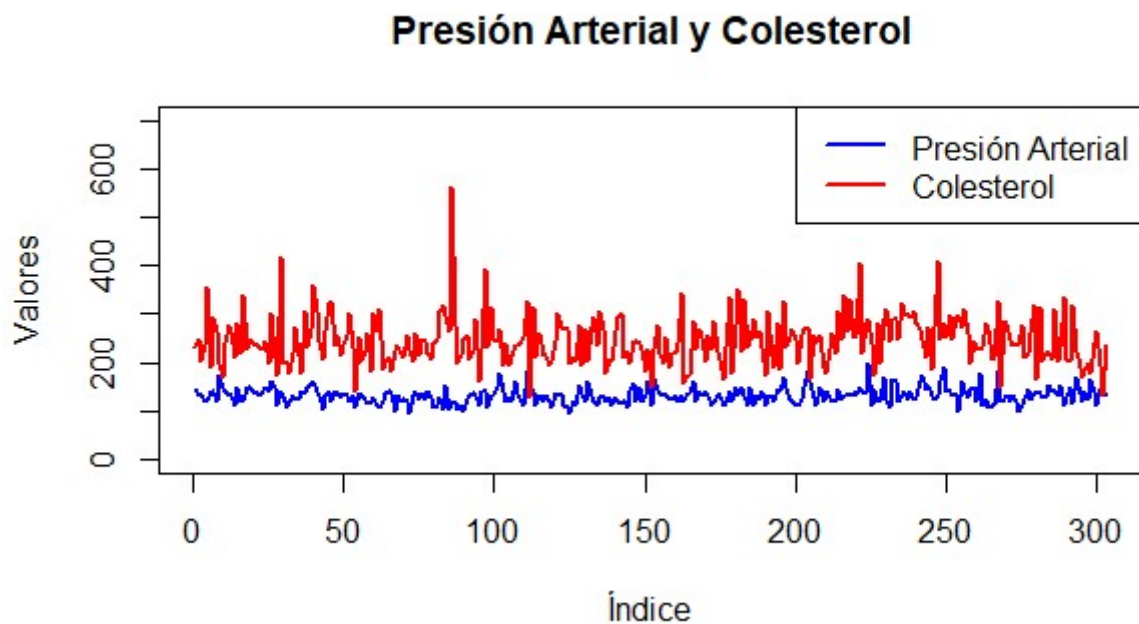


A una base de datos de series de tiempo a elección, aplicar test de cointegración de acuerdo con los pasos seguidos en clase. Interpretar los resultados.

Tomando de la base de datos de kaggle, “Heart Attack Analysis & Prediction Dataset” las variables (trtbps) presión arterial en reposo (en mm Hg) y (chol) colesterol en mg/dl obtenido mediante el sensor de IMC

Primero graficamos las dos series



Analizando estacionalidad vemos que son estacionarias por que el p valor es menor al 5%

Augmented Dickey-Fuller Test

```
data: chol
Dickey-Fuller = -5.5914, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

Phillips-Perron Unit Root Test

```
data: chol
Dickey-Fuller = -17.637, Truncation lag parameter = 5, p-value = 0.01
```

KPSS Test for Level Stationarity

```
data: chol
KPSS Level = 0.13555, Truncation lag parameter = 5, p-value = 0.1
```

Augmented Dickey-Fuller Test

```
data: trtbps
Dickey-Fuller = -6.2659, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

Phillips-Perron Unit Root Test

```
data: trtbps
Dickey-Fuller = -15.772, Truncation lag parameter = 5, p-value = 0.01
```

KPSS Test for Level Stationarity

```
data: trtbps
KPSS Level = 0.51141, Truncation lag parameter = 5, p-value = 0.0391
```

Analizando cointegración vemos que al menos uno de los valores propios es significativamente diferente de cero lo que indica la presencia de cointegración en los datos.

```
#####
# Johansen-Procedure #
#####
```

