

# Análisis Exploratorio de los Datos ACTIVIDADES

Alicia Perdices Guerra

19 de mayo, 2021

## Contents

- **1. ANÁLISIS EXPLORATORIO POR PAISES.**
  - 1.1 EN RELACIÓN CON LAS ACTIVIDADES SANITARIAS
    - \* 1.1.1 Análisis Descriptivo
    - \* 1.1.2 Visualización y Distribución de la variable “Value”
      - OCUPACIÓN DE CAMA HOSPITALARIA.
      - DIALISIS Y TRASPLANTE.
    - \* 1.1.3 Normalidad de la variable “Value”
      - OCUPACIÓN DE CAMA HOSPITALARIA.
      - DIALISIS Y TRASPLANTE.

**1. ANÁLISIS EXPLORATORIO POR PAISES** Se procede en primer lugar a cargar todos los archivos para poder realizar el análisis.

```
ocupacion_cama<-read.csv("C:/temp/OcupacionCamaHospitalaria_clean.csv",sep= ",")
dia_tras<-read.csv("C:/temp/Pacientes_Dialisis_Trasplantes_clean.csv",sep= ",")
```

### 1.1.- EN RELACIÓN CON LAS ACTIVIDADES SANITARIAS

- **1.1.1 Análisis Descriptivo**

Se procede a realizar el análisis descriptivo:

```
summary(ocupacion_cama)
```

```
##      TIME      GEO      ICHA_HC      UNIT
## Min.   :2010   Length:310   Length:310   Length:310
## 1st Qu.:2012   Class :character Class :character Class :character
## Median :2014   Mode  :character Mode  :character Mode  :character
## Mean   :2014
## 3rd Qu.:2017
## Max.   :2019
##      Value      Value_imp
## Min.   :45.60   Mode :logical
## 1st Qu.:70.22   FALSE:229
## Median :74.72   TRUE :81
## Mean   :74.60
## 3rd Qu.:79.80
## Max.   :93.80
```

```
summary(dia_tras)
```

```
##      TIME      GEO      UNIT      ICD9CM
## Min.   :2005   Length:1740   Length:1740   Length:1740
## 1st Qu.:2007   Class :character   Class :character   Class :character
## Median :2010   Mode  :character   Mode  :character   Mode  :character
## Mean   :2010
## 3rd Qu.:2012
## Max.   :2014
##      Value      Value_imp
## Min.    : 0.00   Mode :logical
## 1st Qu.: 52.23   FALSE:981
## Median : 107.24   TRUE :759
## Mean    : 6166.82
## 3rd Qu.: 4705.00
## Max.    :91718.00
```

Se filtra el dataframe para que obtener el ratio de pacientes en Diálisis y Trasplantados por Países (Nº por cada 100000 habitantes). En el archivo de ocupación de cama obtendremos el porcentaje de ocupación de camas por Países.

```
dia_tras_ratio<-filter(dia_tras,UNIT=="Per hundred thousand inhabitants")
dia_tras_ratio_dia<-filter(dia_tras_ratio,ICD9CM=="Haemodialysis")
dia_tras_ratio_tras<-filter(dia_tras_ratio,ICD9CM=="Transplantation of kidney")

#Generamos los ficheros filtrados para utilizarlos en el siguiente análisis.

write.csv(dia_tras_ratio_dia,
          file="Ratio_Pacientes_Dialisis_Analisis.csv", row.names = FALSE)
write.csv(dia_tras_ratio_tras,
          file="Ratio_Pacientes_Trasplantados_Analisis.csv", row.names = FALSE)
write.csv(ocupacion_cama,
          file="Ocupacion_Cama_Analisis.csv", row.names = FALSE)
```

Se reescalan los datos:

```
#Pacientes en Diálisis
=====
dia_tras_ratio_dia["Value_norm"]<-
  rescale(dia_tras_ratio_dia$Value, to=c(0,1))
#Pacientes Trasplantados
=====
dia_tras_ratio_tras["Value_norm"]<-
  rescale(dia_tras_ratio_tras$Value, to=c(0,1))
#Ocupación de Camas
=====
ocupacion_cama["Value_norm"]<-
  rescale(ocupacion_cama$Value, to=c(0,1))
```

- 1.1.2 Visualización y Distribución de la variable “Value”

