SLNB-Data

LU

2023-01-07

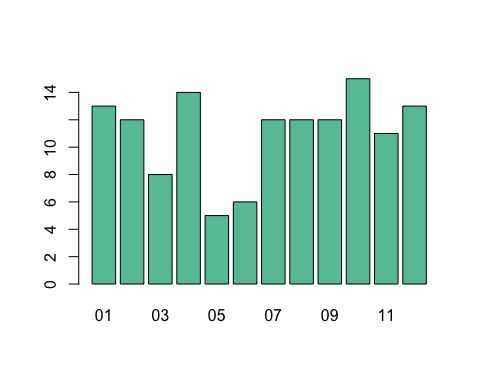
# Summaries of our Data

We have altogether 133 patients. Among them are 93 females and 40 males.

## Hospitalizations by Month

For each patient find the month when s/he was hospitalized (stacionēšanas datuma mēnesis). Our test data cover the entire year 2015, so our selection should not be *biased by month* (nav sistemātiskas novirzes - visiem gada mēnešiem bija vienādas izredzes nonākt datu izlasē). We still do have a tiny bias, since not all months have equal number of days, but we ignore that.

Draw a barplot - the number of patients by month. (Joslu diagramma.)



We formulate null hypothesis as follows: “Hospitalization months can be assumed to be uniform - there are no months that are overrepresented or underrepresented”. There could be reasons why null hypothesis could fail (if patients undergo more testing immediately beforeor after their summer vacations; or, perhaps, they were scheduled to a different month during their initial registration procedure).

Our data cannot find any difference from the uniform distribution. We calculate [Pearson chi-squared test](https://en.wikipedia.org/wiki/Pearson%27s_chi-squared_test#Calculating_the_test-statistic) to find out the difference between actual data and the theoretical/uniform distribution.

# our actual patients per month (January to December)  
patientsByMonth <- as.vector(table(df$month))  
# distribution, if the distribution is uniform  
uniformByMonth <- rep(133/12, rep=12)  
chiSquareExpr <- sum((patientsByMonth - uniformByMonth)^2/uniformByMonth)  
chiSquareExpr

## [1] 9.646617

# Find the probability for 12-1 = 11 degrees of freedom  
dchisq(chiSquareExpr, df=11)

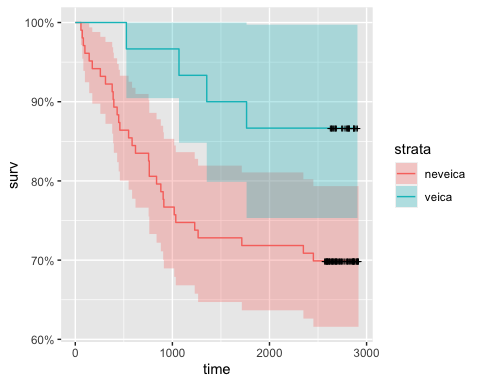
## [1] 0.09129136

The probability that the Chi-squared expression chiSquareExpr is more or equal to the computed value 9.6 is 0.09129136. It is more than 0.05, so we cannot reject the null hypothesis. There is about 9% chance that uniformly distributed patients would be distributed like in our actual distribution or even more unequally.

If the number of observations were considerably larger, we could, in fact, notice small differences between months.

## Survival Data

In this section we draw some Kaplan-Meier diagrams see [Survival analysis by R](https://rviews.rstudio.com/2017/09/25/survival-analysis-with-r/) and other survival-related correlations.



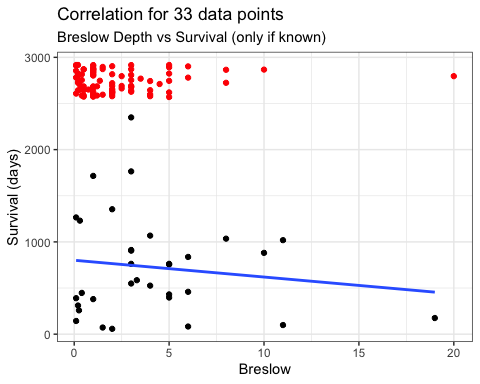
Chi-squared criterion - what is our confidence that undergoing the sentinel lymph node biopsy (SLNB) does indeed improve the survival time has not been done yet.

## Breslow’s Depth and Survival Time

There are 98 patients with their survival time (*bezprogresijas dzīvildze*) unknown - they have no further progression events for their illness or, perhaps, are *censored data* (izvākti jeb nepilnīgi novērojumu dati). They are shown as dots at the top of the correlation diagram, but their data is not used to compute Pearson’s correlation.

Furthermore, there exist 8 patients with their Breslow’s Depth unknown. These data points are not shown.  
Altogether there are 100 patients for which either Breslow Depth or survival is not known.

## `geom\_smooth()` using formula = 'y ~ x'



Numeric value of Pearson Correlation coefficient:

result <- cor(dfCleared$BreslowNum, dfCleared$time, method = "pearson")  
   
# Print the result  
cat("Pearson correlation coefficient is:", result)

## Pearson correlation coefficient is: -0.1364634

This correlation is very weak, but it does not account for the fact that there are considerably more patients with low Breslow Depths who are still surviving (and their exact survival time is not known). If we set their survival time to be December 31, 2022, we get a different picture:

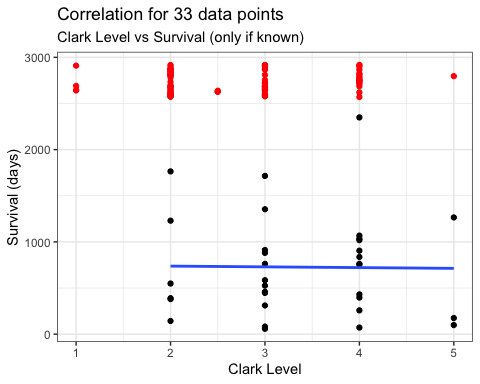
missingPoints0 <- which(is.na(df$BreslowNum))  
dfCleared0 <- df[setdiff(1:nrow(df), missingPoints0),]  
dfMissing0 <- df[missingPoints0,]  
  
  
result <- cor(dfCleared0$BreslowNum, dfCleared0$time, method = "pearson")  
   
# Print the result  
cat("Pearson correlation coefficient is:", result)

## Pearson correlation coefficient is: -0.2620625

As the actual survival times become known, the correlation coefficient is likely to take an even more negative value.

## Clark Levels and Survival

## `geom\_smooth()` using formula = 'y ~ x'



With Clark Level the situation is similar to Breslow Depth: It has very little effect upon those patients who had survival time less than some 3000 days; on the other hand there are considerably more people with low Clark Levels who are still surviving.