Test type: trace statistic , with linear trend

Eigenvalues (lambda):
[1] 0.3121749 0.2791126

Values of teststatistic and critical values of test:

	test	10pct	5pct	1pct
r <= 1		98.51	6.50	8.18 11.65
r = 0		211.15	15.66	17.95 23.52

Eigenvectors, normalised to first column:
(These are the cointegration relations)

	chol.l2	trtbps.l2
chol.l2	1.00000	1.000000
trtbps.l2	-1.16945	7.244876

Weights W:
(This is the loading matrix)

	chol.l2	trtbps.l2
chol.d	-0.80525747	-0.13585402
trtbps.d	0.09616881	-0.09933801

Trabajo Final n°2

A una base de datos a elección, aplicar regresión lineal Normal ó Modelo Lineal generalizado :

- Comprobar supuestos de ser el caso
- Ajustar el modelo
- Realizar las pruebas de bondad de ajuste del modelo.

Para la data heart donde:

Edad: Edad de la persona

Sexo: Género de la persona.

Cp: Tipo de dolor en el pecho Tipo de dolor en el pecho

Trtbps: presión arterial en reposo (en mm Hg)

Chol: colesterol en mg/dl obtenido mediante el sensor de IMC

Fbs: (azúcar en sangre en ayunas > 120 mg/dl) (1 = verdadero; 0 = falso)

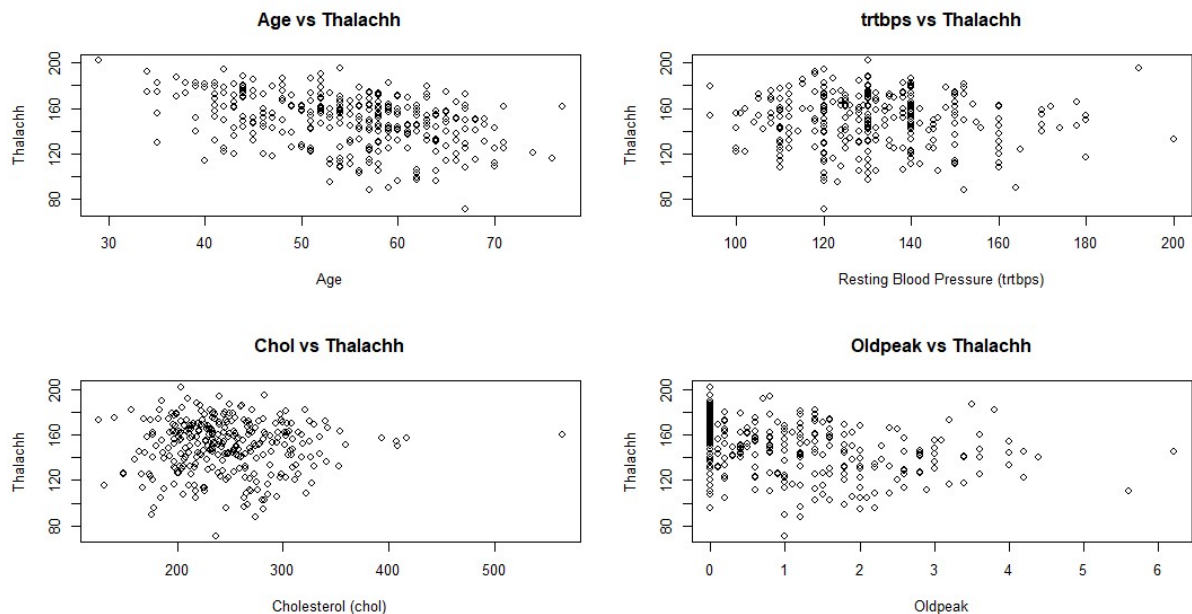
Retecg: resultados electrocardiográficos en reposo

