

# Análisis Exploratorio de datos en Europa: Regresión Logística

*Alicia Perdices Guerra*

*11 de junio, 2021*

## Contents

- **1. ANÁLISIS EXPLORATORIO. MODELOS DE REGRESIÓN.**
  - **1.1 ANÁLISIS EXPLORATORIO. FILTRADO DE LA INFORMACIÓN PARA ITALIA, ALEMANIA, REINO UNIDO, FRANCIA Y ESPAÑA**
    - \* 1.1.1 Se filtra información (De todos los archivos disponibles). Se genera variable con la información necesaria en relación con Gasto, Recursos, Actividades Sanitarias, Cuidados Domiciliarios, Estado de Salud, Prevención y Mortalidad.
    - \* 1.1.2 Modelo de regresión lineal que explique la mortalidad (Mortalidad Prevenible, Mortalidad Tratable) en función del número de Médicos de Atención Primaria, y de Recursos como disposición de Unidades de Resonancia Magnética y Scanners para diagnóstico, Detección de Cáncer de Mama, Cervix Uterino y Vacunación.
    - \* 1.1.3 Modelo de regresión logística que estime la probabilidad de una variable cualitativa (Se transforma Mortalidad\_Tratable y Prevenible en variable cualitativa, con valores de 0 si el país no supera la media muestral para esa variable y 1 en caso contrario), en función de variables cuantitativas (Médicos de Atención Primaria, de Recursos como disposición de Unidades de Resonancia Magnética y Scanners para diagnóstico, y Detección de Cáncer de Mama, Cervix Uterino y Vacunación)
    - \* 1.1.4. Predicciones.

### 1.1 ANÁLISIS EXPLORATORIO. FILTRADO DE LA INFORMACIÓN POR PAÍSES.

Se procede en primer lugar a cargar todos los archivos para poder realizar el análisis.

```
camas_disponibles<-read.csv("C:/temp/Camas_Disponibles_Analisis.csv",sep= ",")
cuidados<-read.csv("C:/temp/Cuidados_Domiciliarios_Analisis.csv",sep= ",")
estado<-read.csv("C:/temp/Estado_Salud_Analisis.csv",sep= ",")
financiacion<-read.csv("C:/temp/Financiacion.csv",sep= ",")
gasto<-read.csv("C:/temp/GastoSanitario_Analisis.csv",sep= ",")
medicos<-read.csv("C:/temp/Medicos_Analisis.csv",sep= ",")
mortalidad<-read.csv("C:/temp/Mortalidad_Analisis.csv",sep= ",")
ocupacion_cama<-read.csv("C:/temp/Ocupacion_Cama_Analisis.csv",sep= ",")
personalSanitarioNoMedicos<-read.csv("C:/temp/PersonalSanitarioNoMedicos_Analisis.csv",sep= ",")
prevencion<-read.csv("C:/temp/Prevencion_Analisis.csv",sep= ",")
ratio_Pacientes_Dialisis<-read.csv("C:/temp/Ratio_Pacientes_Dialisis_Analisis.csv",sep= ",")
ratio_Pacientes_Trasplantados<-read.csv("C:/temp/Ratio_Pacientes_Trasplantados_Analisis.csv",sep= ",")
recursos_tecnologicos<-read.csv("C:/temp/Recursos_Tecnologicos_Analisis.csv",sep= ",")
tecnologia_medica<-read.csv("C:/temp/Tecnologia_Medica_Analisis.csv",sep= ",")
```

#### 1.1.1 Filtrado de la información para los Países foco de estudio

Estudio de regresión de la mortalidad tratable y prevenible en Europa en relación con el número de Médicos de Atención Primaria, y de Recursos como disposición de Unidades de Resonancia Magnética y Scanners para diagnóstico.

*#Se estudia la información para los distintos países en el rango de años 2011-2017,  
 #ya que no existe información de todas las variables estudiadas para todos los años(2009-2019).  
 #solo para los años 2011-2017*

```
financiacion<-filter(financiacion,
  (TIME==2011)|
  (TIME==2012)|
  (TIME==2013)|
  (TIME==2014)|
  (TIME==2015)|
  (TIME==2016)|
  (TIME==2017))
financiacion<-filter(financiacion,
  (GEO=="Spain")|
  GEO=="Italy"|
  GEO=="United Kingdom"|
  GEO=="France"|
  GEO=="Germany (until 1990 former territory of the FRG)")
nrow(financiacion)
```

```
## [1] 35
```

```
head(financiacion)
```

```
##   TIME                                GEO          ICHA11_HF
## 1 2011 Germany (until 1990 former territory of the FRG) All financing schemes
## 2 2011                                Spain All financing schemes
## 3 2011                                France All financing schemes
## 4 2011                                Italy  All financing schemes
## 5 2011                                United Kingdom All financing schemes
## 6 2012 Germany (until 1990 former territory of the FRG) All financing schemes
##      Value Value_imp
## 1 290266.00    FALSE
## 2  97532.09    FALSE
## 3 230575.03    FALSE
## 4 142676.00     TRUE
## 5 229998.79     TRUE
## 6 297784.00    FALSE
```

```
camas_disponibles<-filter(camas_disponibles,
  (TIME==2011)|
  (TIME==2012)|
  (TIME==2013)|
  (TIME==2014)|
  (TIME==2015)|
  (TIME==2016)|
  (TIME==2017))
camas_disponibles<-filter(camas_disponibles,
  (GEO=="Spain")|
  GEO=="Italy"|
  GEO=="United Kingdom"|
  GEO=="France"|
  GEO=="Germany (until 1990 former territory of the FRG)")
nrow(camas_disponibles)
```

```
## [1] 35
```

```
head(camas_disponibles)
```

```
##      TIME                                GEO      UNIT
## 1 2011 Germany (until 1990 former territory of the FRG) Number
## 2 2011                                Spain Number
## 3 2011                                France Number
## 4 2011                                Italy Number
## 5 2011                                United Kingdom Number
## 6 2012 Germany (until 1990 former territory of the FRG) Number
##              FACILITY  Value Value_imp
## 1 Available beds in hospitals (HP.1) 672573      FALSE
## 2 Available beds in hospitals (HP.1) 142394      FALSE
## 3 Available beds in hospitals (HP.1) 414204      FALSE
## 4 Available beds in hospitals (HP.1) 208854      FALSE
## 5 Available beds in hospitals (HP.1) 181972      FALSE
## 6 Available beds in hospitals (HP.1) 670443      FALSE
```

```
estado<-filter(estado,
               (TIME==2011)|
               (TIME==2012)|
               (TIME==2013)|
               (TIME==2014)|
               (TIME==2015)|
               (TIME==2016)|
               (TIME==2017))
estado<-filter(estado,
               Pais=="Spain"|
               Pais=="Italy"|
               Pais=="United Kingdom"|
               Pais=="France"|
               Pais=="Germany (until 1990 former territory of the FRG)")

nrow(estado)
```

```
## [1] 35
```

```
head(estado)
```

```
##      TIME                                Pais Estado_males
## 1 2011 Germany (until 1990 former territory of the FRG)      57.8
## 2 2011                                Spain      65.4
## 3 2011                                France      62.7
## 4 2011                                Italy      63.6
## 5 2011                                United Kingdom      65.2
## 6 2012 Germany (until 1990 former territory of the FRG)      57.3
##      Estado_females
## 1          58.6
## 2          65.6
## 3          63.6
## 4          62.5
## 5          65.2
## 6          57.9
```

```
gasto<-filter(gasto,
              (TIME==2011)|
```

```

      (TIME==2012) |
      (TIME==2013) |
      (TIME==2014) |
      (TIME==2015) |
      (TIME==2016) |
      (TIME==2017))
gasto<-filter(gasto,
  Pais=="Spain" |
  Pais=="Italy" |
  Pais=="United Kingdom" |
  Pais=="France" |
  Pais=="Germany (until 1990 former territory of the FRG)")

nrow(gasto)

