
**University of Granada**

# Exploring the data of 121 patients who underwent glaucoma surgery. Conclusions.

*Multivariate Statistics*

Author:
Quintín Mesa Romero

# 1 Introduction

Glaucoma is a leading cause of irreversible blindness worldwide. It is a disease of a progressive optic neuropathy with loss of retinal neurons and their axons, which can result in blindness in case of untreatment[1][2]. In short, it is a group of diseases that kill retinal ganglion cells [2].

The strongest known risk factor is high IOP (Intraocular Pressure), but it is not the only factor responsible for glaucoma [2]. In fact, people with myopia greater than five dioptres, people aged 60 years or more, people with thin cornea, and even people with different skin type, such as africans or afro-caribbean are more likely to develop glaucoma. Of course, having family history multiplies the risk of developing the desease [3].

Given that there are about 80 million people suffering from glaucoma, and it is estimated that over 112 million individuals will have it by 2040, it is reasonable to think that there is a treatment or an operation to reverse the disease before it is too late [4]. Indeed, there is a surgery based on laser technology which is being applied to people with glaucoma[5].

From a research point of view, in relation to this surgery, we wonder whether the pre-surgery condition is related, in any way, to the long-term progression. For this purpose, we have studied a dataset with information of 121 patients who underwent glaucoma surgery using laser technology, in which variables have been meassured before and after the surgery over a three-month period at different time intervals.

Studying this relationship is crucial, primarily because it can inform better treatment strategies and patient outcomes, and clinicians can better predict which patients are more likely to benefit from laser surgery versus those who may need alternative or additional interventions, in order to improve their lives.

# 2 Methods and techniques

## 2.1 Data Collection, Preparation and Cleaning

The data from the 121 patients have been stored in an Excel file. This dataset includes clinical variables related to presurgical conditions and post-surgical outcomes after the laser-based surgery. The dataset has been processed, prepared and cleaned using the **R** statistical environment.

First of all, we have to load the library **readxl**:

```
# Load necessary libraries
```{r}
library(readxl)  # for reading excel files
```
```

Now, we are ready to load the dataset. For this purpose, we use the function **read_excel**:

```
#Load the dataset
```{r}
data_glaucoma<-read_excel("Glaucoma_DB.xlsx", sheet="DATOS")
```
```

Once the dataset is loaded, it is advisable to check the structure of the data, and for this, we use **str**:

```r
# Check the structure of the dataset
```{r}
str(data_glaucoma)
```
```

```
tibble [121 × 19] (S3: tbl_df/tbl/data.frame)
 $ OJO            : num [1:121] 0 1 0 1 0 1 0 0 0 1 ...
 $ TIPO_GLAUCOMA  : num [1:121] 0 NA 1 2 2 1 1 3 1 1 ...
 $ N_IMPACTOS     : num [1:121] 112 108 123 131 156 125 178 164 109 116
...
 $ CUADRANTES     : chr [1:121] "4" "4" "4" "4" ...
 $ ENERGIA_IMPACTO: chr [1:121] "1.5" "1.2" "1.1000000000000001" "1.5" ...
 $ ENERGIA_TOTAL  : num [1:121] 174 128 133 191 182 170 249 301 109 238
...
 $ CIRUJIA_PREVIA : num [1:121] NA 1 1 1 1 0 0 0 1 0 ...
 $ PIO_PRE_SLT    : num [1:121] 31 29 36 14 14 30 36 25 23 22 ...
 $ PIO_1_SEMANA   : num [1:121] 0 23 30 0 0 0 0 0 0 22 ...
 $ PIO_1_MES      : chr [1:121] "0" "19" "30" "21" ...
 $ PIO_3_MES      : num [1:121] 0 24 30 14 17 20 19 0 16 20 ...
 $ FARMACOS_PRE   : chr [1:121] "3" "3" "1" "1" ...
 $ FARMACOS_1_MES : num [1:121] 0 4 4 0 0 3 3 0 2 0 ...
 $ FARMACOS_3_MES : num [1:121] 0 4 4 0 0 3 3 0 2 0 ...
 $ DOLOR          : num [1:121] 0 1 1 1 1 1 1 0 0 1 ...
 $ SEXO           : num [1:121] NA 0 0 1 1 1 1 0 1 0 ...
 $ EDAD           : num [1:121] 0 56 56 49 49 74 74 65 60 82 ...
 $ PIO_NORMAL     : chr [1:121] "0" "19" "30" "21" ...
 $ PIO_NORMAL_CAT : num [1:121] 1 0 1 0 0 0 0 0 1 0 ...
```

Also, in order to not to alter the original dataset, it is advisable to make a copy and work with it in the future. We will load the variable names from the dataset too. For that, we make:

```r
# copy of the dataset
```{r}
data_glaucoma_copy <- data_glaucoma
attach(data_glaucoma_copy)

```
```