Se visualiza la variable “Value” en función de TIME, y los distintos tipos de Actividades Sanitarias.

```

#Actividades Sanitarias
#=====

#GRÁFICAS DE BARRAS
#=====
#Gráfica de barras del ratio de pacientes en Diálisis."
plot1=ggplot(data=dia_tras_ratio_dia)+
  geom_col(aes(x=TIME,y=Value))+
  theme(axis.text.x = element_text(angle = 45))+
  scale_y_continuous(limit=c(0,3000))+
  ggtitle("Pacientes Diálisis /100.000hab")+
  theme (plot.title = element_text(size=rel(0.5), hjust = 0.5))

#Gráfica de barras del ratio de pacientes Trasplantados."
plot2=ggplot(data=dia_tras_ratio_tras)+
  geom_col(aes(x=TIME,y=Value))+
  theme(axis.text.x = element_text(angle = 45))+
  scale_y_continuous(limit=c(0,3000))+
  ggtitle("Pacientes Trasplantados /100.000hab")+
  theme (plot.title = element_text(size=rel(0.5), hjust = 0.5))

#Gráfica de barras del porcentaje de Ocupación de Cama."
plot3=ggplot(data=ocupacion_cama)+
  geom_col(aes(x=TIME,y=Value))+
  theme(axis.text.x = element_text(angle = 45))+
  scale_y_continuous(limit=c(0,3000))+
  ggtitle("Ocupación de Cama (%))+
  theme (plot.title = element_text(size=rel(0.5), hjust = 0.5))

#GRÁFICAS DE PUNTOS
#=====
##Gráfica de puntos del ratio de pacientes en Diálisis por Países.""
plot4=ggplot(data=dia_tras_ratio_dia)+
  geom_point(aes(x=GEO,y=Value))+
  theme(axis.text.x = element_text(size= 5,angle = 30,vjust=1,hjust = 1))+
  scale_y_continuous(limit=c(0,200))+
  ggtitle("Pacientes en Diálisis /100.000 hab")+
  theme (plot.title = element_text(size=rel(0.5),hjust=0.5))

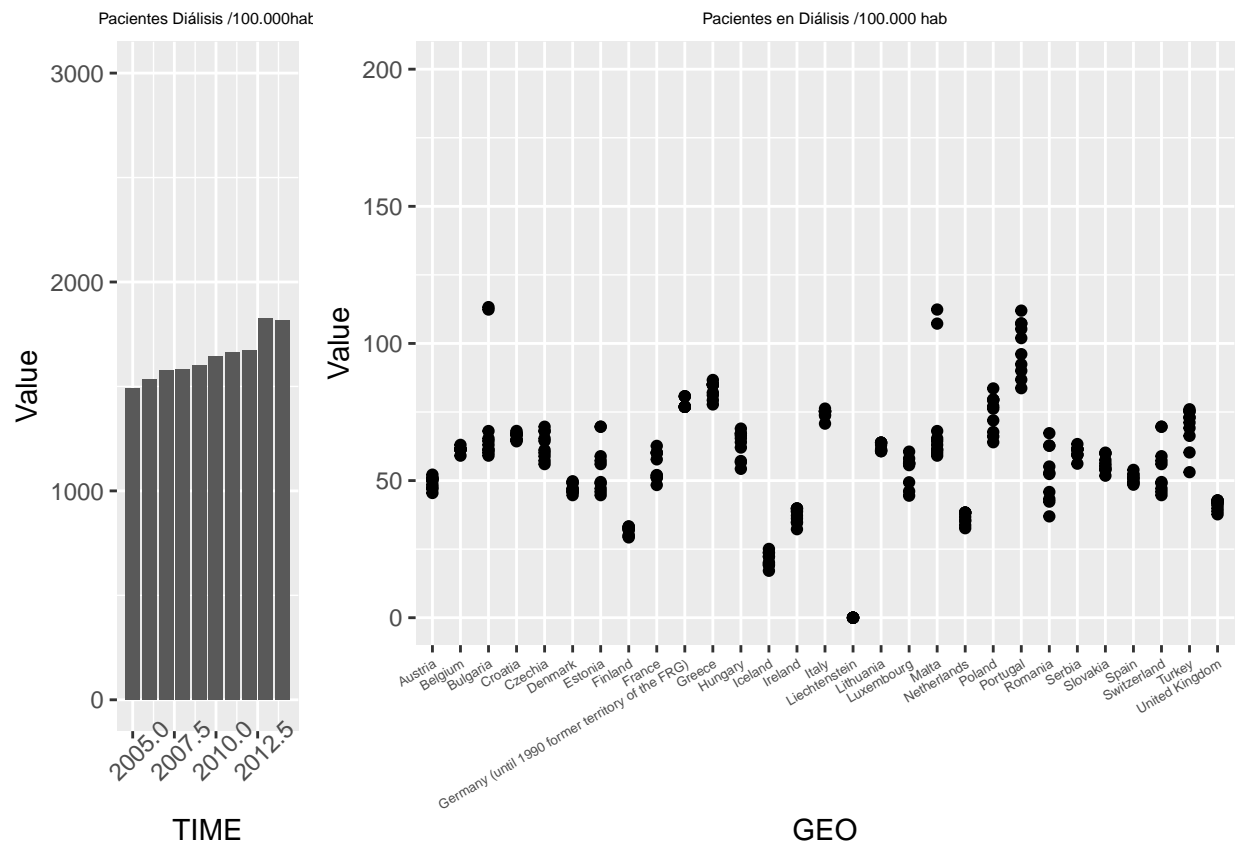
##Gráfica de puntos del ratio de pacientes Trasplantados por Países.""
plot5=ggplot(data=dia_tras_ratio_tras)+
  geom_point(aes(x=GEO,y=Value))+
  theme(axis.text.x = element_text(size= 5,angle = 30,vjust=1,hjust = 1))+
  scale_y_continuous(limit=c(0,200))+
  ggtitle("Pacientes Trasplantados /100.000 hab")+
  theme (plot.title = element_text(size=rel(0.5),hjust=0.5))

##Gráfica de puntos del porcentaje de Ocupación de Cama.""
plot6=ggplot(data=ocupacion_cama)+
  geom_point(aes(x=GEO,y=Value))+
  theme(axis.text.x = element_text(size= 5,angle = 30,vjust=1,hjust = 1))+
  scale_y_continuous(limit=c(0,200))+
  ggtitle("Ocupación de Cama %")+

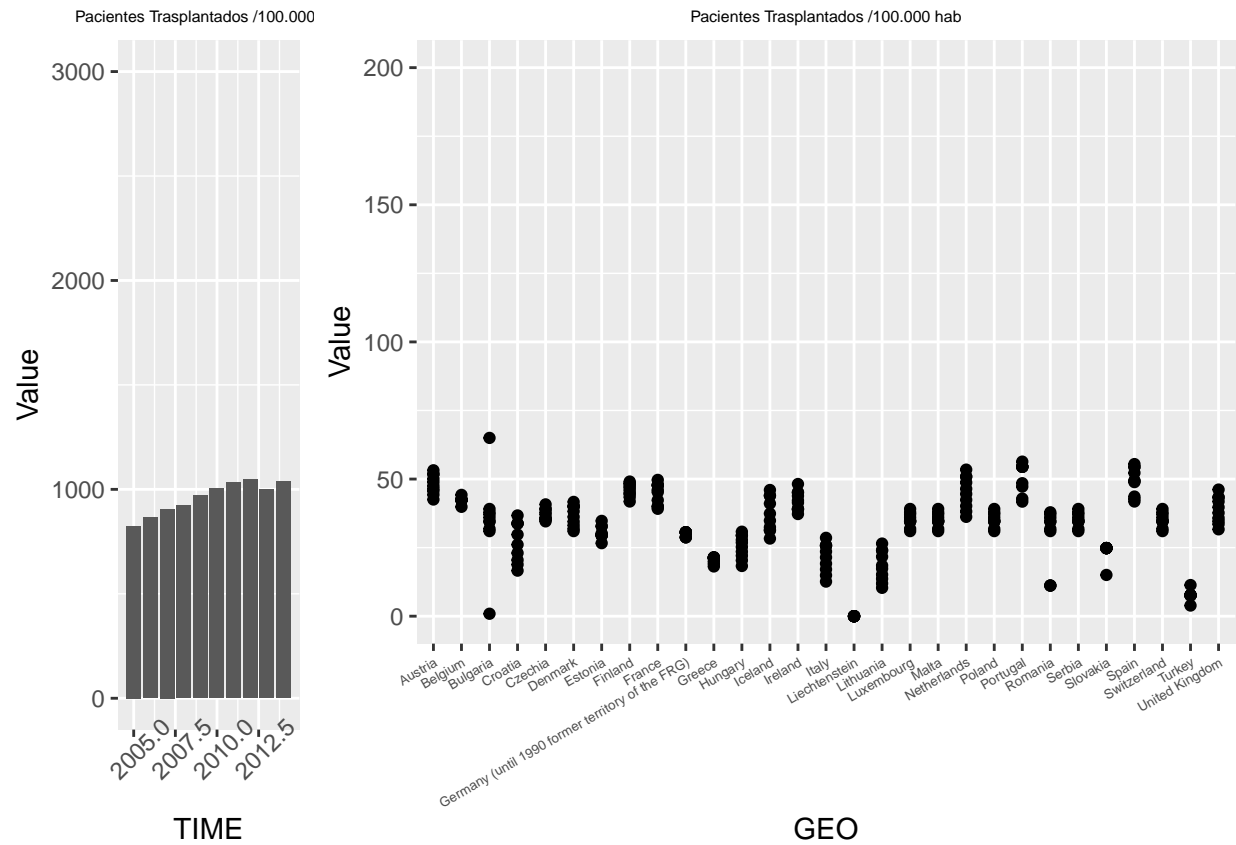
```

```
theme (plot.title = element_text(size=rel(0.5),hjust=0.5))

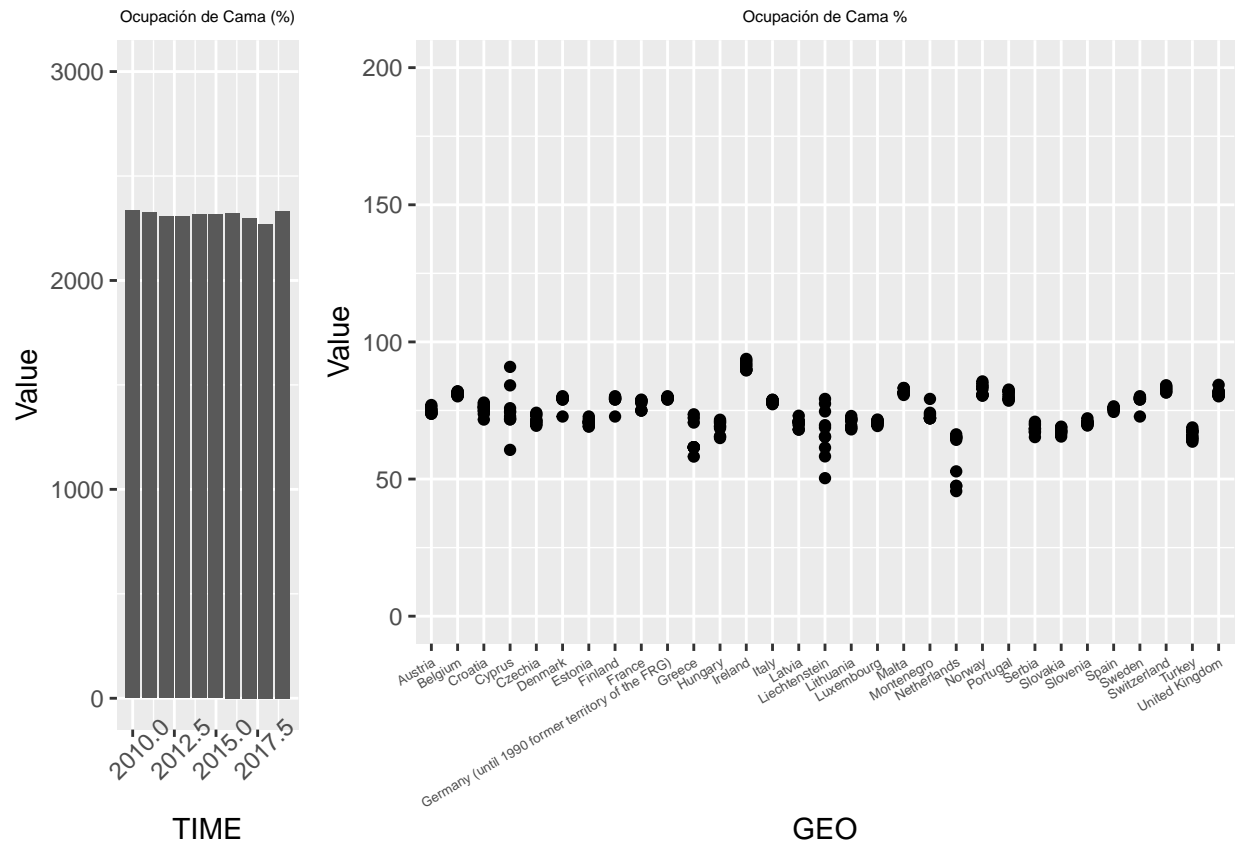
grid.arrange(plot1,plot4,widths=c(1,3), ncol=2)
```



```
grid.arrange(plot2,plot5,widths=c(1,3), ncol=2)
```



```
grid.arrange(plot3,plot6,widths=c(1,3), ncol=2)
```