```

```
## [1] 35
```

```
head(gasto)
```

```

##   TIME                                     Pais
## 1 2011 Germany (until 1990 former territory of the FRG)
## 2 2011                                     Spain
## 3 2011                                     France
## 4 2011                                     Italy
## 5 2011                               United Kingdom
## 6 2012 Germany (until 1990 former territory of the FRG)
##   Gasto_Cuidado_Rehabilitacion Gasto_Cuidado_Rehabilitacion_Ingresados
## 1                150588.00                80315.00
## 2                55849.51                23772.32
## 3                121151.32                64954.38
## 4                79760.00                41034.00
## 5                126551.13                53271.07
## 6                154062.00                82352.00
##   Gasto_Hospitales Gasto_Hospitales_Mentales Gasto_Hospitales_Especiales
## 1          84816.00             1613.05             8364.00
## 2          41164.82              774.08             2172.17
## 3          87269.89            13272.43              331.86
## 4          65633.00             981.36              66.39
## 5          93210.11             981.36              66.39
## 6          86930.00            1039.49             8686.00

```

```

medicos<-filter(medicos,
  (TIME==2011) |
  (TIME==2012) |
  (TIME==2013) |
  (TIME==2014) |
  (TIME==2015) |
  (TIME==2016) |
  (TIME==2017))
medicos<-filter(medicos,
  Pais=="Spain" |
  Pais=="Italy" |
  Pais=="United Kingdom" |
  Pais=="France" |
  Pais=="Germany (until 1990 former territory of the FRG)")

```

```
nrow(medicos)
```

```
## [1] 35
```

```
head(medicos)
```

```
##    TIME                                Pais
## 1 2011 Germany (until 1990 former territory of the FRG)
## 2 2011                                Spain
## 3 2011                                France
## 4 2011                                Italy
## 5 2011                                United Kingdom
## 6 2012 Germany (until 1990 former territory of the FRG)
##   Generalist_medical_practitioners Specialist_medical_practitioners
## 1                                75252                                235971
## 2                                34722                                108759
## 3                                93140                                106780
## 4                                53856                                177625
## 5                                49838                                120479
## 6                                75390                                242000
##   General_paediatricians AGE_Total AGE_Less_35 AGE_35_44 AGE_45_54 AGE_55_64
## 1                   12854    311223       57748     36100     91591    114365
## 2                   12037    179267       43700     44867     49333    34933
## 3                    7287    199920       17865     37929     61853    69758
## 4                   17475    379930       50260     54414    104373    123815
## 5                    9170     30327        3458       8143      9135     7464
## 6                   13047    317390       58379     37274     88334    120468
```

```
prevencion<-filter(prevencion,
  (TIME==2011) |
  (TIME==2012) |
  (TIME==2013) |
  (TIME==2014) |
  (TIME==2015) |
  (TIME==2016) |
  (TIME==2017))
prevencion<-filter(prevencion,
  Pais=="Spain" |
  Pais=="Italy" |
  Pais=="United Kingdom" |
  Pais=="France" |
  Pais=="Germany (until 1990 former territory of the FRG)")
```

```
nrow(prevencion)
```

```
## [1] 35
```

```
head(prevencion)
```

```
##    TIME                                Pais Deteccion_cancer_mama
## 1 2011 Germany (until 1990 former territory of the FRG)          71.3
## 2 2011                                Spain          77.1
## 3 2011                                France          75.4
## 4 2011                                Italy          70.1
## 5 2011                                United Kingdom          70.1
## 6 2012 Germany (until 1990 former territory of the FRG)          71.3
```

```
## Deteccion_cancer_cervix_uterino Vacunacion
## 1 76.90 56.1
## 2 67.70 57.7
## 3 73.59 55.2
## 4 76.90 62.4
## 5 67.70 74.0
## 6 73.59 58.6
```

```
recursos_tecnologicos<-filter(recursos_tecnologicos,
  (TIME==2011) |
  (TIME==2012) |
  (TIME==2013) |
  (TIME==2014) |
  (TIME==2015) |
  (TIME==2016) |
  (TIME==2017))
recursos_tecnologicos<-filter(recursos_tecnologicos,
  Pais=="Spain" |
  Pais=="Italy" |
  Pais=="United Kingdom" |
  Pais=="France" |
  Pais=="Germany (until 1990 former territory of the FRG)")

nrow(recursos_tecnologicos)
```

```
## [1] 35
```

```
head(recursos_tecnologicos)
```

```
## TIME Pais
## 1 2011 Germany (until 1990 former territory of the FRG)
## 2 2011 Spain
## 3 2011 France
## 4 2011 Italy
## 5 2011 United Kingdom
## 6 2012 Germany (until 1990 former territory of the FRG)
## Operation_theatres_in_hospital Surgical_day_care_places
## 1 1241 53
## 2 4202 2838
## 3 11337 14110
## 4 6092 8013
## 5 3636 2177
## 6 1258 73
## Psychiatric_day_care_place Oncological_day_care_place
## 1 15457 983
## 2 3366 6957
## 3 28424 6052
## 4 631 3056
## 5 3366 1460
## 6 16773 1047
## Geriatric_day_care_places
## 1 1949
## 2 2297
## 3 628
## 4 206
```

```
## 5          1949
## 6          2013

tecnologia_medica<-filter(tecnologia_medica,
  (TIME==2011)|
  (TIME==2012)|
  (TIME==2013)|
  (TIME==2014)|
  (TIME==2015)|
  (TIME==2016)|
  (TIME==2017))
tecnologia_medica<-filter(tecnologia_medica,
  Pais=="Spain"|
  Pais=="Italy"|
  Pais=="United Kingdom"|
  Pais=="France"|
  Pais=="Germany (until 1990 former territory of the FRG)")
nrow(tecnologia_medica)
```

```
## [1] 35
```

```
head(tecnologia_medica)
```

```
##   TIME                                     Pais
## 1 2011 Germany (until 1990 former territory of the FRG)
## 2 2011                                     Spain
## 3 2011                                     France
## 4 2011                                     Italy
## 5 2011                               United Kingdom
## 6 2012 Germany (until 1990 former territory of the FRG)
##   Computed_Tomography_Scanners Gamma_cameras Lithotriptors
## 1                2688.00             97             84
## 2                778.00             286             81
## 3                816.00             345             34
## 4               1937.00             680             34
## 5                536.29              97             34
## 6               2735.00             119             63
##   Magnetic_Resonance_Imaging_Units Angiography_units
## 1                2317.00              76
## 2                 643.00             249
## 3                 489.00              70
## 4               1435.00             759
## 5                 440.52              70
## 6               2305.00             81
```

```
mortalidad<-filter(mortalidad,
  Pais=="Spain"|
  Pais=="Italy"|
  Pais=="United Kingdom"|
  Pais=="France"|
  Pais=="Germany (until 1990 former territory of the FRG)")
nrow(mortalidad)
```

```
## [1] 35
```

Una vez analizado la información sobre las causas de Mortalidad, Recursos, Prevención, Actividades, Estado

de Salud, Cuidados Domiciliarios y Gastos Sanitarios de forma individual por años y países, se procede a establecer relaciones entre las distintas variables por países (se ha escogido los 5 países con mayor gasto y recursos sanitarios para hacer el estudio de forma conjunta. Se procede a estimar un modelo de regresión logística que explique unas variables en función de otras para después poder hacer predicciones sobre hipótesis de datos nuevos, por ejemplo si existe relación entre la mortalidad prevenible de un país y el número de médicos disponibles y qué mortalidad cabría esperar si aumentamos o disminuimos estos recursos.

- **1.1.2 Modelo de regresión lineal/logística que explique la mortalidad (Mortalidad Prevenible, Mortalidad Tratable) en función del número de Médicos de Atención Primaria ,de Recursos como disposición de Unidades de Resonancia Magnética y Scanners para diagnóstico, y actividades de Diagnóstico de Cancer\_mama , Cáncer de Cervix Uterino, y Vacunación**

Se analiza en primer lugar si tiene sentido estimar un modelo de regresión simple entre variables y qué variables serían las adecuadas.