Thalachh: frecuencia cardíaca máxima alcanzada

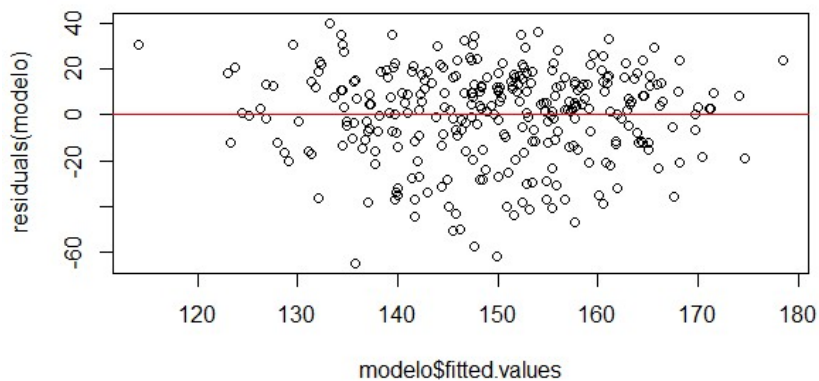
Exng: angina inducida por el ejercicio (1 = sí; 0 = no)

Oldpeak: pico anterior

Para hacer la regresión lineal usamos las variables continuas solamente por lo que usaremos las variables edad, presión, colesterol y puko anterior para calcular la frecuencia cardíaca.



Al graficar residuos no vemos patrones



Autocorrelación

Verificamos con el test de Durbin-Watson para ver si hay autocorrelación en los residuos

Durbin-Watson test

```
data: modelo
DW = 1.742, p-value = 0.01165
alternative hypothesis: true autocorrelation is greater than 0
```

El DW está cerca de 2, lo que sugiere que no hay una fuerte autocorrelación en los residuos. Como el p-valor es menor que 0.05, tenemos suficiente evidencia para aceptar la hipótesis nula de que hay autocorrelación en los residuos.

Normalidad

Shapiro-wilk normality test

```
data: residuos
W = 0.96241, p-value = 4.624e-07
```

Un valor p menor que 0.05 indica que hay evidencia suficiente para aceptar la hipótesis nula de que los residuos no siguen una distribución normal.

Homocedasticidad

studentized Breusch-Pagan test

```
data: modelo
BP = 3.4019, df = 4, p-value = 0.4929
```

Como el valor p es mayor que el nivel de significancia, no hay suficiente evidencia para rechazar la hipótesis nula, lo que sugiere que la homocedasticidad no es violada.

No multicolinealidad

```
          age   trtbps      chol  oldpeak
1.155049 1.112385 1.052664 1.068081
```

Como los valores están por bajo 5 son indica la ausencia de multicolinealidad.

Interpretando los Coeficientes

Interpretando el modelo tenemos que la edad la presión arterial y el pico son significativos y este modelo explica el 23.63% de la variación.

Manteniendo constantes todas las demás variables, se espera que la frecuencia cardiaca disminuya en 0.96792 unidades por cada unidad de aumento en la edad.

Manteniendo constantes todas las demás variables, se espera que la frecuencia cardiaca aumente en 0.13980 unidades por cada unidad de aumento en la presión arterial en reposo.

Manteniendo constantes todas las demás variables, se espera que la variable de frecuencia cardiaca aumente en 0.03289 unidades por cada unidad de aumento en el nivel de colesterol.

Manteniendo constantes todas las demás variables, se espera que la variable frecuencia cardiaca disminuya en 5.68717 unidades por cada unidad de aumento en el pico anterior.

```
Call:
lm(formula = thalachh ~ age + trtbps + chol + oldpeak)

Residuals:
    Min       1Q   Median       3Q      Max
-64.714 -11.864   3.267  13.882  39.837

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  181.67887    10.63340   17.086 < 2e-16 ***
age          -0.96792     0.13630   -7.101 9.11e-12 ***
trtbps         0.13980     0.06927    2.018  0.0445 *
chol           0.03289     0.02280    1.443  0.1502
oldpeak       -5.68717     1.02525   -5.547 6.41e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 20.02 on 298 degrees of freedom
Multiple R-squared:  0.2464,    Adjusted R-squared:  0.2363
F-statistic: 24.36 on 4 and 298 DF,  p-value: < 2.2e-16
```