Se obtienen los 5 países con mayor número de Pacientes en Diálisis, Trasplantados (/100000 hb), y mayor ratio de ocupación de camas por años y en cómputo global.

- Pacientes en Diálisis

```
#####
#Para "Pacientes en Diálisis" #
#####

#Se filtra por Año
y_2005<-filter(dia_tras_ratio_dia, TIME==2005)
y_2006<-filter(dia_tras_ratio_dia, TIME==2006)
y_2007<-filter(dia_tras_ratio_dia, TIME==2007)
y_2008<-filter(dia_tras_ratio_dia, TIME==2008)
y_2009<-filter(dia_tras_ratio_dia, TIME==2009)
y_2010<-filter(dia_tras_ratio_dia, TIME==2010)
y_2011<-filter(dia_tras_ratio_dia, TIME==2011)
y_2012<-filter(dia_tras_ratio_dia, TIME==2012)
y_2013<-filter(dia_tras_ratio_dia, TIME==2013)
y_2014<-filter(dia_tras_ratio_dia, TIME==2014)

#Se ordena por "Value"

actividad_5países_2005<-y_2005[with(y_2005, order(-y_2005$Value)),]
actividad_5países_2006<-y_2006[with(y_2006, order(-y_2006$Value)),]
actividad_5países_2007<-y_2007[with(y_2007, order(-y_2007$Value)),]
```

```

actividad_5países_2008<-y_2008[with(y_2008, order(-y_2008$Value)),]
actividad_5países_2009<-y_2009[with(y_2009, order(-y_2009$Value)),]
actividad_5países_2010<-y_2010[with(y_2010, order(-y_2010$Value)),]
actividad_5países_2011<-y_2011[with(y_2011, order(-y_2011$Value)),]
actividad_5países_2012<-y_2012[with(y_2012, order(-y_2012$Value)),]
actividad_5países_2013<-y_2013[with(y_2013, order(-y_2013$Value)),]
actividad_5países_2014<-y_2014[with(y_2014, order(-y_2014$Value)),]