- La relación entre el predictor y la variable respuesta ha de ser lineal. Se puede comprobar con diagrama de dispersión y graficando los residuos.
- Los residuos deben distribuirse normalmente en torno a 0. Se puede comprobar con la distribución de quantiles (normal Q-Q plot) o test de hipótesis de normalidad.
- La variabilidad de los residuos debe ser constante (homocedasticidad): la variabilidad de los datos entorno a la recta de regresión ha de ser aproximadamente constante. Se comprueba graficando los residuos o mediante el test de Breusch-Pagan.

```
head(mortalidad)
```

```
##      TIME                                Pais Mortalidad_Prevenible
## 1 2011 Germany (until 1990 former territory of the FRG)      127461.0
## 2 2011                                Spain      47858.0
## 3 2011                                France      78950.5
## 4 2011                                Italy      68165.5
## 5 2011                        United Kingdom      82226.0
## 6 2012 Germany (until 1990 former territory of the FRG)     125457.5
##      Mortalidad_Tratable Mortalidad_A00_B99 Mortalidad_G00_R50
## 1              74029.0              7411              15593
## 2              27292.0              2946              11834
## 3              35604.5              5094              11842
## 4              41124.5              5423              10691
## 5              48861.0              2555              17160
## 6              72005.5              8232              15598
##      Mortalidad_pneumonia Other_sepsis
## 1              18020              7411
## 2              7819              2946
## 3              9971              5094
## 4              8211              5423
## 5              27968              2555
## 6              17776              8232
```

```
europa_regresion1<-
```

```
  select(mortalidad,TIME,Pais,Mortalidad_Prevenible,Mortalidad_Tratable)
head(europa_regresion1)
```

```
##      TIME                                Pais Mortalidad_Prevenible
## 1 2011 Germany (until 1990 former territory of the FRG)      127461.0
## 2 2011                                Spain      47858.0
## 3 2011                                France      78950.5
## 4 2011                                Italy      68165.5
```



```
## 5 2011 United Kingdom 82226.0
## 6 2012 Germany (until 1990 former territory of the FRG) 125457.5
## Mortalidad_Tratable
## 1 74029.0
## 2 27292.0
## 3 35604.5
## 4 41124.5
## 5 48861.0
## 6 72005.5
```

```
nrow(europa_regresion1)
```

```
## [1] 35
```

```
europa_regresion2<-
  select(tecnologia_medica,Computed_Tomography_Scanners,Magnetic_Resonance_Imaging_Units)
head(europa_regresion2)
```

```
## Computed_Tomography_Scanners Magnetic_Resonance_Imaging_Units
## 1 2688.00 2317.00
## 2 778.00 643.00
## 3 816.00 489.00
## 4 1937.00 1435.00
## 5 536.29 440.52
## 6 2735.00 2305.00
```

```
nrow(europa_regresion2)
```

```
## [1] 35
```

```
europa_regresion3<-
  select(medicos,Generalist_medical_practitioners)
head(europa_regresion3)
```

```
## Generalist_medical_practitioners
## 1 75252
## 2 34722
## 3 93140
## 4 53856
## 5 49838
## 6 75390
```

```
nrow(europa_regresion3)
```

```
## [1] 35
```

```
europa_regresion4<-
  select(prevenccion,Deteccion_cancer_mama,Deteccion_cancer_cervix_uterino,Vacunacion)
europa_regresion<-data.frame(europa_regresion1,europa_regresion2,europa_regresion3,europa_regresion4)
head(europa_regresion)
```

```
## TIME Pais Mortalidad_Prevenible
## 1 2011 Germany (until 1990 former territory of the FRG) 127461.0
## 2 2011 Spain 47858.0
## 3 2011 France 78950.5
## 4 2011 Italy 68165.5
## 5 2011 United Kingdom 82226.0
## 6 2012 Germany (until 1990 former territory of the FRG) 125457.5
## Mortalidad_Tratable Computed_Tomography_Scanners
```

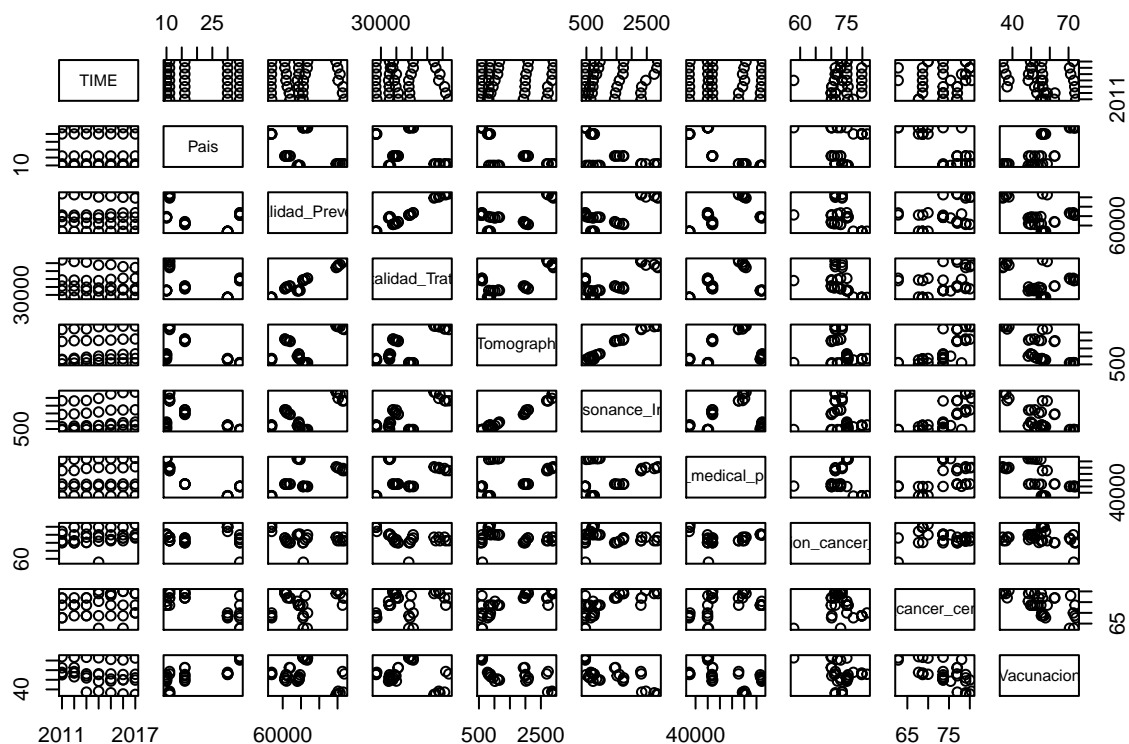
## 1	74029.0	2688.00	
## 2	27292.0	778.00	
## 3	35604.5	816.00	
## 4	41124.5	1937.00	
## 5	48861.0	536.29	
## 6	72005.5	2735.00	
##	Magnetic_Resonance_Imaging_Units	Generalist_medical_practitioners	
## 1	2317.00	75252	
## 2	643.00	34722	
## 3	489.00	93140	
## 4	1435.00	53856	
## 5	440.52	49838	
## 6	2305.00	75390	
##	Deteccion_cancer_mama	Deteccion_cancer_cervix_uterino	Vacunacion
## 1	71.3	76.90	56.1
## 2	77.1	67.70	57.7
## 3	75.4	73.59	55.2
## 4	70.1	76.90	62.4
## 5	70.1	67.70	74.0
## 6	71.3	73.59	58.6

- Se comprueba si el predictor y la variable respuesta tienen una relación lineal mediante un diagrama de dispersión.