It is easy to see that there are missinng values, so we have to handle them. We can use the function **colSums** with the argument **is.na(data_glaucoma)**, which tells us the number NA values that there are for each variable:

```r
# Check for missing values
```{r}
missing_values <- colSums(is.na(data_glaucoma))
print(missing_values)
```
```

```
            OJO   TIPO_GLAUCOMA      N_IMPACTOS      CUADRANTES
              4               2               0               0
ENERGIA_IMPACTO   ENERGIA_TOTAL  CIRUJIA_PREVIA     PIO_PRE_SLT
              0               0              36               0
   PIO_1_SEMANA       PIO_1_MES       PIO_3_MES    FARMACOS_PRE
              0               0               0               0
 FARMACOS_1_MES  FARMACOS_3_MES           DOLOR            SEXO
              0               0              60              61
           EDAD      PIO_NORMAL  PIO_NORMAL_CAT
              0               0               0
```

There are different options for handling **NA** data. One could be simply removing the rows with missing values, but it might lead to the loss of valuable data and even statistical power. Other option is to replace missing values with the mean or the **median**. We will used the last one; we will replace the missing data by the median of the corresponding variable. To do this, we have made a for loop in which for each variable with missing values, each of these NA values are replaced by the median of the rest of the values of the corresponding variable.

```r
# replacing missing values by the median

```{r}
for (col_name in names(missing_values[missing_values > 0])) {
    data_glaucoma_copy[[col_name]][is.na(data_glaucoma_copy[[col_name]])] <- median(data_glaucoma_copy[[col_name]], na.rm = TRUE)
}
```
```

We check the change:

```{r}
head(data_glaucoma_copy,10)
```

A tibble: 10 × 19

| OJO <dbl> | TIPO_GLAUCOMA <dbl> | N_IMPACTOS <dbl> | CUADRANTES <chr> | ENERGIA_IMPACTO <chr> | ENERGIA_TOTAL <dbl> | CIRUJIA_PREVIA <dbl> | PIO_PRE_SLT <dbl> | PIO_1_SEMANA <dbl> | PIO_1_MES <chr> |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 112 | 4 | 1.5 | 174 | 1 | 31 | 0 | 0 |
| 1 | 4 | 108 | 4 | 1.2 | 128 | 1 | 29 | 23 | 19 |
| 0 | 1 | 123 | 4 | 1.1000000000000001 | 133 | 1 | 36 | 30 | 30 |
| 1 | 2 | 131 | 4 | 1.5 | 191 | 1 | 14 | 0 | 21 |
| 0 | 2 | 156 | 4 | 1.2 | 182 | 1 | 14 | 0 | 16 |
| 1 | 1 | 125 | 4 | 1.4 | 170 | 0 | 30 | 0 | 18 |
| 0 | 1 | 178 | 4 | 1.4 | 249 | 0 | 36 | 0 | 20 |
| 0 | 3 | 164 | 4 | 1.9 | 301 | 0 | 25 | 0 | 18 |
| 0 | 1 | 109 | 4 | 1 | 109 | 1 | 23 | 0 | 10 |
| 1 | 1 | 116 | 4 | 2.2000000000000002 | 238 | 0 | 22 | 22 | 20 |

1-10 of 10 rows | 1-10 of 19 columns

```{r}
colSums(is.na(data_glaucoma_copy))
```

```
            OJO   TIPO_GLAUCOMA       N_IMPACTOS      CUADRANTES ENERGIA_IMPACTO   ENERGIA_TOTAL
              0               0               0               0               0               0
CIRUJIA_PREVIA     PIO_PRE_SLT    PIO_1_SEMANA      PIO_1_MES       PIO_3_MES    FARMACOS_PRE
              0               0               0               0               0               0
FARMACOS_1_MES  FARMACOS_3_MES          DOLOR            SEXO            EDAD     PIO_NORMAL
              0               0               0               0               0               0
PIO_NORMAL_CAT
              0
```

Then, we convert the **categorical** variables of the dataset into **factor** type:

```{r}
data_glaucoma_copy$SEXO <- factor(data_glaucoma_copy$SEXO, levels = c("1", "0"),labels = c("Hombre", "Mujer"))

data_glaucoma_copy$OJO <- factor(data_glaucoma_copy$OJO, levels = c("0","1"), labels = c("Izdo", "dcho"))

data_glaucoma_copy$TIPO_GLAUCOMA <- factor(data_glaucoma_copy$TIPO_GLAUCOMA, levels = c("0","1","2", "3","4", "5", "6", "7", "8", "9", "10",
"11", "12", "13", "14","15","16"), labels = c("PIGMENTARIO", "GPPA", "GPAA", "GLAUCO PIGMENT","GPAC", "G.PSEUDOEX", "SD DISPERSION
PIGMENTARIA","GLAUCOMA CONGENITO", "GLAUCOMA POSTRABECULAR", "HTO", "CPAC", "GSAA", "GS PXE", "GCS","HTO PIGMENTARI", "GPAA?miópico",
"GPAA?"))

data_glaucoma_copy$CIRUJIA_PREVIA <- factor(data_glaucoma_copy$CIRUJIA_PREVIA, levels = c("1", "0"), labels = c("No", "Si"))

data_glaucoma_copy$DOLOR <- factor(data_glaucoma_copy$DOLOR, levels = c("0","1"), labels = c("Si", "No"))