```

*#Se crea una tabla para cada año sobre las Actividades Sanitarias  
 #(Pacientes en Diálisis) de los 5 Países con un mayor número de ellos( por cada 100000 habitantes).*

```

kable(actividad_5países_2005[0:5,c(2,5)],
      col.names = c("País", "P.Diálisis/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes en Diálisis en 2005")

```

Table 1: Países con mayor Número de Pacientes en Diálisis en 2005

	País	P.Diálisis/100000 habitantes
20	Portugal	83.69
8	Greece	77.76
5	Germany (until 1990 former territory of the FRG)	76.91
12	Italy	70.77
11	Croatia	66.84

```

kable(actividad_5países_2006[0:5,c(2,5)],
      col.names = c("País", "P.Diálisis/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes en Diálisis en 2006")

```

Table 2: Países con mayor Número de Pacientes en Diálisis en 2006

	País	P.Diálisis/100000 habitantes
20	Portugal	86.82
5	Germany (until 1990 former territory of the FRG)	80.74
8	Greece	79.33
12	Italy	73.98
11	Croatia	66.70

```

kable(actividad_5países_2007[0:5,c(2,5)],
      col.names = c("País", "P.Diálisis/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes en Diálisis en 2007")

```

Table 3: Países con mayor Número de Pacientes en Diálisis en 2007

	País	P.Diálisis/100000 habitantes
20	Portugal	90.06
8	Greece	81.15
5	Germany (until 1990 former territory of the FRG)	76.91

	País	P.Diálisis/100000 habitantes
12	Italy	73.70
11	Croatia	68.10

```
kable(actividad_5países_2008[0:5,c(2,5)],
      col.names = c("País", "P.Diálisis/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes en Diálisis en 2008")
```

Table 4: Países con mayor Número de Pacientes en Diálisis en 2008

	País	P.Diálisis/100000 habitantes
20	Portugal	92.39
8	Greece	82.10
19	Poland	77.10
5	Germany (until 1990 former territory of the FRG)	76.91
12	Italy	75.25

```
kable(actividad_5países_2009[0:5,c(2,5)],
      col.names = c("País", "P.Diálisis/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes en Diálisis en 2009")
```

Table 5: Países con mayor Número de Pacientes en Diálisis en 2009

	País	P.Diálisis/100000 habitantes
20	Portugal	96.09
8	Greece	84.99
5	Germany (until 1990 former territory of the FRG)	76.91
12	Italy	76.24
29	Turkey	75.15

```
kable(actividad_5países_2010[0:5,c(2,5)],
      col.names = c("País", "P.Diálisis/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes en Diálisis en 2010")
```

Table 6: Países con mayor Número de Pacientes en Diálisis en 2010

	País	P.Diálisis/100000 habitantes
20	Portugal	101.95
8	Greece	86.69
19	Poland	79.35
5	Germany (until 1990 former territory of the FRG)	76.91
29	Turkey	76.00

```
kable(actividad_5países_2011[0:5,c(2,5)],
      col.names = c("País", "P.Diálisis"),
      caption = "Países con mayor Número de Pacientes en Diálisis en 2011")
```



Table 7: Países con mayor Número de Pacientes en Diálisis en 2011

	País	P.Diálisis
20	Portugal	105.27
8	Greece	84.99
19	Poland	83.59
5	Germany (until 1990 former territory of the FRG)	76.91
29	Turkey	75.27

```
kable(actividad_5países_2012[0:5,c(2,5)],
      col.names = c("País", "P.Diálisis/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes en Diálisis en 2012")
```

Table 8: Países con mayor Número de Pacientes en Diálisis en 2012

	País	P.Diálisis/100000 habitantes
20	Portugal	107.30
8	Greece	84.99
5	Germany (until 1990 former territory of the FRG)	76.91
19	Poland	76.23
12	Italy	75.25

```
kable(actividad_5países_2013[0:5,c(2,5)],
      col.names = c("País", "P.Diálisis/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes en Diálisis en 2013")
```

Table 9: Países con mayor Número de Pacientes en Diálisis en 2013

	País	P.Diálisis/100000 habitantes
2	Bulgaria	112.36
20	Portugal	111.97
16	Malta	107.23
8	Greece	84.99
5	Germany (until 1990 former territory of the FRG)	80.74

```
kable(actividad_5países_2014[0:5,c(2,5)],
      col.names = c("País", "P.Diálisis/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes en Diálisis en 2014")
```

Table 10: Países con mayor Número de Pacientes en Diálisis en 2014

	País	P.Diálisis/100000 habitantes
2	Bulgaria	113.25
16	Malta	112.36
20	Portugal	107.30
8	Greece	84.99

	País	P.Diálisis/100000 habitantes
5	Germany (until 1990 former territory of the FRG)	80.74

A continuación, se agrupa toda la información (P. Diálisis) por países en una tabla:

```
a1<-group_by(dia_tras_ratio_dia,GEO)#Se agrupa por paises
#Se selecciona las variables Pais y Value (P.Diálisis/100000 hab)
a2<-select(a1,GEO:Value)
#Se muestra la información por cada país, con Value=suma de
#los valores de cada país en los 10 años.
a3<-(summarize(a2,suma=sum(Value)/10))
a4<-data.frame(a3)#Se convierte la información en un dataframe.
#Se ordena el DataFrame por la variable Suma de forma descendente.
a5<-a4[with(a4,order(-a4$suma)),]

#Se crea una tabla con toda la informacion
kable(a5[0:5,c(1,2)],
      col.names = c("País","P.Diálisis/100000 habitantes"),
      caption = "Países con la mayor media de Pacientes en Diálisis en 2005-2014")
```

Table 11: Países con la mayor media de Pacientes en Diálisis en 2005-2014

	País	P.Diálisis/100000 habitantes
22	Portugal	98.284
11	Greece	83.198
10	Germany (until 1990 former territory of the FRG)	78.059
15	Italy	74.619
21	Poland	74.506