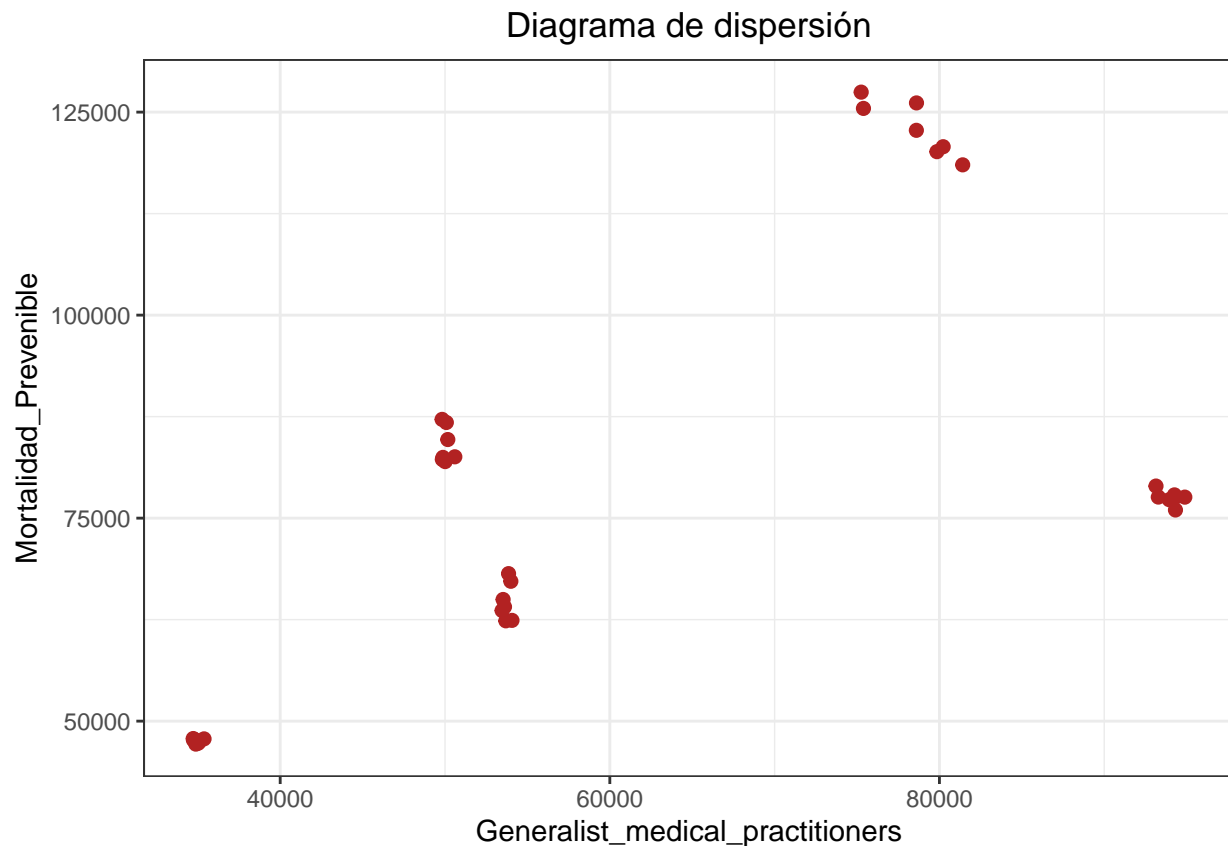
#### Mortalidad\_Prevenible,Generalist\_medical\_practitioners

*#Gráficos de dispersión del conjunto de variables*

`pairs(europa_regresion)`



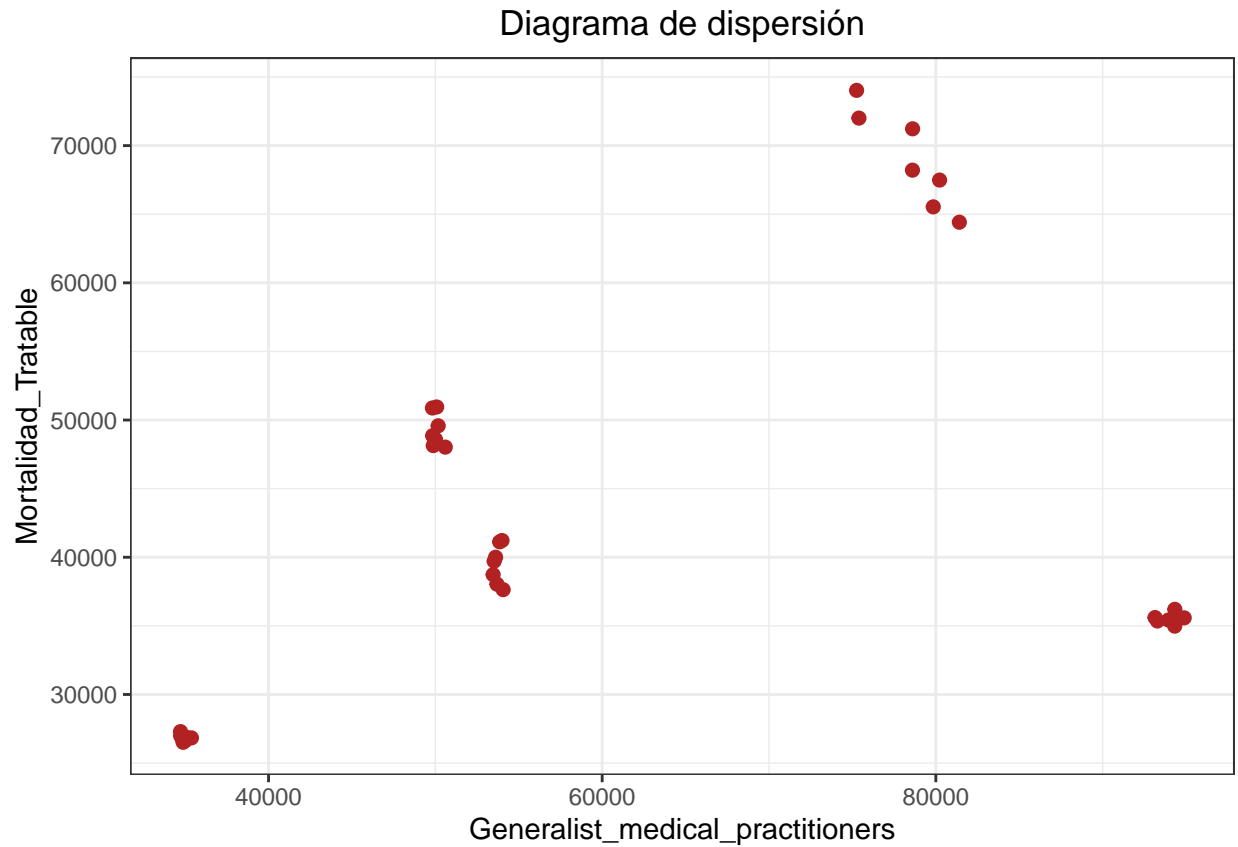
```
ggplot(data = europa_regresion,
       mapping = aes(x = Generalist_medical_practitioners , y = Mortalidad_Prevenible)) +
geom_point(color = "firebrick", size = 2) +
labs(title = "Diagrama de dispersión",
     x = "Generalist_medical_practitioners", y = "Mortalidad_Prevenible") +
theme_bw() +
theme(plot.title = element_text(hjust = 0.5))
```



No se observa una dependencia lineal, por lo que el modelo de regresión lineal no estaría justificado.

**Mortalidad\_Tratable, Generalist\_medical\_practitioners**

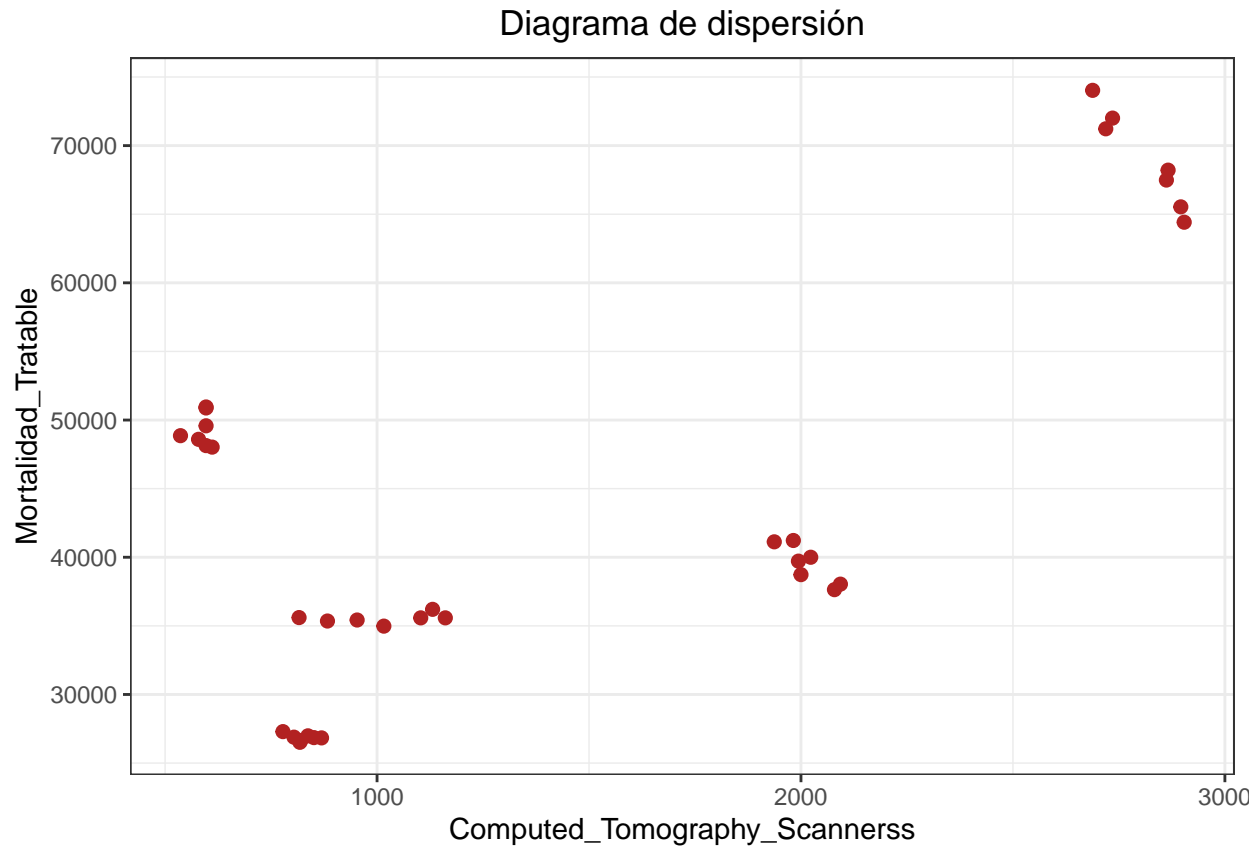
```
#Gráficos de dispersión del conjunto de variables
ggplot(data = europa_regresion,
       mapping = aes(x = Generalist_medical_practitioners , y = Mortalidad_Tratable)) +
geom_point(color = "firebrick", size = 2) +
labs(title = "Diagrama de dispersión",
     x = "Generalist_medical_practitioners", y = "Mortalidad_Tratable") +
theme_bw() +
theme(plot.title = element_text(hjust = 0.5))
```