```

Once all these changes have been applied to the data, we obtain:

```{r}
str(data_glaucoma_copy)
```

```
tibble [121 × 19] (S3: tbl_df/tbl/data.frame)
 $ OJO           : Factor w/ 2 levels "Izdo","dcho": 1 2 1 2 1 2 1 1 1 2 ...
 $ TIPO_GLAUCOMA : Factor w/ 17 levels "PIGMENTARIO",..: 1 5 2 3 3 2 2 4 2 2 ...
 $ N_IMPACTOS    : num [1:121] 112 108 123 131 156 125 178 164 109 116 ...
 $ CUADRANTES    : chr [1:121] "4" "4" "4" "4" ...
 $ ENERGIA_IMPACTO: chr [1:121] "1.5" "1.2" "1.1000000000000001" "1.5" ...
 $ ENERGIA_TOTAL : num [1:121] 174 128 133 191 182 170 249 301 109 238 ...
 $ CIRUJIA_PREVIA : Factor w/ 2 levels "No","Si": 1 1 1 1 1 2 2 2 1 2 ...
 $ PIO_PRE_SLT   : num [1:121] 31 29 36 14 14 30 36 25 23 22 ...
 $ PIO_1_SEMANA  : num [1:121] 0 23 30 0 0 0 0 0 0 22 ...
 $ PIO_1_MES     : chr [1:121] "0" "19" "30" "21" ...
 $ PIO_3_MES     : num [1:121] 0 24 30 14 17 20 19 0 16 20 ...
 $ FARMACOS_PRE  : chr [1:121] "3" "3" "1" "1" ...
 $ FARMACOS_1_MES : num [1:121] 0 4 4 0 0 3 3 0 2 0 ...
 $ FARMACOS_3_MES : num [1:121] 0 4 4 0 0 3 3 0 2 0 ...
 $ DOLOR         : Factor w/ 2 levels "Si","No": 1 2 2 2 2 2 2 1 1 2 ...
 $ SEXO          : Factor w/ 2 levels "Hombre","Mujer": 1 2 2 1 1 1 1 2 1 2 ...
 $ EDAD          : num [1:121] 0 56 56 49 49 74 74 65 60 82 ...
 $ PIO_NORMAL    : chr [1:121] "0" "19" "30" "21" ...
 $ PIO_NORMAL_CAT : num [1:121] 1 0 1 0 0 0 0 0 1 0 ...
```

A tibble: 121 × 19

| OJO <fct> | TIPO_GLAUCOMA <fct> | N_IMPACT... <dbl> | CUADRANT... <chr> | ENERGIA_IMPACTO <chr> | ENERGIA_TOT... <dbl> | CIRUJIA_PREVIA <fct> | PIO_PRE_SLT <dbl> | PIO_1_SEMA... <dbl> |
|---|---|---|---|---|---|---|---|---|
| Izdo | PIGMENTARIO | 112 | 4 | 1.5 | 174.0 | No | 31 | 0 |
| dcho | GPAC | 108 | 4 | 1.2 | 128.0 | No | 29 | 23 |
| Izdo | GPPA | 123 | 4 | 1.1000000000000001 | 133.0 | No | 36 | 30 |
| dcho | GPAA | 131 | 4 | 1.5 | 191.0 | No | 14 | 0 |
| Izdo | GPAA | 156 | 4 | 1.2 | 182.0 | No | 14 | 0 |
| dcho | GPPA | 125 | 4 | 1.4 | 170.0 | Si | 30 | 0 |
| Izdo | GPPA | 178 | 4 | 1.4 | 249.0 | Si | 36 | 0 |
| Izdo | GLAUCO PIGMENT | 164 | 4 | 1.9 | 301.0 | Si | 25 | 0 |
| Izdo | GPPA | 109 | 4 | 1 | 109.0 | No | 23 | 0 |
| dcho | GPPA | 116 | 4 | 2.2000000000000002 | 238.0 | Si | 22 | 22 |

### 2.2 Outliers Detection and Treatment

## 3 Results

## 4 Conclusions

## 5 References


[1 ] Glaucoma https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8473801/

[2 ] https://eyes.arizona.edu/sites/default/files/glaucoma.pdf

[3 ] https://www.clinicbarcelona.org/en/assistance/diseases/glaucoma/risk-factors-and-causes

[4 ] https://www.glaucomapatients.org/basic/statistics/, https://glaucoma.org/articles/glaucoma-worldwide-a-growing-concern

[5 ] https://glaucoma.org/treatment/laser