- Pacientes Trasplantados

```
#####
#Para "Pacientes en Diálisis"          #
#####

#Se filtra por Año
y_2005<-filter(dia_tras_ratio_tras, TIME==2005)
y_2006<-filter(dia_tras_ratio_tras, TIME==2006)
y_2007<-filter(dia_tras_ratio_tras, TIME==2007)
y_2008<-filter(dia_tras_ratio_tras, TIME==2008)
y_2009<-filter(dia_tras_ratio_tras, TIME==2009)
y_2010<-filter(dia_tras_ratio_tras, TIME==2010)
y_2011<-filter(dia_tras_ratio_tras, TIME==2011)
y_2012<-filter(dia_tras_ratio_tras, TIME==2012)
y_2013<-filter(dia_tras_ratio_tras, TIME==2013)
y_2014<-filter(dia_tras_ratio_tras, TIME==2014)

#Se ordena por "Value"
```

```

actividad_5países_2005<-y_2005[with(y_2005, order(-y_2005$Value)),]
actividad_5países_2006<-y_2006[with(y_2006, order(-y_2006$Value)),]
actividad_5países_2007<-y_2007[with(y_2007, order(-y_2007$Value)),]
actividad_5países_2008<-y_2008[with(y_2008, order(-y_2008$Value)),]
actividad_5países_2009<-y_2009[with(y_2009, order(-y_2009$Value)),]
actividad_5países_2010<-y_2010[with(y_2010, order(-y_2010$Value)),]
actividad_5países_2011<-y_2011[with(y_2011, order(-y_2011$Value)),]
actividad_5países_2012<-y_2012[with(y_2012, order(-y_2012$Value)),]
actividad_5países_2013<-y_2013[with(y_2013, order(-y_2013$Value)),]
actividad_5países_2014<-y_2014[with(y_2014, order(-y_2014$Value)),]

#Se crea una tabla para cada año sobre las Actividades Sanitarias
#(Pacientes en Diálisis) de los 5 Países con un mayor número de ellos( por cada 100000 habitantes).

kable(actividad_5países_2005[0:5,c(2,5)],
      col.names = c("País","P.Trasplantados/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes trasplantados en 2005")

```

Table 12: Países con mayor Número de Pacientes trasplantados en 2005

	País	P.Trasplantados/100000 habitantes
18	Austria	42.54
9	Spain	41.82
23	Finland	41.82
20	Portugal	41.78
1	Belgium	39.88

```

kable(actividad_5países_2006[0:5,c(2,5)],
      col.names = c("País","P.Trasplantados/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes trasplantados en 2006")

```

Table 13: Países con mayor Número de Pacientes trasplantados en 2006

	País	P.Trasplantados/100000 habitantes
18	Austria	44.30
23	Finland	43.43
20	Portugal	42.87
9	Spain	42.74
1	Belgium	42.41

```

kable(actividad_5países_2007[0:5,c(2,5)],
      col.names = c("País","P.Trasplantados/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes trasplantados en 2007")

```

Table 14: Países con mayor Número de Pacientes trasplantados en 2007

	País	P.Trasplantados/100000 habitantes
20	Portugal	47.26
18	Austria	45.67
23	Finland	44.53
1	Belgium	44.23
9	Spain	43.61

```
kable(actividad_5países_2008[0:5,c(2,5)],
      col.names = c("País", "P.Trasplantados/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes trasplantados en 2008")
```

Table 15: Países con mayor Número de Pacientes trasplantados en 2008

	País	P.Trasplantados/100000 habitantes
9	Spain	48.88
20	Portugal	48.49
18	Austria	46.48
23	Finland	44.94
1	Belgium	42.41

```
kable(actividad_5países_2009[0:5,c(2,5)],
      col.names = c("País", "P.Trasplantados/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes trasplantados en 2009")
```

Table 16: Países con mayor Número de Pacientes trasplantados en 2009

	País	P.Trasplantados/100000 habitantes
20	Portugal	54.48
9	Spain	49.45
18	Austria	47.55
23	Finland	45.98
10	France	45.24

```
kable(actividad_5países_2010[0:5,c(2,5)],
      col.names = c("País", "P.Trasplantados/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes trasplantados en 2010")
```

Table 17: Países con mayor Número de Pacientes trasplantados en 2010

	País	P.Trasplantados/100000 habitantes
20	Portugal	56.35
9	Spain	49.28

	País	P.Trasplantados/100000 habitantes
18	Austria	48.90
23	Finland	46.87
17	Netherlands	46.44

```
kable(actividad_5países_2011[0:5,c(2,5)],
      col.names = c("País", "P.Trasplantados/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes trasplantados en 2011")
```

Table 18: Países con mayor Número de Pacientes trasplantados en 2011

	País	P.Trasplantados/100000 habitantes
20	Portugal	54.48
9	Spain	52.25
18	Austria	50.10
17	Netherlands	48.74
10	France	47.79

```
kable(actividad_5países_2012[0:5,c(2,5)],
      col.names = c("País", "P.Trasplantados/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes trasplantados en 2012")
```

Table 19: Países con mayor Número de Pacientes trasplantados en 2012

	País	P.Trasplantados/100000 habitantes
20	Portugal	54.48
9	Spain	54.39
18	Austria	51.78
17	Netherlands	50.89
10	France	49.69

```
kable(actividad_5países_2013[0:5,c(2,5)],
      col.names = c("País", "P.Trasplantados/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes trasplantados en 2013")
```

Table 20: Países con mayor Número de Pacientes trasplantados en 2013

	País	P.Trasplantados/100000 habitantes
9	Spain	55.46
20	Portugal	54.48
17	Netherlands	53.43
18	Austria	53.19
23	Finland	49.09

```
kable(actividad_5países_2014[0:5,c(2,5)],
      col.names = c("País", "P.Trasplantados/100000 habitantes"),
      caption = "Países con mayor Número de Pacientes trasplantados en 2014")
```

Table 21: Países con mayor Número de Pacientes trasplantados en 2014

	País	P.Trasplantados/100000 habitantes
2	Bulgaria	65.00
20	Portugal	54.48
9	Spain	54.39
18	Austria	51.78
17	Netherlands	50.89

A continuación, se agrupa toda la información (P. Trasplantados) por países en una tabla:

```
a1<-group_by(dia_tras_ratio_tras,GEO)#Se agrupa por países
#Se selecciona las variables País y Value (P.Trasplantados/100000 hab)
a2<-select(a1,GEO:Value)
#Se muestra la información por cada país, con Value=suma de
#los valores de cada país en los 10 años.
a3<-(summarize(a2,suma=sum(Value)/10))
a4<-data.frame(a3)#Se convierte la información en un dataframe.
#Se ordena el DataFrame por la variable Suma de forma descendente.
a5<-a4[with(a4,order(-a4$suma)),]

#Se crea una tabla con toda la información
kable(a5[0:5,c(1,2)],
      col.names = c("País", "P.Trasplantados/100000 habitantes"),
      caption = "Países con la mayor media de Pacientes Trasplantados en 2005-2014")
```

Table 22: Países con la mayor media de Pacientes Trasplantados en 2005-2014

	País	P.Trasplantados/100000 habitantes
22	Portugal	50.915
26	Spain	49.227
1	Austria	48.229
8	Finland	46.087
20	Netherlands	45.197