No se observa una dependencia lineal, por lo que el modelo de regresión lineal no estaría justificado.

Mortalidad\_Tratable,Computed\_Tomography\_Scanners

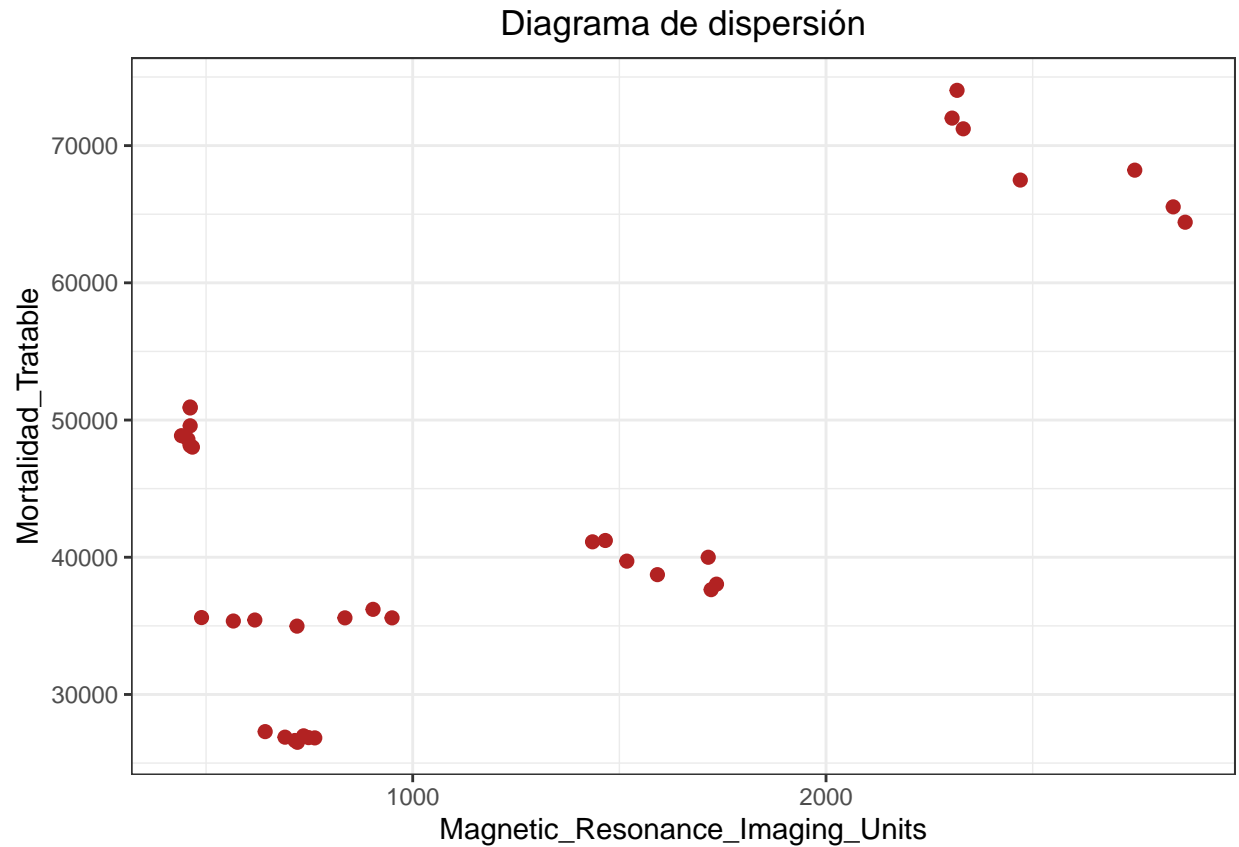
```
ggplot(data = europa_regresion,
       mapping = aes(x =Computed_Tomography_Scanners , y = Mortalidad_Tratable)) +
  geom_point(color = "firebrick", size = 2) +
  labs(title = "Diagrama de dispersión",
       x = "Computed_Tomography_Scannerss", y = "Mortalidad_Tratable") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5))
```



No se observa una dependencia lineal, por lo que el modelo de regresión lineal no estaría justificado.

**Mortalidad\_Tratable, Magnetic\_Resonance\_Imaging\_Units**

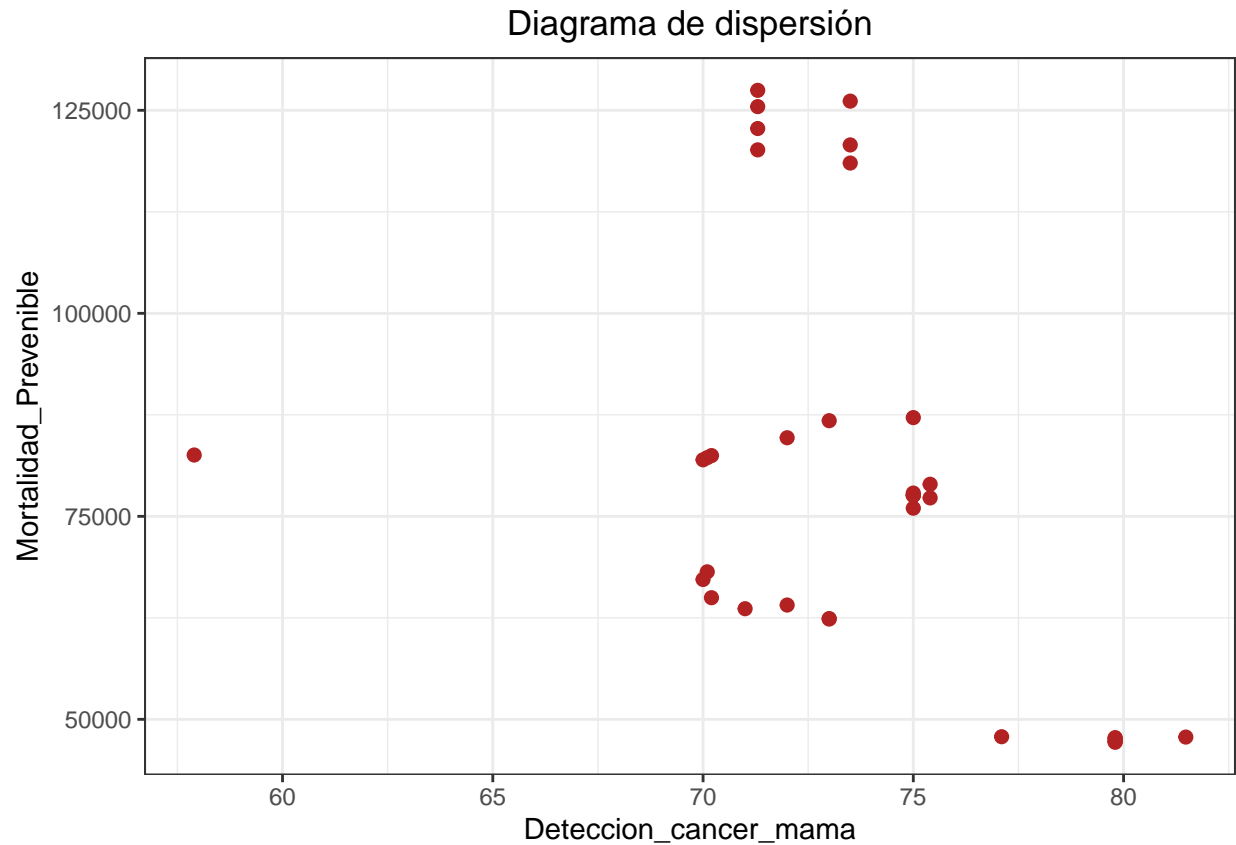
```
ggplot(data = europa_regresion,
       mapping = aes(x =Magnetic_Resonance_Imaging_Units, y = Mortalidad_Tratable)) +
geom_point(color = "firebrick", size = 2) +
labs(title = "Diagrama de dispersión",
     x = "Magnetic_Resonance_Imaging_Units", y = "Mortalidad_Tratable") +
theme_bw() +
theme(plot.title = element_text(hjust = 0.5))
```



No se observa una dependencia lineal, por lo que el modelo de regresión lineal no estaría justificado.

**Mortalidad\_Prevenible,Deteccion\_cancer\_mama**

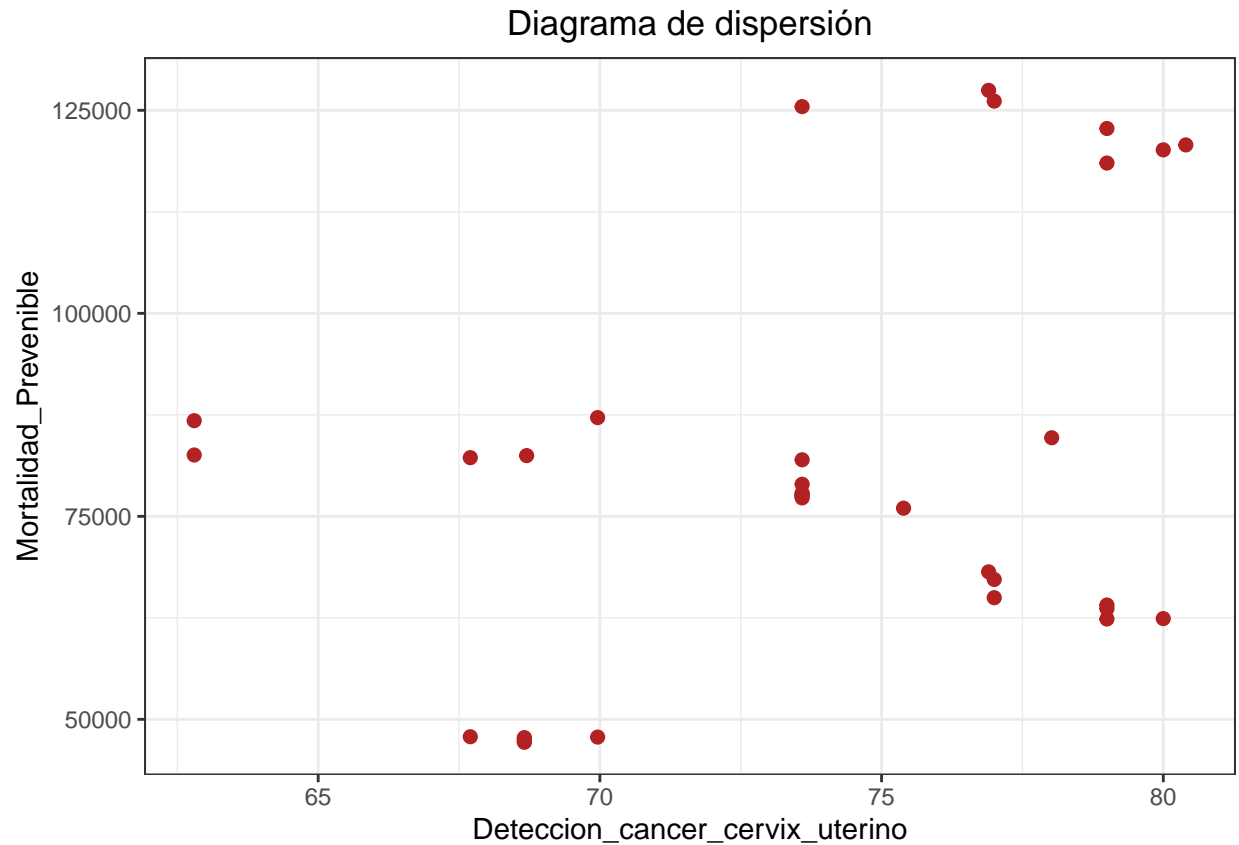
```
ggplot(data = europa_regresion,
       mapping = aes(x =Deteccion_cancer_mama, y = Mortalidad_Prevenible)) +
  geom_point(color = "firebrick", size = 2) +
  labs(title = "Diagrama de dispersión",
       x = "Deteccion_cancer_mama", y = "Mortalidad_Prevenible") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5))
```



No se observa una dependencia lineal, por lo que el modelo de regresión lineal no estaría justificado.

**Mortalidad\_Prevenible,Deteccion\_cancer\_cervix\_uterino**

```
ggplot(data = europa_regresion,
  mapping = aes(x =Deteccion_cancer_cervix_uterino, y = Mortalidad_Prevenible)) +
  geom_point(color = "firebrick", size = 2) +
  labs(title = "Diagrama de dispersión",
    x = "Deteccion_cancer_cervix_uterino", y = "Mortalidad_Prevenible") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5))
```

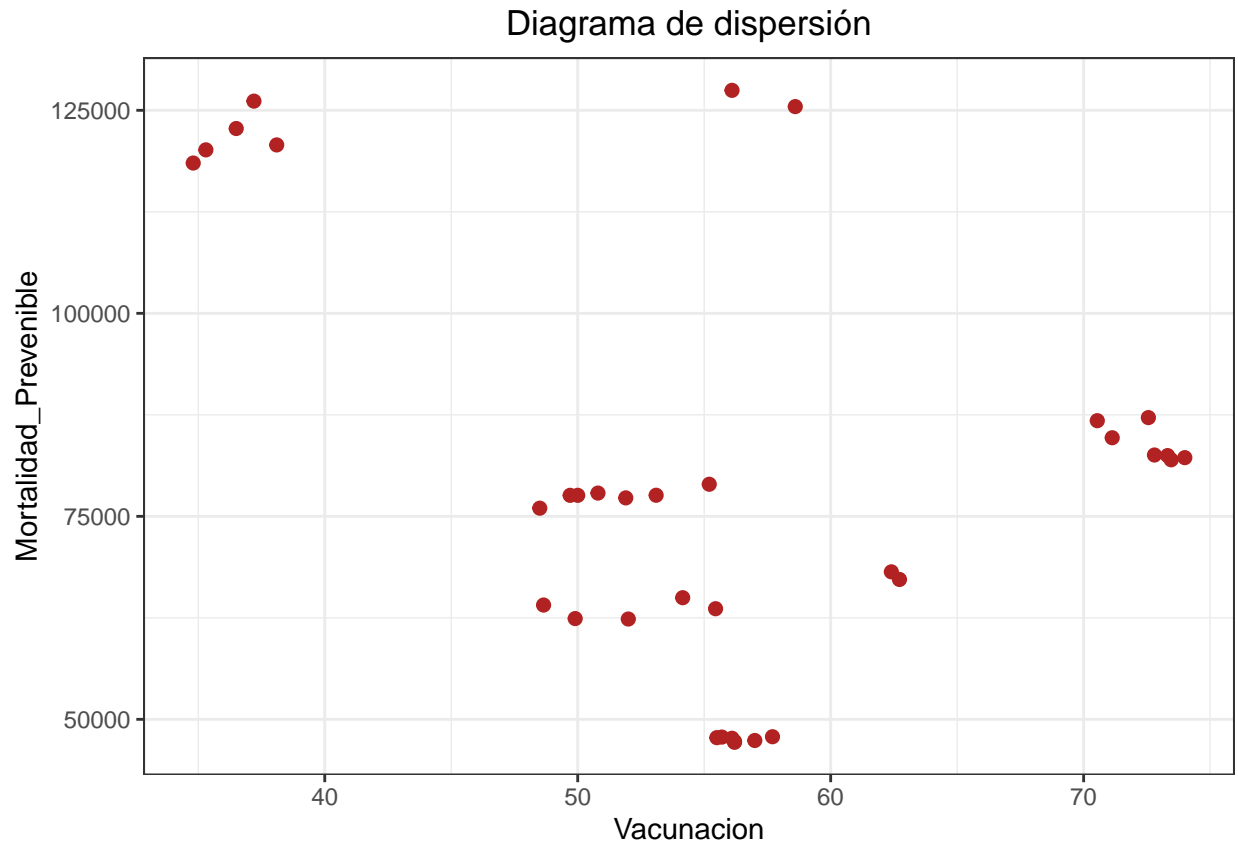


No se observa una dependencia lineal, por lo que el modelo de regresión lineal no estaría justificado.