#### • Ocupación de Cama Hospitalaria

```
#####
#Para "Ocupación de Cama Hospitalaria" #
#####

#Se filtra por Año
y_2010<-filter(ocupacion_cama, TIME==2010)
```

```

y_2011<-filter(ocupacion_cama, TIME==2011)
y_2012<-filter(ocupacion_cama, TIME==2012)
y_2013<-filter(ocupacion_cama, TIME==2013)
y_2014<-filter(ocupacion_cama, TIME==2014)
y_2015<-filter(ocupacion_cama, TIME==2015)
y_2016<-filter(ocupacion_cama, TIME==2016)
y_2017<-filter(ocupacion_cama, TIME==2017)
y_2018<-filter(ocupacion_cama, TIME==2018)
y_2019<-filter(ocupacion_cama, TIME==2019)

```

*#Se ordena por "Value"*

```

actividad_5países_2010<-y_2010[with(y_2010, order(-y_2010$Value)),]
actividad_5países_2011<-y_2011[with(y_2011, order(-y_2011$Value)),]
actividad_5países_2012<-y_2012[with(y_2012, order(-y_2012$Value)),]
actividad_5países_2013<-y_2013[with(y_2013, order(-y_2013$Value)),]
actividad_5países_2014<-y_2014[with(y_2014, order(-y_2014$Value)),]
actividad_5países_2015<-y_2015[with(y_2015, order(-y_2015$Value)),]
actividad_5países_2016<-y_2016[with(y_2016, order(-y_2016$Value)),]
actividad_5países_2017<-y_2017[with(y_2017, order(-y_2017$Value)),]
actividad_5países_2018<-y_2018[with(y_2018, order(-y_2018$Value)),]
actividad_5países_2019<-y_2019[with(y_2019, order(-y_2019$Value)),]

```

*#Se crea una tabla para cada año sobre la Ocupación de Cama Hospitalaria de los 5 Países con un mayor p*

```

kable(actividad_5países_2010[0:5,c(2,5)], col.names = c("País","O.Cama %"),
      caption = "Países con mayor % de Ocupación de Cama en 2010")

```

Table 23: Países con mayor % de Ocupación de Cama en 2010

	País	O.Cama %
6	Ireland	91.40
26	Norway	85.60
28	United Kingdom	84.35
12	Cyprus	84.17
27	Switzerland	82.83

```

kable(actividad_5países_2011[0:5,c(2,5)], col.names = c("País","O.Cama %"),
      caption = "Países con mayor % de Ocupación de Camas en 2011")

```

Table 24: Países con mayor % de Ocupación de Camas en 2011

	País	O.Cama %
6	Ireland	91.9
12	Cyprus	90.9
27	Switzerland	84.2
26	Norway	83.7
17	Malta	83.2

```
kable(actividad_5países_2012[0:5,c(2,5)], col.names = c("País","O.Cama %"),
      caption = "Países con mayor % de Ocupación de Cama en 2012")
```

Table 25: Países con mayor % de Ocupación de Cama en 2012

	País	O.Cama %
6	Ireland	92.60
26	Norway	84.80
17	Malta	83.20
27	Switzerland	81.52
1	Belgium	81.11

```
kable(actividad_5países_2013[0:5,c(2,5)], col.names = c("País","O.Cama %"),
      caption = "Países con mayor % de Ocupación de Cama en 2013")
```

Table 26: Países con mayor % de Ocupación de Cama en 2013

	País	O.Cama %
6	Ireland	93.80
26	Norway	83.80
27	Switzerland	82.81
17	Malta	80.66
1	Belgium	80.44

```
kable(actividad_5países_2014[0:5,c(2,5)], col.names = c("País","O.Cama %"),
      caption = "Países con mayor % de Ocupación de Cama en 2014")
```

Table 27: Países con mayor % de Ocupación de Cama en 2014

	País	O.Cama %
6	Ireland	93.30
27	Switzerland	83.18
26	Norway	82.80
17	Malta	81.79
1	Belgium	80.40

```
kable(actividad_5países_2015[0:5,c(2,5)], col.names = c("País","O.Cama %"),
      caption = "Países con mayor % de Ocupación de Cama en 2015")
```

Table 28: Países con mayor % de Ocupación de Cama en 2015

	País	O.Cama %
6	Ireland	89.90
27	Switzerland	83.59
17	Malta	81.73
20	Portugal	80.87



	País	O.Cama %
26	Norway	80.40

```
kable(actividad_5países_2016[0:5,c(2,5)], col.names = c("País","O.Cama %"),
      caption = "Países con mayor % de Ocupación de Cama en 2016")
```

Table 29: Países con mayor % de Ocupación de Cama en 2016

	País	O.Cama %
6	Ireland	89.60
27	Switzerland	83.91
20	Portugal	81.73
1	Belgium	81.43
28	United Kingdom	81.43

```
kable(actividad_5países_2017[0:5,c(2,5)], col.names = c("País","O.Cama %"),
      caption = "Países con mayor % de Ocupación de Cama en 2017")
```

Table 30: Países con mayor % de Ocupación de Cama en 2017

	País	O.Cama %
6	Ireland	89.80
20	Portugal	82.62
27	Switzerland	82.12
1	Belgium	81.84
28	United Kingdom	81.84

```
kable(actividad_5países_2018[0:5,c(2,5)], col.names = c("País","O.Cama %"),
      caption = "Países con mayor % de Ocupación de Cama en 2018")
```

Table 31: Países con mayor % de Ocupación de Cama en 2018

	País	O.Cama %
6	Ireland	90.70
20	Portugal	82.06
1	Belgium	81.99
28	United Kingdom	81.99
27	Switzerland	81.65

```
kable(actividad_5países_2019[0:5,c(2,5)], col.names = c("País","O.Cama %"),
      caption = "Países con mayor % de Ocupación de Cama en 2019")
```

Table 32: Países con mayor % de Ocupación de Cama en 2019

	País	O.Cama %
6	Ireland	89.80
27	Switzerland	82.12
20	Portugal	82.06
28	United Kingdom	81.99
1	Belgium	81.84

A continuación, se agrupa toda la información (% Ocupación de Cama) por países en una tabla:

```
a1<-group_by(ocupacion_cama,GEO)#Se agrupa por paises
#Se selecciona las variables Pais y Value
a2<-select(a1,GEO:Value)
#Se muestra la información por cada país, con Value=suma de
#los valores de cada país en los 10 años.
a3<-(summarize(a2,suma=sum(Value)/10))
a4<-data.frame(a3)#Se convierte la información en un dataframe.
#Se ordena el DataFrame por la variable Suma de forma descendente.
a5<-a4[with(a4,order(-a4$suma)),]

#Se crea una tabla con toda la informacion
kable(a5[0:5,c(1,2)],
      col.names = c("País","O. Cama %"),
      caption = "Países con la mayor media en porcentaje de Ocupación de cama en 2010-2019")
```