**Mortalidad\_Prevenible, Vacunacion**

```
ggplot(data = europa_regresion,
       mapping = aes(x = Vacunacion, y = Mortalidad_Prevenible)) +
  geom_point(color = "firebrick", size = 2) +
  labs(title = "Diagrama de dispersión",
       x = "Vacunacion", y = "Mortalidad_Prevenible") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5))
```





No se observa una dependencia lineal, por lo que el modelo de regresión lineal no estaría justificado.

No tiene sentido pensar en un modelo de regresión lineal para las variables e información disponible para el estudio.

#### • 1.1.3 Modelos de regresión logística

Modelo de regresión logística que estime la probabilidad de una variable cualitativa (Se transforma Mortalidad\_Tratable y Prevenible en variable cualitativa, con valores de 0 si el país no supera la media muestral para esa variable y 1 en caso contrario), en función de variables cuantitativas (Médicos de Atención Primaria, y de Recursos como disposición de Unidades de Resonancia Magnética y Scanners para diagnóstico)

- se transforma la variable **Mortalidad\_Tratable** y **Mortalidad\_Prevenible** en cualitativas (con valores 0 y 1).

```
#Media muestral de la variable Mortalidad_Tratable en todos los paises Europeos.(variable mortalidad)
mean_mt<-mean(mortalidad$Mortalidad_Tratable)

#Media muestral de la variable Mortalidad_Prevenible en todos los paises Europeos.(variable mortalidad)
mean_mp<-mean(mortalidad$Mortalidad_Prevenible)

#Se transforma Mortalidad_Tratable en variable cualitativa.
europa_regresion$Mortalidad_Tratable_01<-ifelse(europa_regresion$Mortalidad_Tratable<mean_mt,0,1)

#Se transforma Mortalidad_Prevenible en variable cualitativa.
europa_regresion$Mortalidad_Prevenible_01<-ifelse(europa_regresion$Mortalidad_Prevenible<mean_mp,0,1)
tail(europa_regresion)
```

```
##      TIME                                     Pais Mortalidad_Prevenible
```

```
## 30 2016 United Kingdom 86777.5
## 31 2017 Germany (until 1990 former territory of the FRG) 118506.5
## 32 2017 Spain 47815.5
## 33 2017 France 77585.5
## 34 2017 Italy 62353.5
## 35 2017 United Kingdom 87159.0
## Mortalidad_Tratable Computed_Tomography_Scanners
## 30 50951.5 596.61
## 31 64417.5 2904.00
## 32 26835.5 869.00
## 33 35583.0 1161.00
## 34 38040.5 2093.00
## 35 50880.0 596.61
## Magnetic_Resonance_Imaging_Units Generalist_medical_practitioners
## 30 461.53 50080
## 31 2869.00 81411
## 32 763.00 35378
## 33 950.00 94889
## 34 1735.00 53691
## 35 461.53 49824
## Deteccion_cancer_mama Deteccion_cancer_cervix_uterino Vacunacion
## 30 73.00 62.80 70.54
## 31 73.50 79.00 34.80
## 32 81.48 69.96 55.70
## 33 75.00 73.59 49.70
## 34 73.00 79.00 52.00
## 35 75.00 69.96 72.56
## Mortalidad_Tratable_01 Mortalidad_Prevenible_01
## 30 1 1
## 31 1 1
## 32 0 0
## 33 0 0
## 34 0 0
## 35 1 1
```

```
table(europa_regresion$Mortalidad_Tratable_01)
```

```
##
## 0 1
## 21 14
```

```
table(europa_regresion$Mortalidad_Prevenible_01)
```

```
##
## 0 1
## 21 14
```

Se ajusta un modelo de regresión logística para predecir la probabilidad de tener una Mortalidad Tratable que supere la media en función de las variables Médicos de Atención Primaria, y de Recursos como disposición de Unidades de Resonancia Magnética y Scanners para diagnóstico.

```
modelo.log.1<-
```

```
glm(Mortalidad_Tratable_01~Generalist_medical_practitioners, family = binomial(), data= europa_regresion)
summary(modelo.log.1)
```

```
##
## Call:
```

```
## glm(formula = Mortalidad_Tratable_01 ~ Generalist_medical_practitioners,
##     family = binomial(), data = europa_regresion)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1087  -0.9853  -0.9306   1.3061   1.3962
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -8.737e-01  1.082e+00  -0.808    0.419
## Generalist_medical_practitioners  7.483e-06  1.629e-05   0.459    0.646
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 47.111  on 34  degrees of freedom
## Residual deviance: 46.899  on 33  degrees of freedom
## AIC: 50.899
##
## Number of Fisher Scoring iterations: 4
```

```
modelo.log.2<-
```

```
  glm(Mortalidad_Prevenible_01~Generalist_medical_practitioners, family = binomial(), data= europa_regresion)
summary(modelo.log.2)
```

```
##
## Call:
## glm(formula = Mortalidad_Prevenible_01 ~ Generalist_medical_practitioners,
##     family = binomial(), data = europa_regresion)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1087  -0.9853  -0.9306   1.3061   1.3962
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -8.737e-01  1.082e+00  -0.808    0.419
## Generalist_medical_practitioners  7.483e-06  1.629e-05   0.459    0.646
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 47.111  on 34  degrees of freedom
## Residual deviance: 46.899  on 33  degrees of freedom
## AIC: 50.899
##
## Number of Fisher Scoring iterations: 4
```

```
modelo.log.3<-
```

```
  glm(Mortalidad_Tratable_01~Computed_Tomography_Scanners, family = binomial(), data= europa_regresion)
summary(modelo.log.3)
```

```
##
## Call:
## glm(formula = Mortalidad_Tratable_01 ~ Computed_Tomography_Scanners,
##     family = binomial(), data = europa_regresion)
##
## Deviance Residuals:
```

```

##      Min      1Q   Median      3Q      Max
## -1.1594 -0.9270 -0.8616   1.0393   1.6059
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -1.2857449   0.7275115  -1.767   0.0772 .
## Computed_Tomography_Scanners  0.0005939  0.0004215   1.409   0.1588
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 47.111  on 34  degrees of freedom
## Residual deviance: 45.055  on 33  degrees of freedom
## AIC: 49.055
##
## Number of Fisher Scoring iterations: 4
modelo.log.4<-
  glm(Mortalidad_Tratable_01~Magnetic_Resonance_Imaging_Units, family = binomial(), data= europa_regres.
summary(modelo.log.4)

##
## Call:
## glm(formula = Mortalidad_Tratable_01 ~ Magnetic_Resonance_Imaging_Units,
##      family = binomial(), data = europa_regresion)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1762  -0.8954  -0.8289   0.9826   1.6462
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -1.4152906   0.6855061  -2.065   0.0390 *
## Magnetic_Resonance_Imaging_Units  0.0008141  0.0004655   1.749   0.0803 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 47.111  on 34  degrees of freedom
## Residual deviance: 43.808  on 33  degrees of freedom
## AIC: 47.808
##
## Number of Fisher Scoring iterations: 4
modelo.log.5<-
  glm(Mortalidad_Tratable_01~Magnetic_Resonance_Imaging_Units+Computed_Tomography_Scanners+
      Generalist_medical_practitioners,family = binomial(), data= europa_regresion)
summary(modelo.log.5)

##
## Call:
## glm(formula = Mortalidad_Tratable_01 ~ Magnetic_Resonance_Imaging_Units +
##      Computed_Tomography_Scanners + Generalist_medical_practitioners,
##      family = binomial(), data = europa_regresion)

```

```
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3398  -0.9674  -0.4235   0.9334   1.5655
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.151e+00  1.227e+00  -0.938   0.3483
## Magnetic_Resonance_Imaging_Units  1.655e-02  8.883e-03   1.864   0.0624 .
## Computed_Tomography_Scanners     -1.431e-02  7.886e-03  -1.814   0.0697 .
## Generalist_medical_practitioners  2.537e-05  2.607e-05   0.973   0.3305
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 47.111  on 34  degrees of freedom
## Residual deviance: 35.810  on 31  degrees of freedom
## AIC: 43.81
##
## Number of Fisher Scoring iterations: 6
modelo.log.6<-
  glm(Mortalidad_Prevenible_01~Magnetic_Resonance_Imaging_Units+Computed_Tomography_Scanners+
      Deteccion_cancer_mama+Deteccion_cancer_cervix_uterino,family = binomial(), data= europa_regresion)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(modelo.log.6)

##
## Call:
## glm(formula = Mortalidad_Prevenible_01 ~ Magnetic_Resonance_Imaging_Units +
##      Computed_Tomography_Scanners + Deteccion_cancer_mama + Deteccion_cancer_cervix_uterino,
##      family = binomial(), data = europa_regresion)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.06977  -0.15860  -0.00237   0.00300   1.17812
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      213.63231  125.20785   1.706   0.0880 .
## Magnetic_Resonance_Imaging_Units  0.08747   0.06646   1.316   0.1881
## Computed_Tomography_Scanners     -0.07580   0.05821  -1.302   0.1928
## Deteccion_cancer_mama           -2.16553   1.18371  -1.829   0.0673 .
## Deteccion_cancer_cervix_uterino  -0.66271   0.53439  -1.240   0.2149
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 47.111  on 34  degrees of freedom
## Residual deviance:  8.978  on 30  degrees of freedom
## AIC: 18.978
```

```
##
## Number of Fisher Scoring iterations: 10
modelo.log.7<-
  glm(Mortalidad_Prevenible_01~Computed_Tomography_Scanners+
      Deteccion_cancer_mama+Deteccion_cancer_cervix_uterino+Vacunacion,family = binomial(),
summary(modelo.log.7)

##
## Call:
## glm(formula = Mortalidad_Prevenible_01 ~ Computed_Tomography_Scanners +
##      Deteccion_cancer_mama + Deteccion_cancer_cervix_uterino +
##      Vacunacion, family = binomial(), data = europa_regresion)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6387  -0.6040  -0.3612   0.6083   2.2451
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      57.549891   35.121239    1.639   0.1013
## Computed_Tomography_Scanners    0.001603   0.001329    1.206   0.2278
## Deteccion_cancer_mama      -0.479908   0.287314   -1.670   0.0949 .
## Deteccion_cancer_cervix_uterino -0.344140   0.203229   -1.693   0.0904 .
## Vacunacion           0.006479   0.091288    0.071   0.9434
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 47.111  on 34  degrees of freedom
## Residual deviance: 30.528  on 30  degrees of freedom
## AIC: 40.528
##
## Number of Fisher Scoring iterations: 6
```

Interpretación:

- Se observa que tanto el protocolo de Deteccion de Cáncer de Mama, Cérvix Uterino, Vacunación y la adquisición de Escáners tomográficos disminuyen la probabilidad del aumento de la mortalidad Tratable y Prevenible.
- Sin Embargo, aunque parece contradictorio las Unidades de Resonancia Magnética y el refuerzo de Medicina de atención primaria, podrían incrementar la probabilidad del aumento de la Mortalidad Tratable y Prevenible.

#### 1.1.4 Predicciones

Se realizan las siguientes predicciones:

- ¿Se superaría la Mortalidad Prevenible, si se aumentan las pruebas diagnósticas de detección de Cáncer de Mama y Cérvix, el porcentaje de población vacunada y el número de escáners?

```
#Se comprueba los valores máximos de cada predictor para realizar las modificaciones oportunas
maximo_deteccion_cm<-max(europa_regresion$Deteccion_cancer_mama)
maximo_deteccion_cm
```

```
## [1] 81.48
```

```
maximo_deteccion_cc<-max(europa_regresion$Deteccion_cancer_cervix_uterino)
maximo_deteccion_cc
```

```
## [1] 80.4
```

```
maximo_vacunacion<-max(europa_regresion$Vacunacion)
maximo_vacunacion
```

```
## [1] 74
```

```
maximo_escaners<-max(europa_regresion$Computed_Tomography_Scanners)
maximo_escaners
```

```
## [1] 2904
```

```
new1<-data.frame(Deteccion_cancer_mama=92,
                  Deteccion_cancer_cervix_uterino=89,
                  Vacunacion=98,
                  Computed_Tomography_Scanners=4000)
p1=predict(modelo.log.7,new1,type="response",se.fit=TRUE)
p1
```

```
## $fit
##          1
## 3.772807e-05
##
## $se.fit
##          1
## 0.0004576144
##
## $residual.scale
## [1] 1
```

- ¿Se superaría la Mortalidad Prevenible, si se disminuyen las pruebas diagnósticas de detección de Cáncer de Mama y Cérvix , el porcentaje de población vacunada y el número de escáners?

```
new2<-data.frame(Deteccion_cancer_mama=50,
                  Deteccion_cancer_cervix_uterino=50,
                  Vacunacion=50,
                  Computed_Tomography_Scanners=1000)
p2=predict(modelo.log.7,new2,type="response",se.fit=TRUE)
p2
```

```
## $fit
## 1
## 1
##
## $se.fit
##          1
## 1.214035e-07
##
## $residual.scale
## [1] 1
```

Interpretación:

- Si se aumentan los recursos, la probabilidad de que aumente la mortalidad prevenible es muy baja.
- Por el contrario si se disminuyen, la probabilidad es máxima.

## Intervalos de Confianza de los modelos

Dado que el modelo logístico modela el logaritmo de ODDs, estas son las unidades en las que se devuelven predicciones. Es necesario convertirlas de nuevo a probabilidad mediante la función `logit`. En R la función `predict()` puede devolver directamente las probabilidades en lugar de los logODDs si se indica el argumento `type="response"`. Pero para calcular intervalos de confianza y que éstos no se salgan del rango  $[0,1]$ , es necesario emplear los logODDs y una vez sustraído o sumado el margen de error, se transforman a probabilidades.

```
#si se indica se.fit=TRUE se devuelve el error estándar de cada predicción junto con  
#el valor de la predicción (fit)
```

```
CI_inferior_p1<-p1$fit-1.96*p1$se.fit  
CI_superior_p1<-p1$fit+1.96*p1$se.fit  
CI_inferior_p1
```

```
##           1  
## -0.0008591962
```

```
CI_superior_p1
```

```
##           1  
## 0.0009346523
```

```
CI_inferior_p2<-p2$fit-1.96*p1$se.fit  
CI_superior_p2<-p2$fit+1.96*p1$se.fit  
CI_inferior_p2
```

```
##           1  
## 0.9991031
```

```
CI_superior_p2
```

```
##           1  
## 1.000897
```

```
#Modelo modelo.log.5
```

```
intervalo_mod_5<-confint(object = modelo.log.5,level = 0.95)
```

```
## Waiting for profiling to be done...
```

```
intervalo_mod_5
```

```
##           2.5 %           97.5 %  
## (Intercept)      -3.6724675555  1.225776e+00  
## Magnetic_Resonance_Imaging_Units  0.0040815839  3.884228e-02  
## Computed_Tomography_Scanners      -0.0341073247 -3.121959e-03  
## Generalist_medical_practitioners -0.0000229954  8.432069e-05
```

```
#Modelo modelo.log.7
```

```
intervalo_mod_7<-confint(object = modelo.log.7,level = 0.95)
```

```
## Waiting for profiling to be done...
```

```
intervalo_mod_7
```

```
##           2.5 %           97.5 %  
## (Intercept)      -1.022370035 139.968044017  
## Computed_Tomography_Scanners      -0.000702269  0.004753664  
## Deteccion_cancer_mama      -1.153260644 -0.001592713  
## Deteccion_cancer_cervix_uterino -0.834542982 -0.008049804  
## Vacunacion        -0.175053084  0.199804199
```