Table 33: Países con la mayor media en porcentaje de Ocupación de cama en 2010-2019

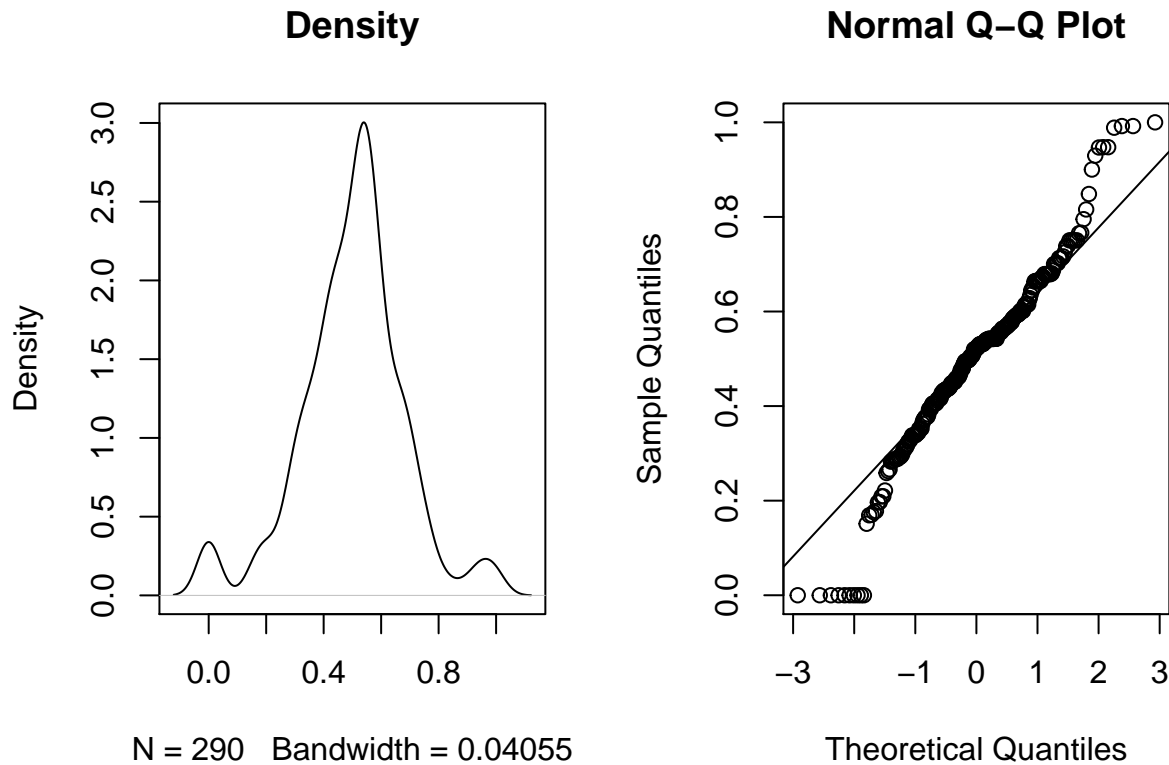
	País	O. Cama %
13	Ireland	91.280
29	Switzerland	82.793
22	Norway	82.390
19	Malta	81.735
31	United Kingdom	81.419

### • 1.1.3 Normalidad de la variable “Value”

Se comprueba con métodos visuales si la variable tiene una distribución normal.

## PACIENTES EN DIÁLISIS

```
par(mfrow=c(1,2))
plot(density(dia_tras_ratio_dia$Value_norm) ,main="Density")
qqnorm(dia_tras_ratio_dia$Value_norm)
qqline(dia_tras_ratio_dia$Value_norm)
```



Para estudiar si una muestra proviene de una población con distribución normal, se disponen de tres herramientas:

- Histograma o Densidad
- Gráficos cuantil cuantil (QQplot)
- Pruebas de hipótesis.

Si en la prueba de Densidad se observa sesgo hacia uno de los lados de la gráfica, sería indicio de que la muestra no proviene de una población normal. Si por otra parte, sí se observa simetría, **NO** se garantiza que la muestra provenga de una población normal. En estos casos sería necesario utilizar otras herramientas como **QQplot y pruebas de hipótesis**.

En la gráfica Densidad de la variable "Value", no se observa claramente sesgo hacia ningún lado, por lo que no se puede descartar normalidad. Se puede confirmar observando la gráfica QQplot en la que la línea que grafica qqline sirve de referencia para interpretar el gráfico. Si se tuviese una muestra distribuida normalmente, se esperaría que los puntos del gráfico cuantil cuantil estuviesen perfectamente alineados con la línea de referencia, y observamos que para este caso, "Value" no se alinea, tan solo un poco en el centro.

Por otro lado, se realizan las pruebas de hipótesis:

- $H_0$ : La muestra proviene de una población normal.
- $H_1$ : La muestra NO proviene de una población normal.

Se aplica la prueba Shapiro-Wilk:

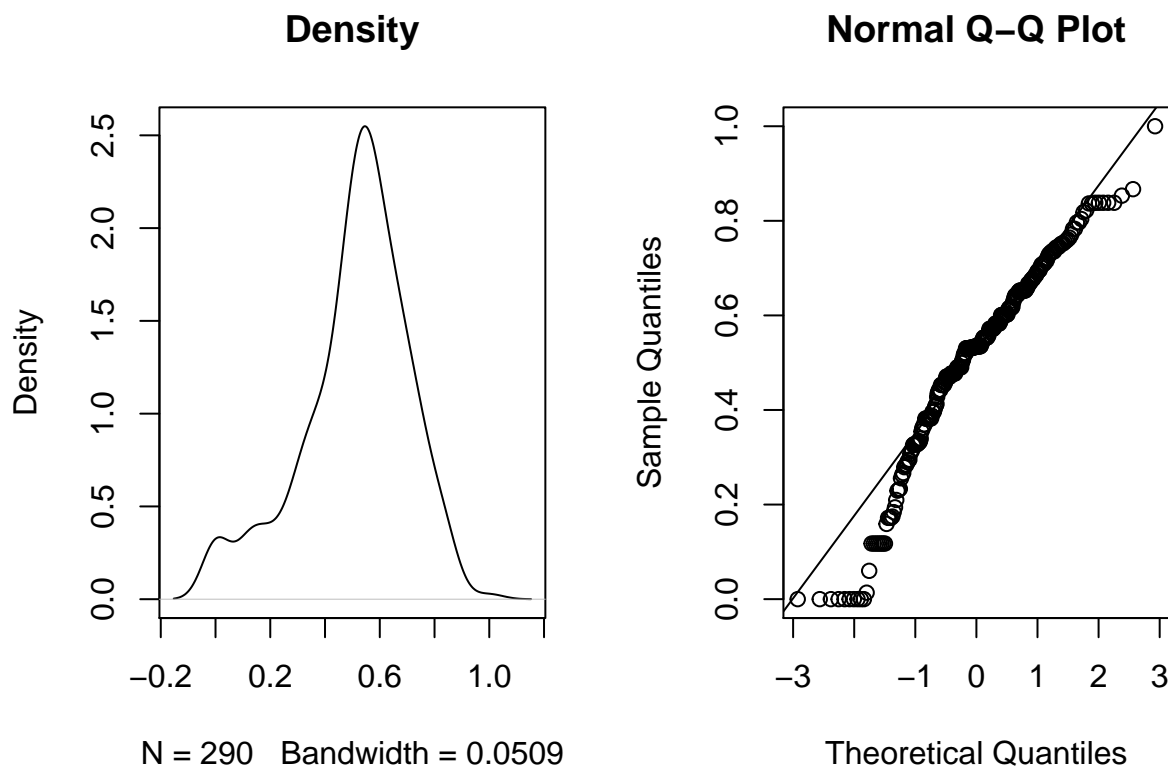
```
shapiro.test(dia_tras_ratio_dia$Value_norm)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  dia_tras_ratio_dia$Value_norm  
## W = 0.9648, p-value = 1.648e-06
```

Se observa un p-value muy pequeño, mucho más pequeño que cualquier nivel de significación ( $\alpha=0.5$ ) por lo que se rechaza la hipótesis nula y asumimos **No normalidad** en la muestra.

### PACIENTES TRASPLANTADOS

```
par(mfrow=c(1,2))  
plot(density(dia_tras_ratio_tras$Value_norm) ,main="Density")  
qqnorm(dia_tras_ratio_tras$Value_norm)  
qqline(dia_tras_ratio_tras$Value_norm)
```



En la gráfica Densidad de la variable “Value”, se observa un cierto grado de sesgo hacia la derecha, por lo que no se considera normalidad. Se puede confirmar observando la gráfica QQplot en la que la línea que grafica qqline sirve de referencia para interpretar el gráfico, no se alinea con los puntos de los valores de la variable “Value”, tan solo un poco en la parte central.

Tras aplicar la prueba Shapiro-Wilk:

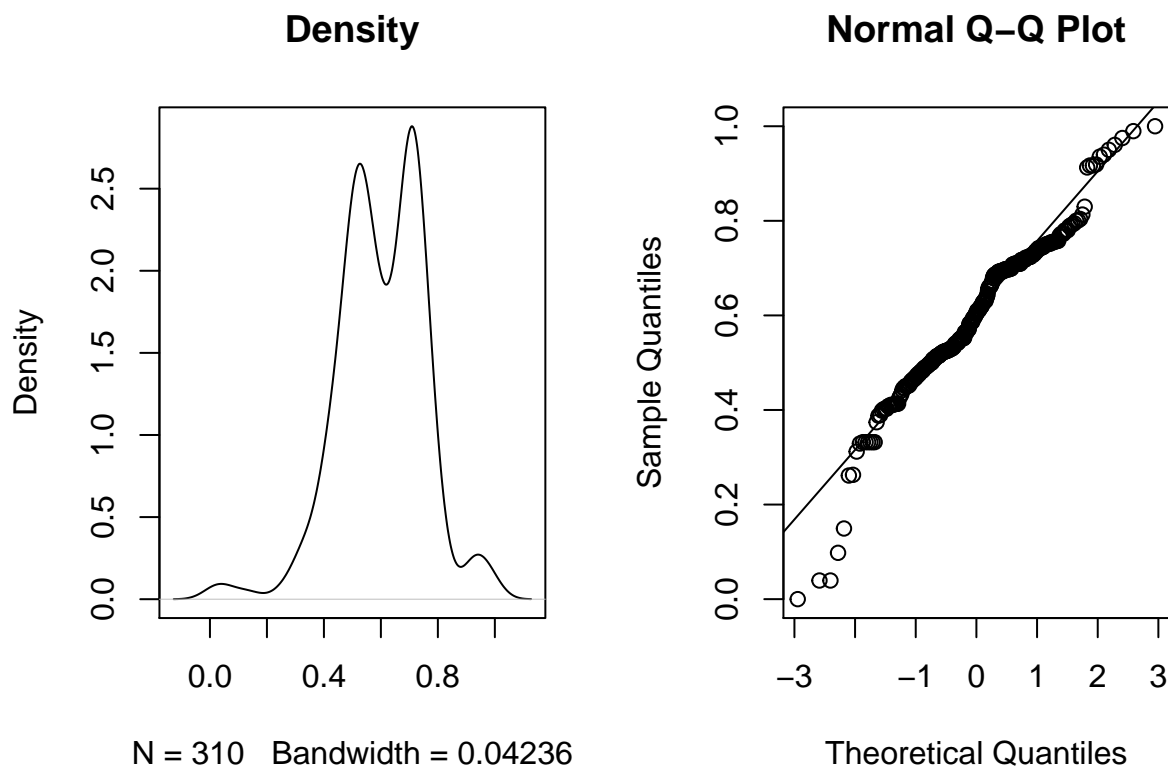
```
shapiro.test(dia_tras_ratio_tras$Value_norm)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  dia_tras_ratio_tras$Value_norm  
## W = 0.9554, p-value = 9.659e-08
```

Se observa un p-value muy pequeño, mucho más pequeño que cualquier nivel de significación ( $\alpha=0.5$ ) por lo que se rechaza la hipótesis nula y asumimos **No normalidad** en la muestra.

### OCUPACIÓN DE CAMA %

```
par(mfrow=c(1,2))  
plot(density(ocupacion_cama$Value_norm) ,main="Density")  
qqnorm(ocupacion_cama$Value_norm)  
qqline(ocupacion_cama$Value_norm)
```



En la gráfica Densidad de la variable “Value”, no se observa sesgo hacia ningún lado , pero la forma de la gráfica no se acerca a lo que se considera normalidad. Se puede confirmar observando la gráfica QQplot en la que la línea que grafica qqline sirve de referencia para interpretar el gráfico, no se alinea con los puntos de los valores de la variable “Value”, tan solo un poco en la parte central.

Tras aplicar la prueba Shapiro-Wilk:

```
shapiro.test(ocupacion_cama$Value_norm)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  ocupacion_cama$Value_norm  
## W = 0.96308, p-value = 4.391e-07
```

Se observa un p-value muy pequeño, mucho más pequeño que cualquier nivel de significación ( $\alpha=0.5$ ) por lo que se rechaza la hipótesis nula y asumimos **No normalidad** en la muestra.